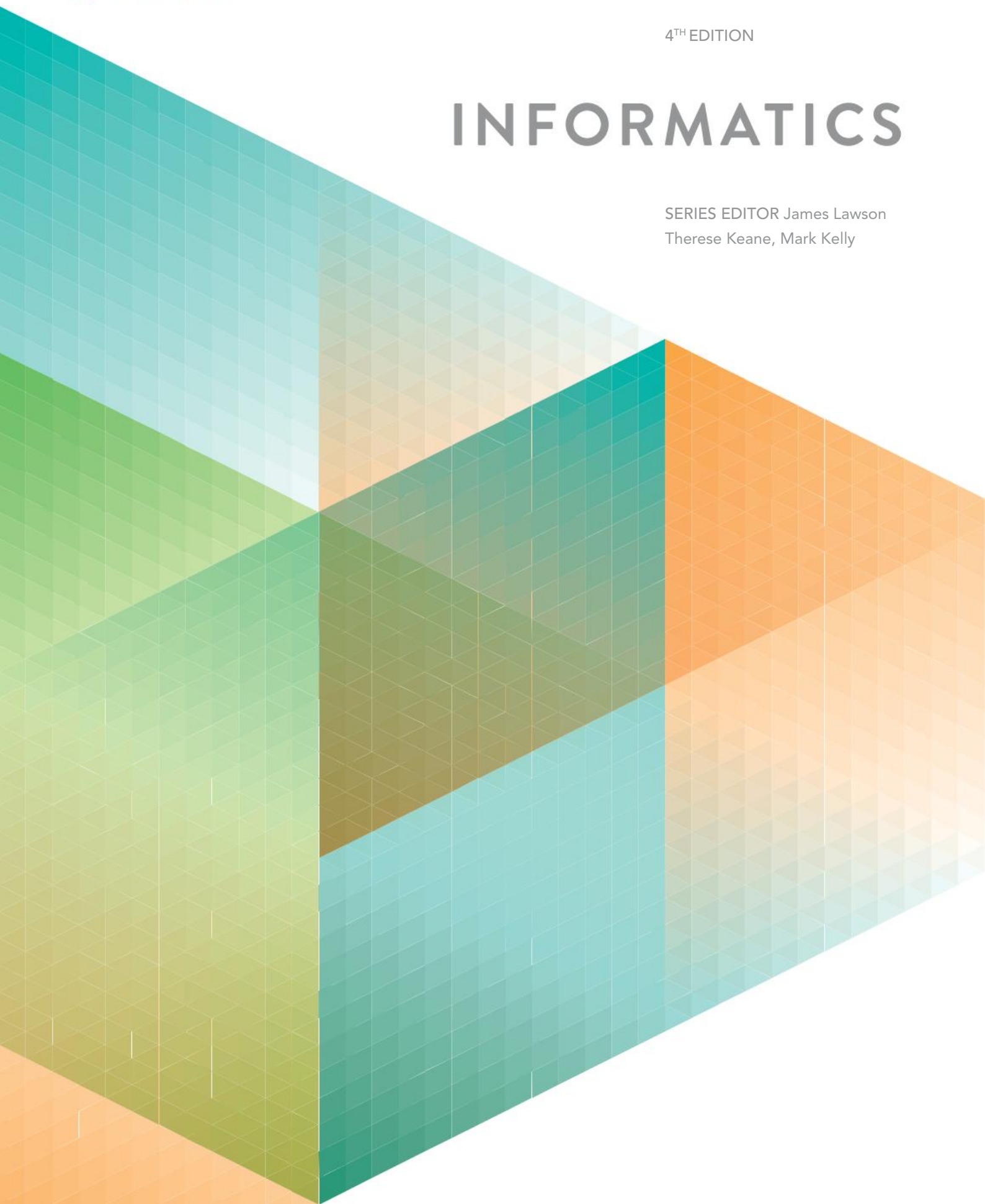
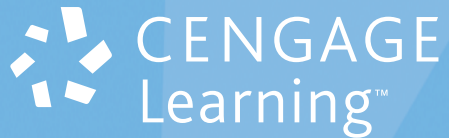


INFORMATICS

SERIES EDITOR James Lawson
Therese Keane, Mark Kelly





COPYRIGHT NOTICE

Copyright in this work is owned by Cengage Learning Australia ("the work"). A condition of purchase of this electronic version of the work is that you agree to respect the copyright in the work, abide by the Copyright Act 1968 and specifically agree not to transfer, sell, assign, misuse, copy or transmit an electronic or other version of the work to any third party.

Please note: This product is accompanied by a licence (single user, network or adoption) governing the terms and conditions of its use.

This is a legal agreement between the you, (the “Customer”) and Cengage Learning Australia Pty Limited (ABN 14 058 280 149) (the “Licensor”) which provides the terms and conditions of this non-exclusive licence and the limited warranty for the Product. Use of the Product indicates an acknowledgement that the Customer has read and agreed to be bound by the terms and conditions of this Agreement. If you do not agree to these terms and conditions, return the Product to the place of purchase within 15 days of the date of purchase (with proof of purchase) for a full refund

1. Licence Grant

You do not receive title to the Product. Copyright in the Product (which includes all images, photographs, video, animations, audio, music and text incorporated in the Product, including all of the accompanying printed material) is owned by the Licensor and/or its suppliers and is protected by Australian copyright laws. The Licensor grants you a non-exclusive licence to use the Product subject to the restrictions and terms set out in this Agreement.

2. A Licence allows you to:

Use the Product on your computer. The Customer represents that they shall in no way place the Product in the public domain or in any way compromise our copyright in the Material. You agree to take reasonable steps to protect our copyright.

3. You may not:

Alter, modify, translate, reverse engineer, decompile, or adapt the software or create derivative works based on the Product. Make further copies by any means technological, electronic, digital whatsoever without the written permission of the Licensor. Rent or transfer all or any part of your rights under this Agreement. Remove or alter any copyright or other proprietary notice or label attached to the software.

4. Termination

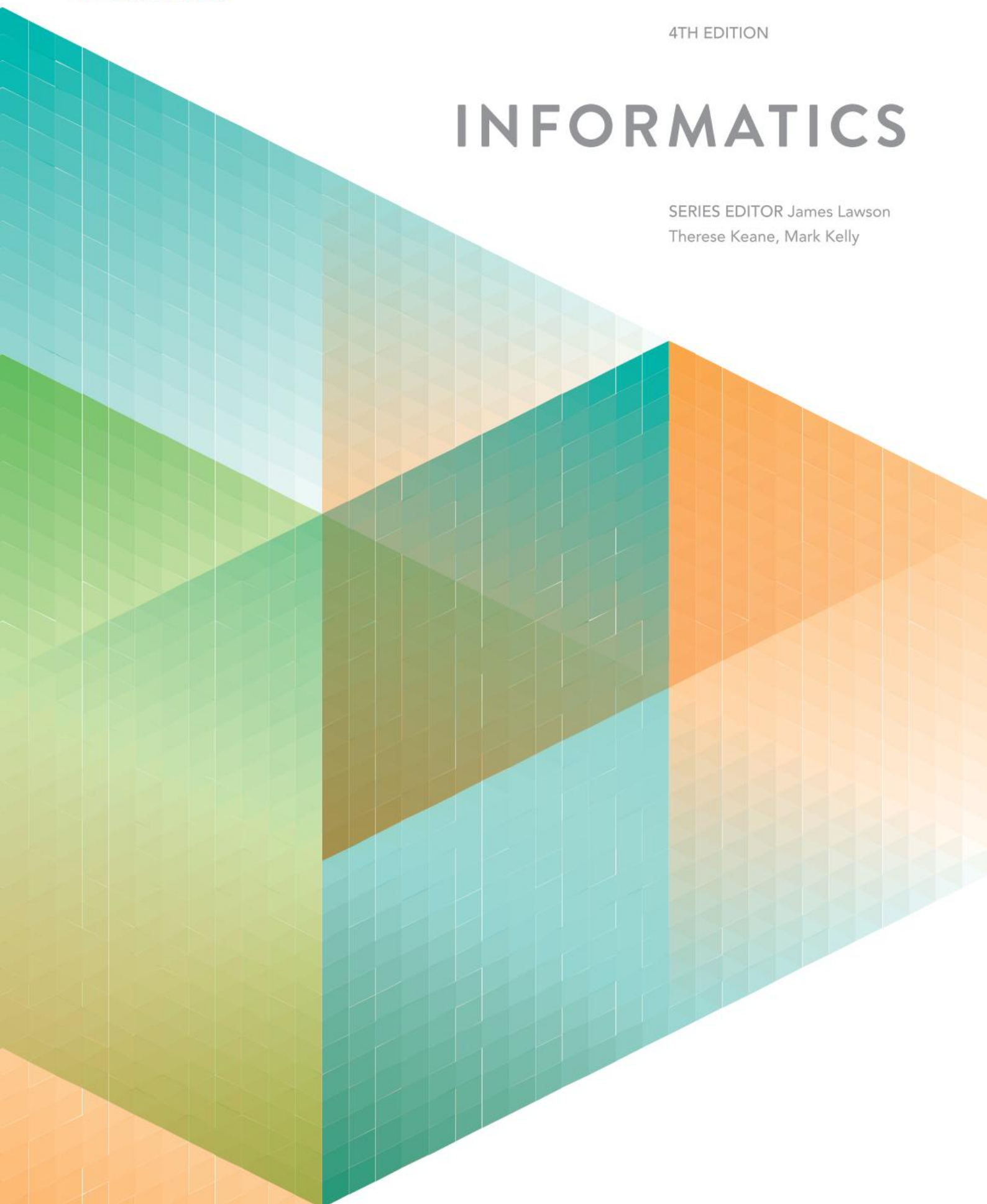
Any failure to comply with the terms and conditions of this agreement will result in the automatic termination of this licence. Upon termination of this licence for any reason, the Customer must destroy or return to the Licensor all copies of the software and accompanying documentation.

5. Warranties

To the extent permitted by law, the Licensor’s liability for any breach of the warranty or any term implied by law into this licence is limited to the lowest cost of replacing the goods, acquiring equivalent goods or having the goods repaired.

INFORMATICS

SERIES EDITOR James Lawson
Therese Keane, Mark Kelly



Informatics VCE Units 3 & 4**4th Edition****James Lawson****Therese Keane****Mark Kelly**

Senior Publishing Editor: Eleanor Gregory

Editor: Nadine Anderson-Conklin

Copyeditors: Gene Anderson-Conklin and Nadine Anderson-Conklin

Proofreader: Adaline Tallarida

Indexer: Russell Brooks

Permissions researcher: Debbie Gallagher

Cover designer: Chris Starr

Text designer: Kim Ferguson

Senior Production Controller: Julie McArthur

Typeset by: Q2A Media

Any URLs contained in this publication were checked for currency during the production process. Note, however, that the publisher cannot vouch for the ongoing currency of URLs.

Acknowledgements

VCE Computing Study Design (2016) © The Victorian Curriculum and Assessment Authority (VCAA). Used with permission.

VCE® is a registered trademark of the VCAA. The VCAA does not endorse or make any warranties regarding this Cengage product. Current and past VCE Study Designs, VCE exams and related content can be accessed directly at www.vcaa.vic.edu.au

© 2016 Cengage Learning Australia Pty Limited

Copyright Notice

This Work is copyright. No part of this Work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means without prior written permission of the Publisher. Except as permitted under the *Copyright Act 1968*, for example any fair dealing for the purposes of private study, research, criticism or review, subject to certain limitations. These limitations include: Restricting the copying to a maximum of one chapter or 10% of this book, whichever is greater; providing an appropriate notice and warning with the copies of the Work disseminated; taking all reasonable steps to limit access to these copies to people authorised to receive these copies; ensuring you hold the appropriate Licences issued by the Copyright Agency Limited ("CAL"), supply a remuneration notice to CAL and pay any required fees. For details of CAL licences and remuneration notices please contact CAL at Level 15, 233 Castlereagh Street, Sydney NSW 2000, Tel: (02) 9394 7600, Fax: (02) 9394 7601

Email: info@copyright.com.auWebsite: www.copyright.com.au

For product information and technology assistance,
in Australia call **1300 790 853**;
in New Zealand call **0800 449 725**

For permission to use material from this text or product, please email
aust.permissions@cengage.com

National Library of Australia Cataloguing-in-Publication Data

Lawson, James (James Austin), 1971– author.

Informatics VCE units 3 & 4 / James Lawson, Therese Keane, Mark Kelly.

4th edition.

9780170364829 (paperback)

Includes index.

For secondary school age.

Computer science--Study and teaching (Secondary)
Information technology--Study and teaching (Secondary)
Victorian Certificate of Education examination.

Keane, Therese, author.

Kelly, Mark John, author.

004

Cengage Learning Australia

Level 7, 80 Dorcas Street

South Melbourne, Victoria Australia 3205

Cengage Learning New Zealand

Unit 4B Rosedale Office Park

331 Rosedale Road, Albany, North Shore 0632, NZ

For learning solutions, visit cengage.com.au

Printed in Australia by Ligare Pty Limited.

1 2 3 4 5 6 7 20 19 18 17 16



CONTENTS

Preface	v
About the authors	vi
How to use this book	vii
Outcomes	viii
Problem-solving methodology	xii
Key concepts	xiv

UNIT 3

1

Introduction

CHAPTER 1 ORGANISATIONS AND DATA MANAGEMENT

2

Collecting data online	3
Why users supply data for online transactions	5
Techniques used by organisations to acquire data online	9
User flow diagrams	11
Techniques for efficient and effective data collection	13
Design principles	15
Protection of rights	20
Characteristics of data types	22
Databases	23
Creating an RDBMS structure	26
Database design tools	36
Test data	41
Developing a RDBMS	44
Testing the solution	49
Protecting valuable data	57
Preparing for Unit 3, Outcome 1	70

CHAPTER 2 DATA ANALYTICS: DRAWING CONCLUSIONS, PART 1

72

What is data?	73
What is a hypothesis?	73
Solution specifications	78

Primary and secondary data	82
Quantitative and qualitative data	85
Data sources	91
Acquiring data	94
Referencing data sources	107
Integrity of data	111
Data types and data structures	121
Next steps	123

CHAPTER 3 DATA ANALYTICS: DRAWING CONCLUSIONS, PART 2

128

Continuing Unit 3, Outcome 2	129
Legal requirements	129
Project management	134
File naming strategies	144
Organising and storing data	146
Identifying patterns and relationships between data	152
Digital systems	165
Data security	170
Next steps	175
Preparing for Unit 3, Outcome 2	180

UNIT 4**183**

Introduction

CHAPTER 4 DATA ANALYTICS: PRESENTING THE FINDINGS**184**

What is a multimodal online solution?	185
An effective MMOS	186
Manipulating data	199
Formats and conventions	202
Design principles	204
Generating design ideas	209
Design tools	220
Documenting the progress of projects	223
Managing files	227
Manipulating data with software	231
Testing	232
Evaluating your solution	237
Assessing your project plan	239
Next steps	240
Preparing for Unit 4, Outcome 1	244

CHAPTER 5 INFORMATION MANAGEMENT **246**

Importance of data and information to organisations	247
Goals and objectives of organisations and information systems	249
Information management strategies	252
Key legislation for storage and disposal of data and information	256
Ethics and information systems	265
Resolving legal and ethical tensions	269

CHAPTER 6 SECURITY MEASURES **274**

Preparing for the School-assessed Coursework (SAC)	275
Threats to data and information	275
Disaster recovery plans	280
Consequences of security failure	289
Components of data and information management	291
Networks and cloud computing	296
Evaluating information management strategies	299
Next steps	300

Index	305

PREFACE

This fourth edition of *Informatics VCE Units 3 & 4* incorporates the changes to the VCAA VCE Computing Study Design that took effect from 2016.

This textbook looks at how individuals and organisations use, and can be affected by, information systems in their daily lives.

We believe that teachers and students require a text that focuses on the **Areas of Study** specified in the **Study Design**, and that presents information in a sequence that allows easy transition from theory into practical assessment tasks. We have therefore written this textbook so that a class can begin at Chapter 1 and work their way systematically through to the end. Students will encounter material relating to the **key knowledge** dot points for each **Outcome** before they reach the special section that describes the Outcome. The Study Design outlines key skills that indicate how the knowledge can be applied to produce a solution to an information problem. These Outcome preparation sections occur regularly throughout the textbook, and flag an appropriate point in the student's development for each Outcome to be completed. The authors have covered all key knowledge dot points for the Outcomes from the Informatics VCE Units 3 & 4 course.

Our approach has been to focus on the key knowledge required for each school-assessed Outcome, and to ensure that students are well prepared for these; however, there is considerable duplication in the Study Design relating to the knowledge required for many of the Outcomes. We have found that, with an Outcomes approach, we are covering material sometimes several times. For example, knowledge of a problem-solving methodology is listed as key knowledge for five different Outcomes. In these cases, we have tried to provide a general coverage in the first instance, and specifically apply the concept to a situation relevant to the related Outcome on subsequent encounters.

The authors assume teachers will develop the required key skills with their students within the context of the key knowledge addressed in this textbook and the resources available to them.

We have incorporated a margin column in the text that provides additional information and reinforcement of key concepts. The margin column also includes activities related to the topics covered in the text, and a consideration of issues relevant to the use of information systems.

Outcome features are included at several points in the book, indicating the nature of the tasks that students are to undertake in the completion of the school-assessed Outcome. The steps required to complete the Outcome are listed, together with advice and suggestions for approaching the task. The output and support material needed for submission are described. Sample tasks and further advice relating to the Outcomes are available at <http://informatics3and4.nelsonnet.com.au>.

The chapters are organised to present the optimum amount of information in the most effective manner. The text is presented in concise, clearly identified sections to guide students through the text. Each chapter is organised into the sections described on page vii.

ABOUT THE AUTHORS

James Lawson is Head of Computing at Trinity Grammar School, and has taught VCE Computing since 1995. He has been an exam assessor for Computing, and has on a number of occasions presented at both VITTA and the History Teachers' Association of Victoria (HTAV).

Dr Therese Keane is a senior lecturer at Swinburne University. Therese has worked in a variety of school settings where she taught IT and was the Director of ICT. She holds a Doctorate in Education focusing on ICT leadership in schools. Therese is a member of the ACS ICT Educators Board, a Committee Member on the DLTV and a former office holder at both VITTA and the ICTEV. She has presented numerous seminars and workshops for teachers involved in teaching IT. Therese has written several textbooks in all units of VCE Information Technology since 1996. Therese's current work involves providing professional development to Computing teachers, delivering workshops and presentations to secondary students and researching the use of technology and computers in schools for teaching and learning purposes.

Mark Kelly learned to program in FORTRAN in 1975, bought his first computer – a Tandy TRS-80 with 4KB of RAM – in 1978 and has been programming and researching computers ever since. He taught VCE IT for 20 years after 10 years teaching English and Psychology. At McKinnon Secondary College he was Systems Manager and author of 'Rupert', the college's student reporting database.

HOW TO USE THIS BOOK

Key knowledge

The key knowledge that is covered in each chapter is listed on the first page. The list includes key knowledge specified in the Outcome related to the chapter.

For the student

The first page of each chapter includes an overview of the chapter's contents so that students are aware of the material they will encounter.

For the teacher

This section outlines how the chapter fits into the overall study of Informatics, and indicates how the material relates to the completion of Outcomes.

Chapters

The major learning material in the chapter is presented as text, photographs and illustrations. The text describes in detail the theory associated with the stated Outcomes of the Informatics VCE Units 3 & 4 curriculum in language that is clear and appropriate for students at this level. The photographs show actual hardware, software and other objects described in the text. Illustrations are used to demonstrate concepts that are more easily explained in this manner.

Throughout the chapter, glossary terms are highlighted in bold, red text. They are defined at the end of the chapter, in **Essential terms**.

Margin column

The margin column contains further explanations that support the main text, weblink icons, additional material outside the Study Design and cross-references to material covered elsewhere in the textbook. Issues relevant to Informatics that will promote classroom discussion are also included in the form of 'Think about Informatics' boxes.

THINK ABOUT INFORMATICS 1.1

Choose a UFD style you prefer and create one that shows how the Rebel online ordering system might work from the point of view of a user.

Chapter summary

The chapter summary at the end of each chapter is divided into two main parts to help you review each chapter.

Essential terms are the glossary terms that have been highlighted throughout the chapter.

Important facts are a list of summaries, ideas, processes and statements relevant to the chapter, in the order in which they occur in the chapter.

Test your knowledge

Short-answer questions are provided to help students when reviewing the chapter material. The questions are grouped and identified with a section of the text to allow the teacher to direct appropriate questions based on material covered in class. Teachers will be able to access answers to these questions at <http://informatics3and4.nelsonnet.com.au>.

Apply your knowledge

Each chapter concludes with a set of questions requiring students to demonstrate that they can apply the theory from the chapter to more complex questions. The style of questions reflects what students can expect in the end-of-year examination. Teachers will be able to access suggested responses to these applications on the teacher website.

Preparing for the Outcomes

This section appears at points in the course where it is appropriate for students to complete an Outcome task. The information provided describes what the students need to do in the Outcome, the suggested steps to be followed in the completion of the task and the material that needs to be submitted for assessment.

NelsonNet

The NelsonNet student website contains:

- multiple-choice quizzes for each chapter, mirroring the VCAA Unit 3 & 4 exam.

The NelsonNet teacher website also contains:

- answers for the **Test your knowledge** and **Apply your knowledge** questions in the book
- sample SACs
- chapter tests
- practice exams for each unit.

Please note that complimentary access to NelsonNet and the NelsonNetBook is only available to teachers who use the accompanying student textbook as a core educational resource in their classroom. Contact your sales representative for information about access codes and conditions.

An open-access weblink page is also provided for all weblinks that appear in the margins throughout the textbook. This is accessible at <http://informatics3and4.nelsonnet.com.au>.

OUTCOMES

OUTCOME	KEY KNOWLEDGE	LOCATION
Unit 3 Area of Study 1 Outcome 1	Organisations and data management On completion of this unit the student should be able to design a solution, develop it using a relational database management system, and diagrammatically represent how users interact with an online solution when supplying data for a transaction.	Chapter 1
Data and information	techniques used by organisations to acquire data through their interactive online solutions and reasons for their choice	pp. 3–4, 9–10
	techniques for efficient and effective data collection	pp. 13–15
	characteristics of data types	pp. 22–3
Digital systems	physical and software security controls used by organisations to protect their data	pp. 57–9
Approaches to problem solving	purposes and structure of an RDBMS, including comparison with flat file databases	pp. 23–7
	naming conventions to support efficient use and maintenance of an RDBMS	p. 36
	a methodology for creating an RDBMS structure: identifying entities, defining tables and fields to represent entities; defining relationships by identifying primary key and foreign key fields; defining data types and field sizes; normalisation to third level	pp. 26–36
	design tools for describing data types and the value of entity–relationship (ER) diagrams for representing the structure of an RDBMS	pp. 26–33
	design principles that influence the functionality and appearance of solutions	pp. 15–19
	design tools for representing solutions	pp. 36–40
	functions and techniques within an RDBMS to efficiently and effectively validate and manipulate data	pp. 41–3
	functions and techniques to retrieve required information through searching, sorting, filtering and querying data sets	pp. 44–8
	methods and techniques for testing that solutions perform as intended	pp. 49–57
Interactions and impact	reasons why organisations acquire data using online facilities, including 24-hour customer access, improved efficiencies through direct data entry by customers, improvements in effectiveness, and access to global markets, marketing opportunities and ongoing services	pp. 3–5
	reasons why users supply data for online transactions, including convenience, variety of choice, reducing costs	pp. 5–9
	techniques used by organisations to protect the rights of individuals and organisations who supply data, including security protocols and stating privacy, shipping and returns policies	pp. 57–9
	user flow diagrams that depict different ways in which users interact with online solutions	pp. 11–13
Key skills	select and apply design tools and techniques for describing data types and representing the structure and functionality of solutions	pp. 36–40
	use RDBMS functions and techniques to construct a relational database to manipulate and validate data	pp. 41–3
	apply functions and techniques to construct queries that efficiently retrieve required information	pp. 44–8
	select and apply testing methods and techniques to confirm whether the solutions operate as intended	pp. 49–57
	use software tools to represent the user interface of the page on which online transactions begin, including data protection, where appropriate	pp. 10, 18, 70
	use software tools to represent the interactions between users and online solutions (user flow diagrams)	pp. 11–13
	annotate user flow diagrams to identify where and why data protection is used.	pp. 11–13

OUTCOME	KEY KNOWLEDGE	LOCATION
Unit 3 Area of Study 2 Outcome 2	Data analytics: drawing conclusions On completion of this unit the student should be able to use a range of appropriate techniques and processes to acquire, prepare, manipulate and interpret complex data to confirm or refute a hypothesis, and formulate a project plan to manage progress.	Chapter 2, Chapter 3
Data and information	primary and secondary data sources (digital and non digital) and methods of data acquisition, including observation, interview and querying of resources	pp. 82–86, 91–107
	suitability of quantitative and qualitative data for manipulation including comparisons (quantitative) and policy formation (qualitative)	pp. 85–6
	data types and data structures relevant to selected software tools	pp. 121–3
	one of the following methods for referencing primary and secondary sources: Harvard, American Psychological Association (latest edition), Chicago, Institute of Electrical and Electronics Engineers (IEEE)	pp. 107–11
	criteria to check the integrity of data including timeliness, authenticity, relevance, accuracy	pp. 111–21
	techniques for coding qualitative data to support manipulation	pp. 86–91
Interactions and impact	key legal requirements for storage and communication of data and information, including privacy, intellectual property and human rights requirements	pp. 129–34
Approaches to problem solving	features of a reasonable hypothesis including a specific statement identifying a prediction and the variables	pp. 73–8
	solution specifications: requirements, including data to support the prediction of the hypothesis, constraints and scope	pp. 78–82
	project management concepts and processes, including milestones and dependencies (concepts), task identification, sequencing, time allocation, resources and documentation using Gantt charts (processes)	pp. 134–44
	file naming conventions to support efficient use of software tools	pp. 144–6
	software functions to organise, manipulate and store data	pp. 146–52
	techniques for identifying patterns and relationships between data	pp. 152–65
Digital systems	roles, functions and characteristics of digital system components used to input, store, communicate and output data and information	pp. 165–70
	physical and software security controls suitable for protecting stored and communicated data	pp. 170–5
Key skills	frame a hypothesis that can be tested	pp. 84, 127, 156
	determine the specifications of the solution	pp. 82, 127
	acquire complex data sets and use a standard referencing system to acknowledge intellectual property	pp. 127, 179
	apply techniques that discriminate data on the basis of its integrity	pp. 127, 148, 179
	select and apply methods to secure stored and communicated data and information	pp. 127, 179
	organise, manipulate and interpret selected data, identifying relationships and patterns to develop a conclusion	pp. 164, 179
	devise and apply a file management plan	p. 179
	prepare project plans using software	p. 179
	select and use digital system components appropriate to project needs	p. 179

OUTCOME	KEY KNOWLEDGE	LOCATION
Unit 4 Area of Study 1 Outcome 1	Data analytics: presenting the findings On completion of this unit the student should be able to design, develop and evaluate a multimodal online solution that confirms or refutes a hypothesis, and assess the effectiveness of the project plan in managing progress.	Chapter 4
Approaches to problem solving	characteristics of information for educating world-wide audiences, including gender and culture inclusiveness, commonality of language, age appropriateness	pp. 186–92
	techniques for generating design ideas	pp. 209–18
	criteria for evaluating alternative design ideas and the effectiveness of solutions	pp. 218–20, 237–40
	characteristics of effective multimodal online solutions	pp. 186–199
	formats and conventions appropriate to multimodal online solutions	pp. 202–4
	design principles that influence the functionality and appearance of multimodal online solutions	pp. 204–9
	design tools for representing a solution's appearance and functionality, including relationships, where appropriate	pp. 220–3
	functions, techniques and procedures for efficiently and effectively manipulating data using software tools	pp. 199–202, 231–2
	manual and electronic validation techniques	pp. 231–2
	functions, techniques and procedures for managing files	pp. 227–30
	techniques for testing that solutions do what is intended	pp. 232–7
	techniques for documenting the progress of projects, including annotations, logs and adjustments to tasks and timeframes	pp. 223–7
	strategies for evaluating the effectiveness of solutions and assessing project plans	pp. 239–40
Key skills	generate alternative design ideas	p. 209
	select preferred design ideas, based on student-developed criteria	p. 209
	select and apply design tools to represent the functionality and appearance of solutions	p. 222
	select and apply software functions, methods, formats, conventions, techniques and design principles to develop multimodal online solutions that operate as intended	p. 209
	monitor and adjust project plans where appropriate	p. 225
	apply criteria to evaluate the effectiveness of multimodal online solutions in communicating conclusions to hypotheses	p. 243
	assess the effectiveness of project plans in managing work practices	p. 243
Unit 4 Area of Study 2 Outcome 2	Information management On completion of this unit the student should be able to compare and contrast the effectiveness of information management strategies used by two organisations to manage the storage and disposal of data and information, and recommend improvements to their current practices.	Chapter 5, Chapter 6
Interactions and impact	reasons why data and information are important to organisations, including meeting the goals and objectives of both organisations and information systems	pp. 247–52
	reasons why information management strategies are important to organisations, including maximising opportunities, minimising risks and fulfilling legal requirements	pp. 252–6
	key legislation that affects how organisations control the storage and disposal of their data and information: the <i>Privacy Act 1988</i>, the <i>Privacy and Data Protection Act 2014</i>, and the <i>Health Records Act 2001</i>	pp. 256–64
	ethical dilemmas arising from information management practices	pp. 265–9
	strategies for resolving legal and ethical tensions between stakeholders arising from information management practices	pp. 269–70

OUTCOME	KEY KNOWLEDGE	LOCATION
	reasons for preparing disaster recovery plans, and their scope, including evacuation, backing up, restoration and test plans	pp. 280–9
	possible consequences for organisations that fail to follow or violate security measures	pp. 289–90
	criteria for evaluating the effectiveness of information management strategies	pp. 299–300
Digital systems	role of people, processes and digital systems in the management of data and information	pp. 291–6
	types and causes of accidental, deliberate and events-based threats to the integrity and security of data and information	pp. 275–80
	physical and software security controls for preventing unauthorised access to data and information and for minimising the loss of data accessed by authorised and unauthorised users	pp. 275–90
	the advantages and disadvantages of using networks and cloud computing for storing and disposing of data and information.	pp. 296–8
Key skills	explain the current information management strategies used by organisations to monitor and control their data and information	p. 273
	identify similarities and differences between the information management strategies of organisations	p. 273
	propose and apply criteria to evaluate the effectiveness of information management strategies	pp. 273, 304
	discuss possible consequences of ineffective information management strategies	pp. 273, 304
	recommend information management strategies to improve current practices.	pp. 273, 304

PROBLEM-SOLVING METHODOLOGY

When an information problem exists, a structured problem-solving methodology is followed to ensure that the most appropriate solution is found and implemented. For the purpose of this course, the problem-solving methodology has four key stages: analysis, design, development and evaluation. Each of these stages can be further broken down into a common set of activities. Each unit may require you to examine a different set of problem-solving stages. It is critical for you to understand the problem-solving methodology because it underpins the entire VCE Computing course.

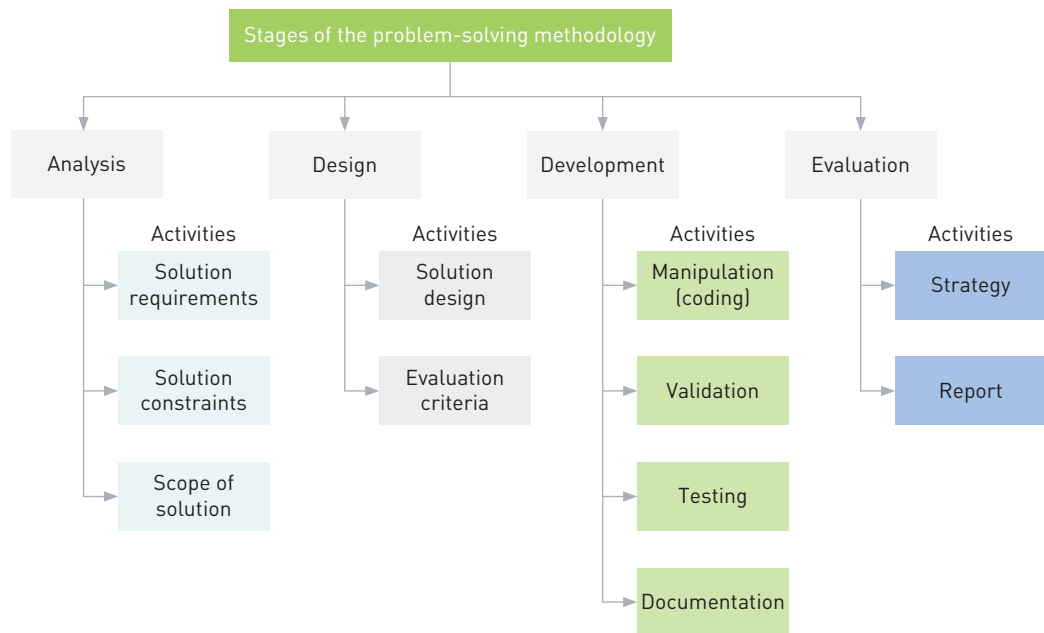


FIGURE 1 The four stages of the problem-solving methodology and their key activities

Analyse the problem

The purpose of analysis is to establish the root cause of the problem, the specific information needs of the organisation involved, limitations on the problem and exactly what a possible solution would be expected to do (the scope). The three key activities are:

- 1** identifying solution requirements – features and functionality that the solution needs to include, information it must produce and data needed to produce this information
- 2** establishing solution constraints – the limitations on solution development that need to be considered. Constraints are classified as economic, technical, social, legal and related to useability.
- 3** defining the scope of the solution – what the solution will and will not be able to do, as well as how the user will benefit.

Design the solution

During the design stage, several alternative design ideas are planned and the most appropriate of these is chosen. Criteria are also created to select the most appropriate ideas and to evaluate the solution's success once it has been implemented. The two key design activities are:

- 1** creating the solution design – it must clearly show a developer what the solution should look like the specific data required and how its data elements should be structured, validated and manipulated. Tools typically used to represent data elements could include data dictionaries, data structure diagrams, input–process–output (IPO) charts, flowcharts, pseudocode (or structured English) and object descriptions. The following tools are also used to show the

relationship between various components of the solution: storyboards, site maps, entity–relationship diagrams, data flow diagrams, structure charts, hierarchy charts and context diagrams. Furthermore, the appearance of the solution needs to be planned so that overall layout, fonts and their colours, for example, can be represented. Layout diagrams and annotated diagrams (or mock-ups) usually fulfil this requirement. A combination of tools from each of these categories will be selected to represent the overall solution design.

- 2 specifying evaluation criteria – during the evaluation stage, the solution is assessed to establish how well it has met its intended objectives. The criteria for evaluation must be created during the design stage so that all personnel involved in the task are aware of the level of performance that ultimately will determine the success or otherwise of the solution. The criteria are based on the solution requirements identified in the analysis stage.

Develop the solution

The solution is created by the developers during this stage from the designs supplied to them. The ‘coding’ takes place, but also checking of input data (validation), testing that the solution works and the creation of user documentation. The four activities involved with development are:

- 1 manipulating or coding the solution – the designs are used to build the electronic solution. The coding will occur here and internal documentation will be included where necessary
- 2 checking the accuracy of input data by way of validation – manual and electronic methods are used; for example, proofreading is a manual validation technique. Electronic validation involves using the solution itself to ensure that data is reasonable. Electronic validation, along with any other formulas, always needs to be tested to ensure that it works properly
- 3 ensuring that a solution works through testing – each formula and function, not to mention validation and even the layout of elements on the screen, need to be tested. Standard testing procedures involve stating what tests will be conducted, identifying test data, stating the expected result, running the tests, stating the actual result and correcting any errors
- 4 documentation allowing users to interact with (or use) the solution – while it can be printed, in many cases it is now designed to be viewed on screen. User documentation normally outlines procedures for operating the solution, as well as generates output (like reports) and basic troubleshooting.

Evaluate the solution

Sometime after a solution has been in use by the end user or client, it needs to be assessed or evaluated to ensure that it has been successful and does actually meet the user’s requirements. The two activities involved in evaluating a solution are:

- 1 working out an evaluation strategy – creating a timeline for when various elements of the evaluation will occur and how and what data will be collected (because it must match the criteria created in the design stage)
- 2 reporting on the success of the solution – providing feedback to the user about how well the solution meets their requirements. This is based on the findings of the data gathered at the beginning of the evaluation stage when compared with the evaluation criteria created during the design stage.

KEY CONCEPTS

Within each VCE Informatics subject are four key concepts whose purpose is to organise course content into themes. These themes are intended to make it easier to teach and make connections between related concepts and to think about information problems. Key knowledge for each Area of Study is categorised into these key concepts, but not all concepts are covered by each Area of Study. The four key concepts are:

- 1 data and information
- 2 digital systems
- 3 approaches to problem solving
- 4 interactions and impact.

Data and information focuses on the acquisition, structure, representation and interpretation of data and information in order to elicit meaning or make deductions. This process needs to be completed in order to create solutions.

Digital systems focus on how hardware and software operate in a technical sense. This also includes networks, applications, the internet and communication protocols. Information systems have digital systems as one of their parts. The other components of an information system are people, data and processes.

Approaches to problem solving focuses on thinking about problems and ways of creating solutions. Computational, design and systems thinking are the three key problem-solving approaches.

Interactions and impact focuses on relationships that exist between different information systems and how these relationships affect the achievement of economic and social goals. Three types of relationships are considered:

- 1 people interacting with other people when collaborating or communicating with digital systems
- 2 how people interact with digital systems
- 3 how information systems interact with other information systems.

This theme also looks at the impact of these relationships on information needs, privacy and personal safety.

INTRODUCTION

Unit 3 of Informatics focuses on the acquisition, management, manipulation and interpretation of data to meet a range of needs.

Students consider why organisations acquire online transaction data from websites and how the data gathering process is structured. They consider the legal requirements of protecting the rights of data providers and why organisations require data to be organised in particular ways. Skills and knowledge developed by students in using a relational database management system (RDBMS), coupled with understanding the process of acquisition of transaction data from websites, are used to produce a solution to an information problem. This requires data to be manipulated and validated, and information to be presented in a suitable format.

Students formulate a hypothesis, then acquire and organise multiple data sets in order to confirm or refute their hypothesis. This provides students with an understanding of how complex data sets can be used for decision making. The data gathered by students will be manipulated and analysed with either a database or spreadsheet application. The project is planned and monitored using appropriate project management skills and tools. The second half of this project is covered in Unit 4.

There are two Areas of Study and Outcomes to be completed in Unit 3.

AREA OF STUDY 1: ORGANISATIONS AND DATA MANAGEMENT

Outcome 1

On completion of this unit you should be able to design a solution, develop it using a relational database management system and diagrammatically represent how users interact with an online solution when supplying data for a transaction.

The focus of this Outcome is on using a RDBMS to produce a solution to an information problem. You will consider how and why data is acquired from websites and use a sample set of data to design and develop a database. You will use appropriate functions and techniques to retrieve required information through searching, sorting, filtering and querying data sets. You will also represent, using diagrams, the user interface of a data acquisition page and create a user flow diagram to trace the interaction of a user with an online transaction.

AREA OF STUDY 2: DATA ANALYTICS: DRAWING CONCLUSIONS

Outcome 2

On completion of this unit you should be able to use a range of appropriate techniques and processes to acquire, prepare, manipulate and interpret complex data to confirm or refute a hypothesis and formulate a project plan to manage progress.

You will formulate a hypothesis, choosing from a wide range of categories, then use data acquisition and manipulation skills to investigate your claim. You must check the integrity of your data and format it correctly. The data from this Outcome is used to create a multimodal solution in Unit 4, Outcome 1. Therefore, when you create a detailed project plan in this Outcome, it must span both Outcomes.

UNIT

3

CHAPTER

1

ORGANISATIONS AND DATA MANAGEMENT

Key knowledge

After completing this chapter, you will be able to demonstrate knowledge of:

Data and information

- techniques used by organisations to acquire data through their interactive online solutions and reasons for their choice
- techniques for efficient and effective data collection
- characteristics of data types

Digital systems

- physical and software security controls used by organisations to protect their data

Approaches to problem solving

- purposes and structure of an RDBMS, including comparison with flat file databases
- naming conventions to support efficient use and maintenance of an RDBMS
- a methodology for creating an RDBMS structure: identifying entities, defining tables and fields to represent entities; defining relationships by identifying primary key and foreign key fields; defining data types and field sizes; normalisation to third level
- design tools for representing the structure of an RDBMS
- design principles that influence the functionality and appearance of solutions
- design tools for representing solutions
- functions and techniques within an RDBMS to efficiently and effectively validate and manipulate data
- functions and techniques to retrieve required information through searching, sorting, filtering and querying data sets
- methods and techniques for testing that solutions perform as intended

Interactions and impact

- reasons why organisations acquire data using online facilities, including 24-hour customer access, improved efficiencies through direct data entry by customers, improvements in effectiveness and access to global markets, marketing opportunities and ongoing services
- reasons why users supply data for online transactions
- techniques used by organisations to protect the rights of individuals and organisations who supply data, including security protocols and stating privacy, shipping and returns policies
- user flow diagrams that depict different ways in which users interact with online solutions

For the student

Many of our financial and social transactions are now carried out online, in an electronic environment, rather than face-to-face. Organisations increasingly seek to gather our personal details via electronic forms and use that data to improve or provide services. Because of the extent to which databases are now used, it is important that you have an understanding of how relational database management systems work, and how to plan and develop their key features.

For the teacher

The focus of this chapter is on how organisations gather data in databases, why they do it and how the data is safeguarded. Students are also introduced to a relational database management system and its key features. This chapter forms the basis of the background needed to prepare students for Unit 3, Outcome 1.

Collecting data online

It is now commonplace for people to interact with online services that require them to enter and share data. Many organisations now rely on their customers to enter and provide their details directly to the organisation's **database** via websites. By removing the extra steps of filling out a paper-based form, returning it to the organisation and entering it into the database, organisations have eliminated some degree of paperwork. In addition, users neither have to wait for forms to be mailed out, nor do they have to go to a specific place to collect them during business hours. Instead, users can access an **online form** at any time of the day and from any location with an internet connection (Figure 1.1).

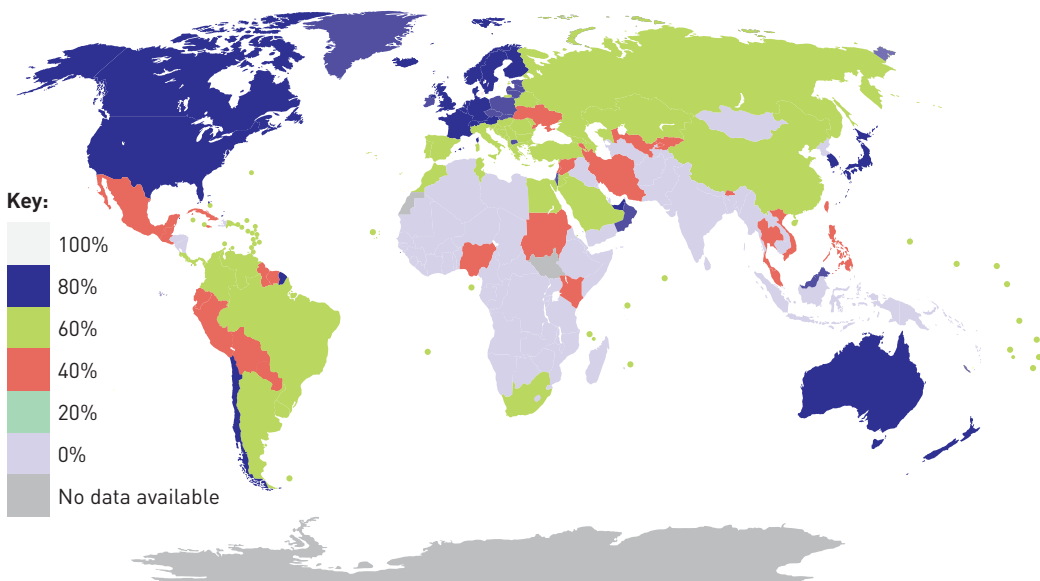


FIGURE 1.1 The International Telecommunications Union (ITU) regularly produces statistics on internet and telecommunications usage throughout the world. This map of the world has been drawn to interpret internet penetration as a percentage of the population of each country in 2012. While developed countries are well-served – with Iceland, Bermuda, Norway, Sweden and Denmark leading the way – much of Africa and some parts of Asia lag behind.

Having removed some of the time and physical barriers that hardcopy forms present, a clear reason for organisations to provide online data facilities is that customers can access websites when it is convenient for them – at any time in any 24-hour period. **Efficiency** improves when customers enter their own data. For example, time and effort are saved because the double handling of forms is no longer required, so money can be saved because extra staff members are not needed for data entry.

As shown in Figure 1.2 on page 4, the input boxes for a typical online form all have labels inside them to guide the user as to the type and format of data to be entered. The travel option icons like 'Flights' and 'Hotels' act in the same manner as radio buttons in that, depending on which one you select, all the data entry fields change. The tick box beside the 'Add a Hotel' option activates the hotel booking boxes when selected. If the 'Hotels' tick box is unticked, those options are hidden.

Effectiveness also improves when customers input their own data. Online forms offer greater accuracy by overcoming transcription issues associated with messy handwriting that is difficult to decipher and thus transcribe. The customer types into the online form directly, so the customer ensures they supply their own correct details.

The screenshot shows the Expedia.com.au homepage. At the top, there's a navigation bar with links like Home, Hotels, Packages, Flights, Cruises, Car Hire, Travel Insurance, Activities, Short Breaks, Last Minute, Deals, Blog, and Rewards. Below this is a search bar with tabs for Flights, Hotels, Flight + Hotel, Cars, Activities, and Cruises. The main form is for flight booking, with fields for 'Flying from' and 'Flying to' (both with location icons), 'Departing' and 'Returning' dates (with date pickers), and a checkbox for 'Add a hotel'. Below the hotel checkbox are fields for 'Check in' and 'Check out' dates, and dropdown menus for 'Rooms', 'Adults (18+)', and 'Children (0-17)'. A yellow 'Search' button is at the bottom left of the form. A promotional banner at the bottom says 'Save up to \$200 on Expedia packages Book Flight + Hotel at the same time*'. On the right side of the image, there is vertical text: 'Screenshot used by permission from Expedia, Inc.'

FIGURE 1.2 The homepage for Expedia, a site for booking travel and accommodation online

Many required fields in online forms are indicated by an asterisk – especially a red asterisk positioned to the left or right of the field.

Many online forms have required boxes that must be completed before the individual can progress further (Figure 1.3). These are generally indicated by an asterisk. Inbuilt mechanisms like these ensure that organisations collect complete data from the user. A potential problem with this direct access is that a customer can input incorrect data (such as a false email address or a nonsensical name), which will not necessarily be picked up by electronic **validation**.

Purchasing goods from a global market opens up economies that were once closed to us unless we travelled and purchased the goods and services overseas, thus creating the

global economy. Online data acquisition can also work for businesses because it can provide them with marketing opportunities with an expanded customer base. Their potential customers can come from anywhere in the world, rather than being limited to their local community.

Online options from businesses do not stop merely at selling goods. A wide range of **ongoing services** are now provided by organisations after a purchase. A common example is the tracking information that allows you to view the

The screenshot shows a form titled 'Download our playbook on Presentation Storyboarding'. The text below the title says: 'It's no secret that a solid presentation story is key to convincing your audience. In this playbook, you'll find our foolproof presentation outline and a storyboarding template to help you plan your next presentation. Fill up the form to download this playbook now!'. The form has a red header with the text 'YOUR GUIDE TO PRESENTATION STORYBOARDING' and 'A SlideCentral Playbook'. Below the header is a 'slide & meet' logo. To the right of the header are input fields for 'First Name', 'Last Name', 'Email', 'Company Name', and 'Country'. At the bottom right is a red 'DOWNLOAD NOW' button. Below the button is a small disclaimer: 'We at SlideCentral Inc. don't warrant we won't be using your email to anyone other than we don't guarantee that we won't send you massive presentation tips though.'

FIGURE 1.3 A simple form used to allow users to download a file about storyboarding presentations. If users wish to download the information then they must provide the requested details.

processing of your order and where it is in transit right up until the moment it arrives at your door (Figure 1.4). Ongoing services may also include the electronic interactions concerning a returned item (online forms and emails) or online booking systems where a range of organisations will be alerted to and respond to your bookings (such as booking a flight and hotel). It could even be a confirmation reminder via email or text message of an upcoming booking or appointment.

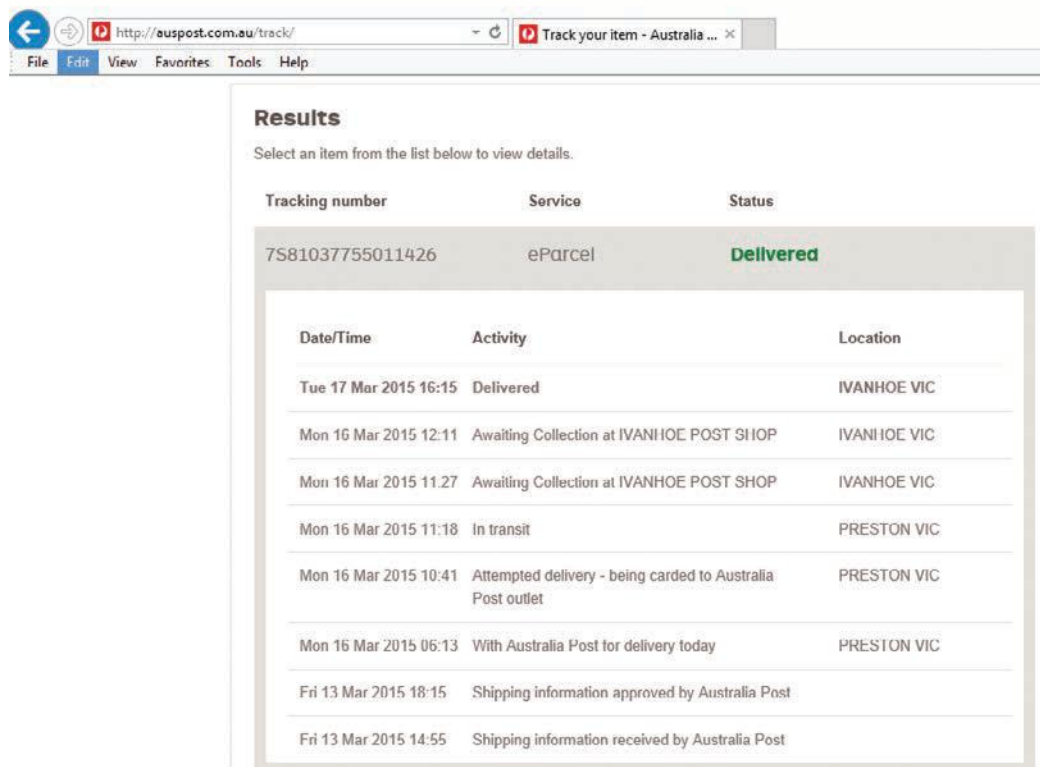


FIGURE 1.4 Most online sellers allow purchased goods to be tracked so a customer can see where their goods are while in transit. This is an ongoing service provided by Australia Post.

Why users supply data for online transactions

There are several key benefits for users supplying data in order to complete online transactions. These benefits include convenience, an increase in variety of choice of goods and services available, and the reduction of costs to the user. Each of these benefits is explained in more detail in the section that follows.

Convenience

Being able to purchase goods and services via websites has changed the habits and practices of many shoppers. While many people enjoy going to a supermarket to browse for products or shopping at the local shopping centre, others prefer the convenience of purchasing goods and services online. It is difficult to categorise the types of people who use these services; however, it is of most benefit to individuals who are housebound, such as the elderly or sick, or those with young children. Increasingly, individuals who are time-poor tend to order their groceries online rather than shop at a supermarket and wait in long queues. Also, banking transactions can be performed and bills paid online by supplying data such as account numbers, bill numbers and credit card or bank details rather than paying in person or by phone (Figure 1.5). All these transactions require the user to supply data to an organisation via websites and online forms, and the data is entered directly into the organisation's database.

Transfers & BPAY

FIGURE 1.5 An example of a funds transfer screen for paying bills (BPAY) or making direct bank deposits. Customers fill out the account details (which can include a credit card account) for the 'from' account (sender) and the 'to' account (recipient). The transaction can be set up as regular using the 'Set up regular payments' option.

In most instances, for extra security, before a bank transfer is made to an external account for the first time, the user must authorise the transaction by entering a code that is sent directly to their mobile phone.

To purchase goods or services, individuals need to supply data to the organisation about their particular request (Figure 1.6). For example, when purchasing clothes online, the following data is first required: size of the item and the colour. Once the data has been entered into the system and the customer has finished placing items into their 'shopping cart', they then proceed to the 'checkout', where further data is required, such as payment details, the individual's name, address and shipping details.

FIGURE 1.6 Shopping at Rebel is very easy, because the transaction is completed online and the data is sent directly to the organisation. Each step in the transaction is clearly numbered so the purchaser can follow their progress, and at any stage the user can return to an earlier step and edit the data they have entered. Figure continues on page 7.

Continue Shopping

1 Review Cart

Edit Details

2 Checkout

Checkout with My Account

Email Address *

Password *

Forgot password?

☐ Remember my details

▶ LOGIN & CHECKOUT

Register an Account

Email Address *

Password *

Confirm Password

☐ Receive special deals, offers and promotional material

▶ REGISTER & CHECKOUT

Checkout as a Guest

You don't have to be registered to shop with Rebel.

You will receive your order summary and confirmation email at the address you provide. If you have registered an account, you will be prompted to login.

▶ CHECKOUT AS GUEST

3 Your Details

4 Payment

3 Your Details

Ensure your details are correct below so we can deliver your products as quickly as possible.

We are unable to ship orders to PO Boxes, Private Parcel Lockers or Defence Force addresses.

Delivery Details

First Name *

Last Name *

Phone *

e.g. 07 3123 4567

Alternate Phone

e.g. 0401 123 456

Address *

Start typing your address and we'll look it up for you to save you time...

Special Delivery Request

None

Please note this is a request only, and we cannot guarantee our couriers will be able to accommodate

Billing Details

☒ Use Delivery Details

Contact Email

You will receive your order summary and confirmation email at this address.

Email Address *

▶ CONFIRM DETAILS

Rebel

FIGURE 1.6 (continued)

FIGURE 1.6 (continued)

4 Payment

Choose your payment method below to complete your purchase.

Only 1 Gift Card can be redeemed per purchase.

Credit Card

VISA MASTERCARD AMERICAN EXPRESS

PayPal

PayPal

Gift Card

rebel

You will be charged the remaining balance:
\$100.00

Credit Card Type *
American Express

Name On Card *
[Text Field]

Card Number *
[Text Field]

Credit Card Expiry Date *
[Month] [Year]

CCV Number *
[Text Field]

Summary

Subtotal	\$100.00
Freight	Free
Total Payable	\$100.00

☐ I agree with the [Terms & Conditions](#) *

SUBMIT PAYMENT

Online Order Enquiry
1300 654 502
8.00am - 5.00pm AEDT

Customer Service
Contact Us
Help Desk

We Accept
VISA MASTERCARD AMERICAN EXPRESS PayPal

Secured By
digicert EV SSL
SECURE
EV SSL Certificates

Copyright © 2015 rebel. All Rights Reserved. Privacy Policy | Terms & Conditions | Site Map | Corporate Site | Careers Web Design Brisbane - Emsu Design

FIGURE 1.6 (continued)

Sometimes alternate versions of goods are sold in different countries because they are customised for the local market, such as clothing sizes and styles or power supplies and communication frequencies for electronic devices. Local laws may also require that goods comply with certain regulations that differ from other parts of the world.

Variety of choice

Users are also keen to supply data in online transactions because it gives them a greater choice of goods and services than they may have available to them locally. For organisations, access to a global market is an important reason for an online presence. The other side of this, of course, is that consumers are the ones creating the global market by their desire and willingness to purchase what they want, regardless of where it originated. For example, Australia often receives fashion clothing a season behind Europe and the United States. The range available is often smaller than what was available to those other countries. But if fashion companies have an online store that ships internationally, Australians can access the same, larger range ahead of the season, rather than only have access to the limited choices available at their local retailers. Similarly, users can compare prices, specifications and quality of goods among several online sellers of technology products that might not be sold in local shops.

Reducing costs

It can be expensive to travel to a specific shop in order to compare products. Fuel and parking costs dissuade consumers from shopping around. If consumers compare products online, it saves money and time. Most online stores deliver, so customers do not need to collect purchases either. In some instances, free delivery is included, although it is important to shop

Price	Description
\$515.95	Apple iPhone 5C 16GB LTE Smart Phone White Shopping Square (NSW) www.shoppingsquare.com.au updated: 09-04-2015
\$519.00	Apple AS NEW iPhone 5 64GB Black BecexTech BecexTech Australia (VIC) www.becextech.com.au updated: 09-04-2015
\$529.00	Apple iPhone 5 64GB Phone UNLOCKED - BLACK Techrific Techrific Australia (VIC) www.techrific.com.au updated: 09-04-2015
\$539.00	iPhone 5C 16GB Blue 16GB, 4G, iOS 7.0.1, 4.0" Touchscreen, 8MP Camera Melbourne Mobiles (VIC) www.melbournemobiles.com.au updated: 08-04-2015
\$539.00	iPhone 5C 16GB Green 16GB, 4G, iOS 7.0.1, 4.0" Touchscreen, 8MP Camera Melbourne Mobiles (VIC) www.melbournemobiles.com.au updated: 08-04-2015
\$539.00	iPhone 5C 16GB Pink 16GB, 4G, iOS 7.0.1, 4.0" Touchscreen, 8MP Camera Melbourne Mobiles (VIC) www.melbournemobiles.com.au updated: 08-04-2015
\$539.00	iPhone 5C 16GB Yellow 16GB, 4G, iOS 7.0.1, 4.0" Touchscreen, 8MP Camera Melbourne Mobiles (VIC) www.melbournemobiles.com.au updated: 08-04-2015
\$539.00	iPhone 5C 16GB White 16GB, 4G, iOS 7.0.1, 4.0" Touchscreen, 8MP Camera Melbourne Mobiles (VIC) www.melbournemobiles.com.au updated: 08-04-2015
\$549.00	Apple iPhone 5 Phone 16GB Unlocked - Refurbished Deals Direct (Online) www.dealsdirect.com.au updated: 09-04-2015
\$559.00	Apple iPhone 5S 16GB Phone Unlocked - Refurbished Deals Direct (Online) www.dealsdirect.com.au updated: 09-04-2015
\$559.00	Apple iPhone 5C 16GB Blue mobileciti MobileCiti (NSW) www.mobileciti.com.au updated: 09-04-2015
\$559.00	Apple iPhone 5 16GB Mobile Phone - White (SPECIAL) Exeltek Solutions (VIC) www.exeltek.com.au updated: 08-04-2015
\$559.99	iPhone 6 Complete LCD with digitizer - Black Macfixit Australia (VIC) www.macfixit.com.au updated: 06-04-2015

FIGURE 1.7 StaticICE compares prices for different products among different retailers, helping consumers to find ways to reduce costs

around because sometimes the price and delivery charge combined may be less than the price of a product with free delivery. Price comparison websites help consumers find the best prices across different retailers without needing to travel (Figure 1.7).

With the free shipping often used by online sellers to attract more customers or generate a greater volume of sales, the seller absorbs the shipping cost as a reduction in their profits. It is important to note that not all businesses can afford to do this.

Techniques used by organisations to acquire data online

Prior to the introduction of web-based database forms, when organisations collected data they tended to do this by asking customers to fill out paper forms (Figure 1.8). The data on these forms would then be typed by a data-entry operator who would transfer the data acquired onto a database. The data-entry operator validated the data to ensure its correctness. Two methods were used: **manual validation**, where the data-entry operator would visually inspect (proofread) the data to check if any mistakes had been made in copying the data; and **machine validation**, in which the database would pick up certain types of errors, such as incorrect numbers entered for a postcode or certain boxes being empty, and then notify the operator of the error. Using a data-entry operator meant that there was scope for mistakes to occur in data entry or input into the database.

Epworth HealthCare

Please tick which Epworth site you are being admitted to:

☐ Richmond Acute ☐ Richmond Rehab
☐ Eastern ☐ Brighton
☐ Clivedon ☐ Camberwell
☐ Freemasons Clarendon St
☐ Freemasons DPC
☐ Freemasons Maternity

Unit Record Number: _____ Adm Number: _____
 Surname: _____
 Given Name: _____
 D.O.B.: _____ Age: _____ Sex: _____
 Medical Officer: _____
 After Patient Identification Label

ADMISSION DETAILS
 (Doctors Secretary to complete - MUST BE COMPLETED)
 Admission Date: _____ Admission Time: _____
 Admitting Dr: _____ Dr Phone: _____
 Procedure: _____ Provisional Item Number(s): _____
 Estimated Length of Stay: _____ days Day Case ☐ Overnight Case ☐

MATERNITY DETAILS
 Estimated Date of Delivery: ____/____/____ Obstetrician: _____

PATIENT DETAILS
 Have you been a patient at Epworth Richmond/Brighton/Eastern/Freemasons/Camberwell/Clivedon? ☐ Yes ☐ No
 Have you stayed in any hospital within the last month? ☐ Yes ☐ No If Yes, Hospital name: _____
 Title: ☐ Mr ☐ Mrs ☐ Miss ☐ Ms ☐ Master ☐ Other: _____
 Surname: _____ Previous Surname: _____
 Given Names: _____ Sex: ☐ Male ☐ Female Date of Birth: ____/____/____
 Country of Birth: _____ Marital Status: _____ Preferred Language: _____
 Residential Address: _____
 Suburb / Town: _____ State: _____ Postcode: _____
 Postal Address: Tick if as per above ☐ Business: _____ Mobile: _____
 Contact No: Home: _____ Aboriginal or Torres Strait Islander: ☐ Yes ☐ No Religion: _____ Tick if No Religion ☐
 Medicare Number: _____ Number beside name on card ☐ Exp Date: ____/____/____
 Pension / Concession No: _____ Exp Date: ____/____/____
 PBS Entitlement Card No: _____ HealthCare Card No: _____

ADMISSION DETAILS

FIGURE 1.8 Part of a hardcopy form used by the Epworth Hospital to collect data about incoming patients

1. Trolley	2. Delivery address	3. Delivery time	4. Payment details	5. Confirmation
4 items \$31.21			Estimated total: \$31.21	

View

	True Foods Traditional Roti Bread 280g	☒	2	\$5.92 \$11.00 per 100 G Remove
Meat, Seafood & Deli (1)				
	Coles RSPCA Approved Chicken Breast Fillets Single Pack approx. 350g Indicated price based on approx. weight. Weight may vary between 300g and 400g.	☒	2	\$9.80 \$14.00 per 100 G Remove
Pantry (2)				
	Passage Foods Passage to Sri Lanka Coconut & Cashew Curry Simmer Sauce 375g	☒	2	\$9.24 \$11.25 per 100 G Remove
	Kraft Smooth Peanut Butter 500g	☒	1	\$6.25 \$11.25 per 100 G Remove
* Trolley sub-total:				\$31.21

FIGURE 1.9 The online checkout at Coles shows the grocery item with the price and quantity next to the selection. The subtotal is dynamic – adjusted as items are added or removed.

Online data acquisition

To make online transactions as easy as possible for users, organisations go to great trouble to design appropriate user experiences (UX) that improve on the manual methods once used, so they use **data acquisition software**. Users come from different backgrounds and experience so the transaction must be designed to guide them through the process as helpfully as possible, confirming what is occurring at each step as well as which steps remain before the transaction is complete (refer back to Figure 1.6). A final confirmation or acknowledgement is also critical so that users know that their data has been sent and that the transaction has been processed and is complete (Figure 1.10).

The role of a user experience or UX designer is to create an online environment that encourages the user to interact with the system with ease by providing an interface that is both usable and accessible.

MS _____ You are logged on to Internet Banking. [LOG OFF](#)

No new messages [Print](#) | [中文](#)

Pay a Bill - Acknowledgement [Help](#)

Transfer	
Debit Account:	094-_____ AUD Day to Day Account
Billers:	000006007 ANZ CARDS 450 _____
Customer Reference Number:	450 _____

Payment	
Amount in AUD:	1,877.23

Payment Date and Frequency	
Payment type:	Pay now
Effective date:	September 9, 2015

Details	
For your account (optional):	ANZ credit card

Transaction reference	
Transaction reference number:	285543X _____
Receipt number:	01538

FIGURE 1.10 This screen confirms to the user that a transaction involving a funds transfer has completed

User flow diagrams

To document how a user will interact with an online solution you can create a **user flow diagram (UFD)**. A UFD depicts the pathways or steps a user will follow to complete a transaction. The UFD may show the 'screens', but not their detail (this is shown in **layout diagrams**) and the different entry points to the online solution. Entry points are not limited to the website or application itself. They can be links from advertisements (such as Google Ads), from search engines, or from any number of social media sites. Typically, the user will find themselves on the relevant page to initiate the intended transaction (for example, purchasing goods or making a booking or reservation of some kind), but they will not have necessarily started at the website homepage.

There are many different ways to draw a UFD. There is no mandated method in the Study Design, although you must use some type of graphics or drawing software. Figures 1.11 to 1.16 demonstrate examples of different UFD styles, but many more exist. Whichever method you use, it is important to show clearly the multiple entry points that could get a user to the beginning of a transaction and the key data flows or interactions they must go through to complete the transaction.

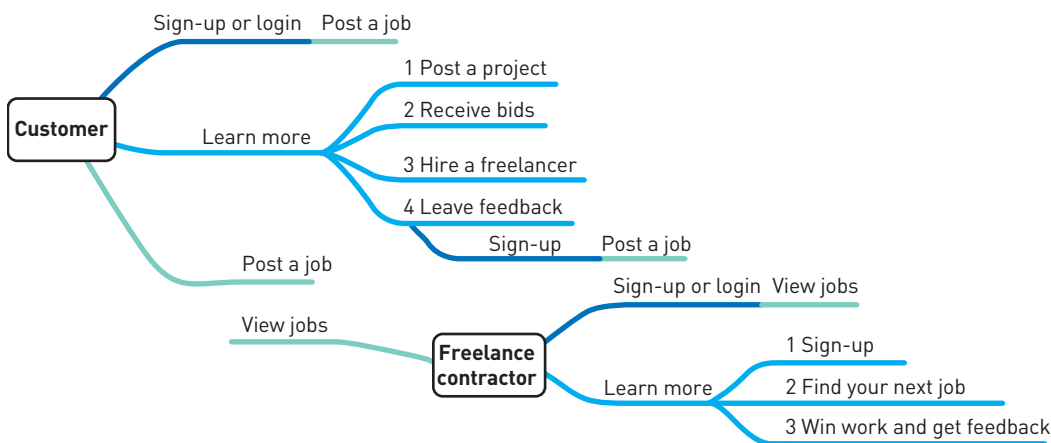


FIGURE 1.11 Job website UFD

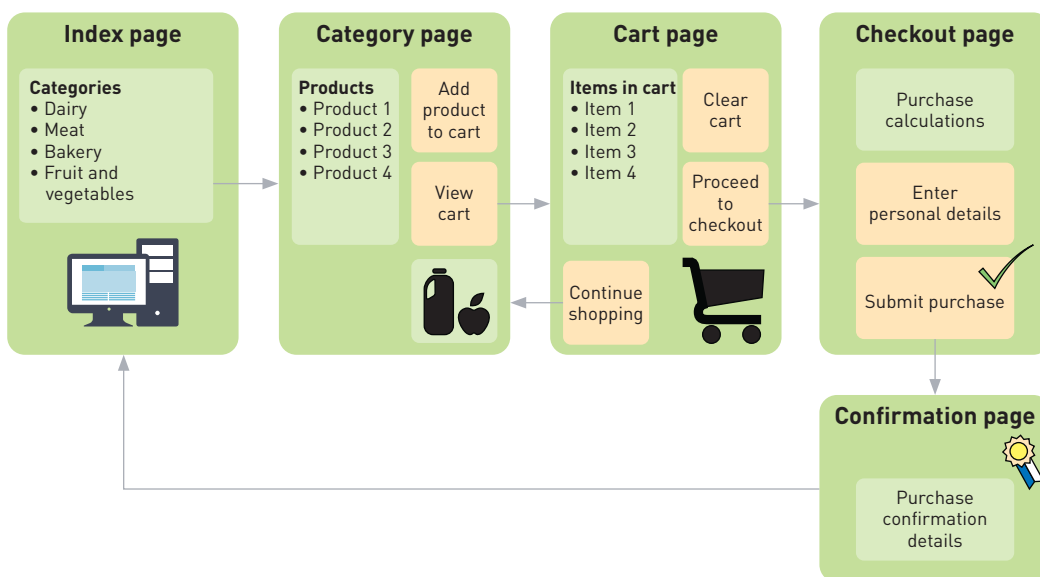


FIGURE 1.12 Purchasing process UFD

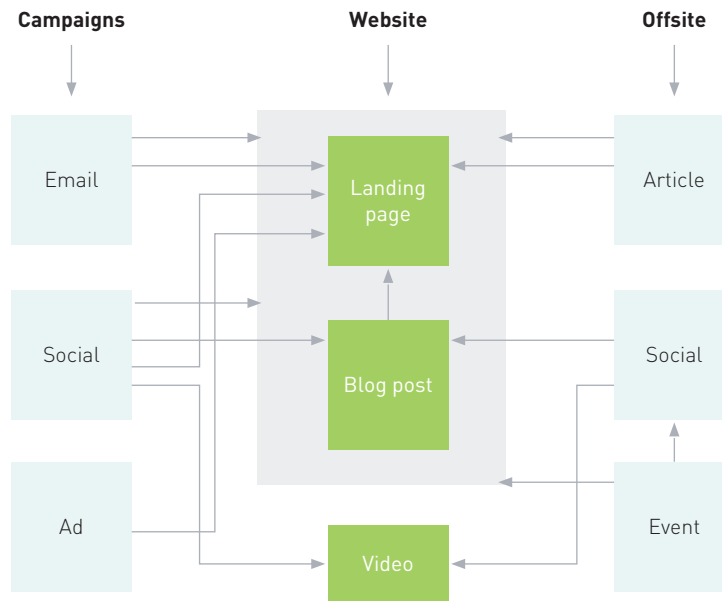


FIGURE 1.13 Social activism website UFD

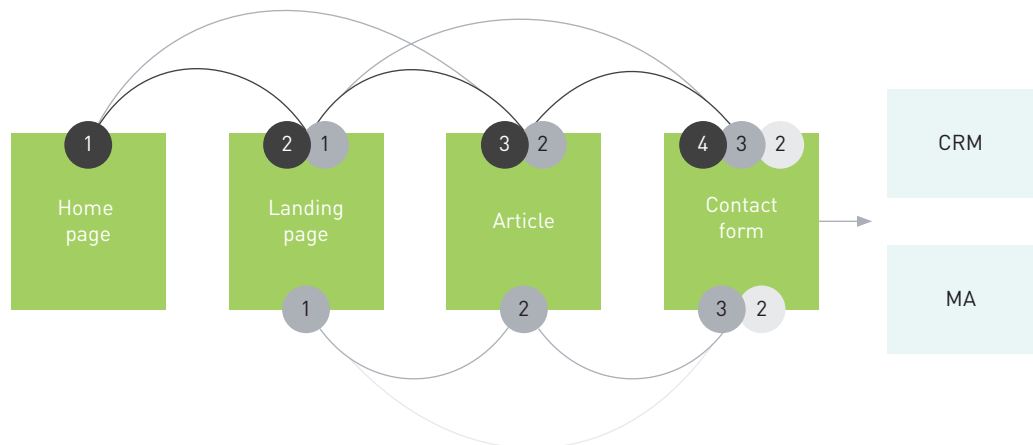


FIGURE 1.14 Customer relations management website UFD

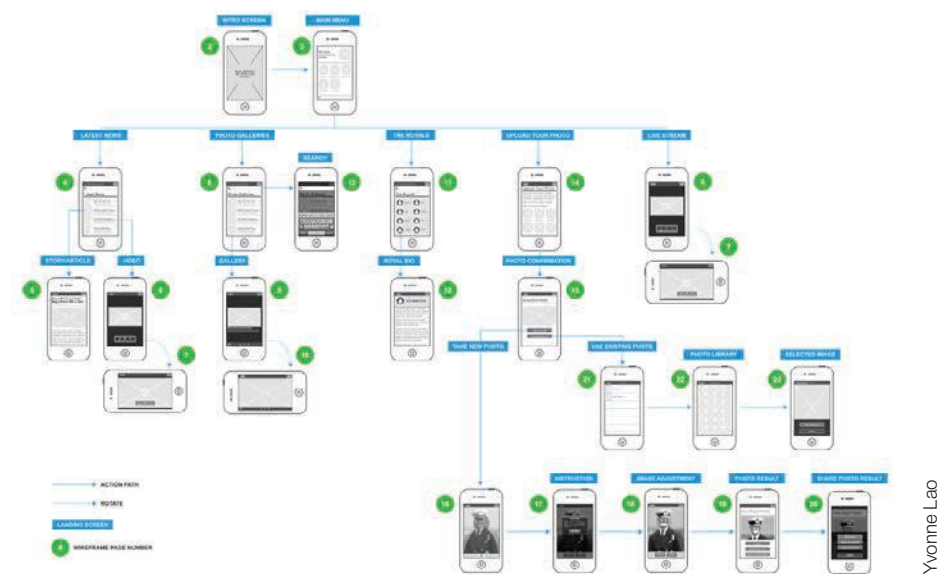


FIGURE 1.15 A UFD that shows an iPhone application flow chart

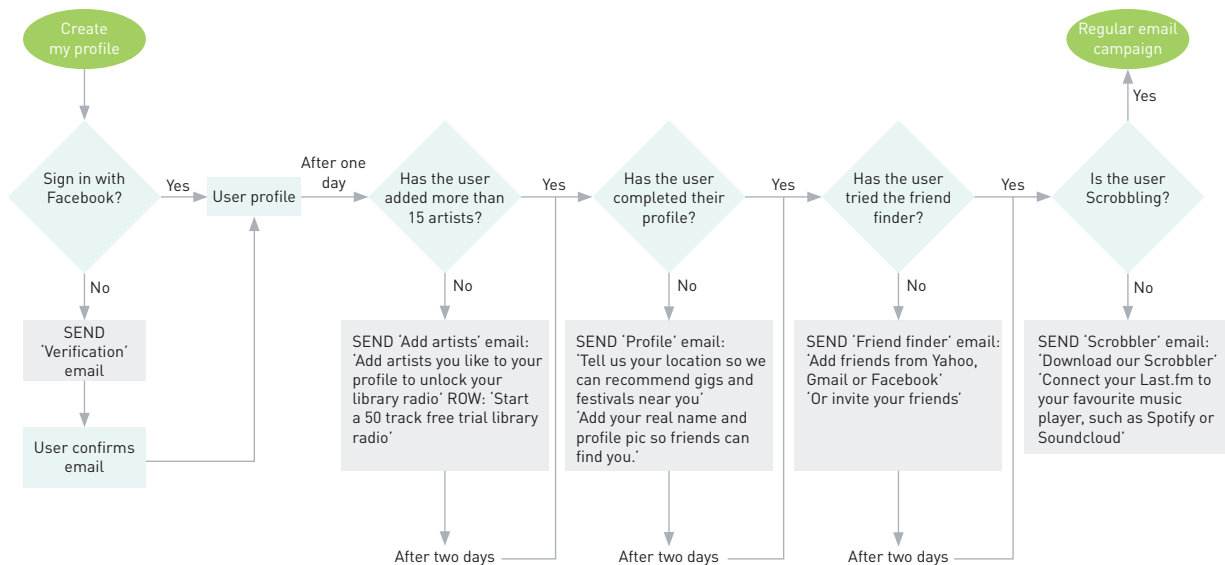


FIGURE 1.16 A UFD showing the Last.fm new starter sign-up email process

Techniques for efficient and effective data collection

To make the transaction as easy as possible, organisations use several different techniques on data input forms. The following are the most common techniques that increase the efficiency and effectiveness of data collection.

- Keep it short: Only include a few questions on each screen and break large sets of questions into separate screens.
- Use clear labels that are easy to understand.
- Validate: Inline existence checks are done by mandatory fields (usually marked with an asterisk), range checking is often done by dropdown lists or radio button groups, data type checks can be done inline on data, such as dates and numbers (accuracy).
- Appropriate field types: For example, use a tick box for True/False answers, or a dropdown list where a set of radio buttons would take up too much space or be unwieldy.
- Text fields: Ensure that any field that a user types into is large enough for all the text to be seen in, to maximise readability.
- Allow the computer to format data, such as phone numbers, to increase readability.
- Ensure error messages are informative: Informative error messages help users to understand what they have done incorrectly and how to fix the error.
- Map clear paths to completion: Set out the form in a logical and linear manner – a single column is preferred.
- Be mobile-friendly: Provide a form that can be accessed from devices other than desktop computers, such as tablets and smartphones.
- Ask relevant questions: What are you asking? Why are you asking it?
- Prepopulate data: Prepopulated data can take details users have already entered and put them into similar fields to save the time and effort of retyping the same details. If the details need to be changed then the user can do so. Home address and shipping address in an online purchase is a good example of this.

THINK ABOUT INFORMATICS 1.1

Choose a UFD style you prefer and create one that shows how the Rebel online ordering system might work from the point of view of a user.

Before any data collection screen is created it needs to be designed, because even a simple feedback form may be expensive to produce. It is important to be able to design a form using a layout diagram and annotate it to show how you have included security and incorporated features that add to the efficiency and effectiveness of the data collection process. Figure 1.17 is an example of the sort of diagram you might produce.

Sign up English (Australia) ▼

First name Last name

Your last name is required.

Yahoo username @yahoo.com.au ▼

Your username is required.

Password Show password

Please create a password.

+61 Mobile number ⓘ

Your mobile number is required.

Birthday Month ▼ Day ▼ Year ▼

Day required.

☐ Male ☒ Female

FIGURE 1.17 Yahoo sign-up screen with validation

rebel

CHOOSE YOUR GIFT CARD NOW

Can't decide on a gift? Why not give a rebel gift card!

rebel gift cards are the ideal solution for those hard to buy for family and friends

Already have a gift card? Check your balance or activate your gift card via the menu options on the left.

Value * 10000

Quantity * 1

The minimum card value is \$10 and the maximum is \$2000.

+ ADD TO CART

FIGURE 1.18 There is a range check on the 'value' field in this order form. There is a clear 'error' message displayed when the value entered is outside the required range.

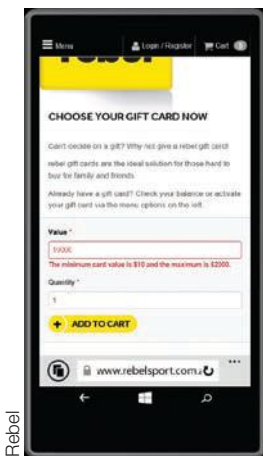


FIGURE 1.19 This is the same screen as in Figure 1.18, but it is being displayed on a mobile device. Notice how the overall layout has changed to suit the screen size.

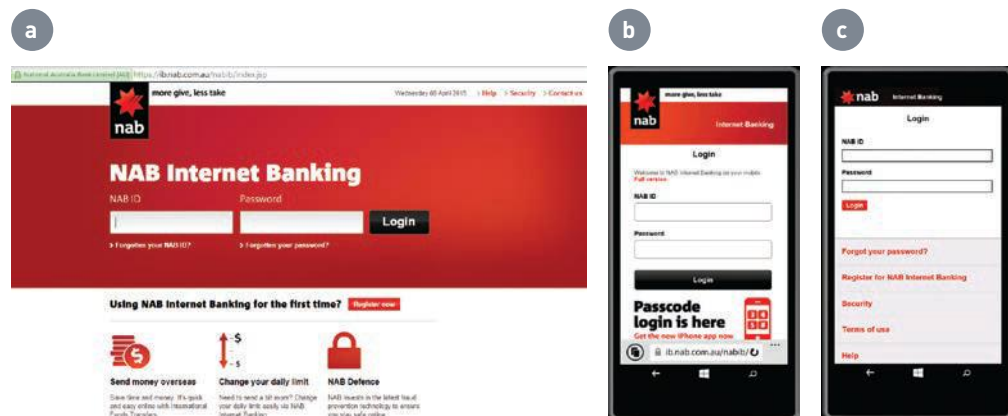


FIGURE 1.20 A comparison of logon screens for internet banking. **a** On the left is the desktop browser version. **b** The mobile version of the website is shown in the centre. **c** On the right is the stand-alone 'app' version. Note that a secure connection has been established on all devices.

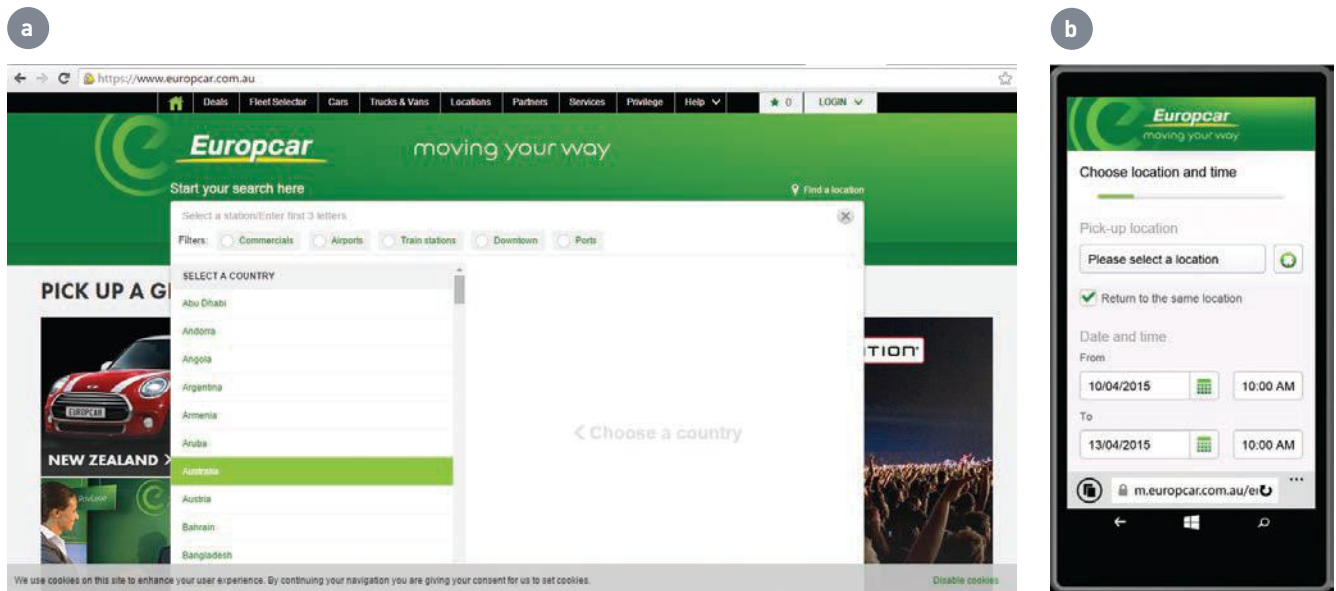


FIGURE 1.21 This is the screen to book a vehicle through Europcar. **a** In the desktop version on the left, you need to select your country from a large dropdown list. **b** However, the mobile version on the right makes use of GPS location services to make selecting your location easier and (hopefully) more accurate.

Design principles

Appearance

From the examples earlier in the chapter, you have seen a range of techniques that can be used for collecting data. To be efficient and effective, a data collection form must be more than simply a few fields stuck on a page. To maximise the user experience, the layout of the form must be carefully considered. For the purposes of this study, you must be familiar with several design principles related to the appearance of a solution. These principles include alignment, repetition, contrast, space and balance (Figure 1.22).

Alignment

The human eye seeks to make connections as it looks at forms. One way to help with creating an invisible 'line' to follow is to apply the principle of alignment. In the context of web forms it refers to justifying text and placing objects such as input boxes and graphics in a way that guides the eye, making it easier for the user to follow the path of instructions. Alignment avoids an unprofessional, random look to the placement of objects on the screen.

In terms of data collection, input boxes are often aligned on the left-hand edge to create an invisible vertical line. Objects aligned across a page allow a user to follow a series of steps from left to right. In Figure 1.23 this is evident from the placement of the MCRLogo image, 'Picking Up Location', 'Pickup Date' and 'Car Type' input boxes down the left-hand side of the form, as well as the horizontal line that runs through the two 'location' and 'date' related boxes, respectively.

THINK ABOUT INFORMATICS 1.2

Visit a range of websites that include data collection. Take screen shots and annotate as many data collection techniques as you can find from the suggestions on this page.

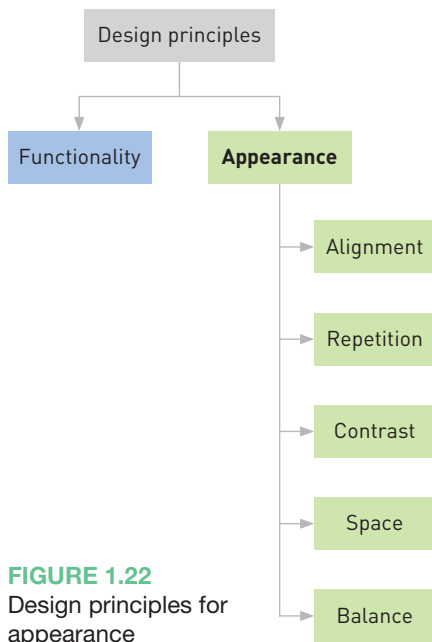


FIGURE 1.22
Design principles for appearance

Repetition

Repetition is used to unify elements of a layout. This is achieved by repeating patterns, textures and page elements. For example, a set of bullet points creates a sense of repetition that connects those points. On a data collection form, repetition usually means that each input box looks similar so that it is easier for a user to understand where details need to be entered, and how. Symbols may also be used consistently. For instance, an asterisk (*) may signify mandatory fields or a thumbnail graphic that reads 'SALE ITEM' may be placed beside each photo in a product gallery.

Contrast

When we position elements that are dissimilar beside each other we achieve contrast. Essentially, the greater the difference between an image or text and their surroundings, the more they stand out. This principle is followed by normally using a light-coloured page background with dark (usually black) text and input box borders. In addition, the title banner of a form is usually designed to contrast with the body. In Figure 1.24, the top banner and logo contrast so they stand out and are distinguishable from the form. The banner in Figure 1.23 has a similar contrasting effect.

Space

Space is the area that separates onscreen objects. Ideally, we space objects so that they are easy to perceive, rather than overlapping and obscuring them. Space is used in two ways. If we minimise space between objects, then we imply that those objects are grouped together. An input box and its label follow this principle. On the other hand, to reduce the look of overcrowding and to frame sets of common objects so that they *appear* grouped, but separate from other groups, we can make use of white space. In Figure 1.23, white space has been used so that each page element can breathe without being cramped. Similarly, the label for the 'Car Type' field and the field itself are positioned with minimal space while there is a great deal of space between this section and the 'Pick Up Date' field. Any of the 'app' examples in this chapter also show optimal use of space.

Balance

All elements of a layout have a visual weight. If the elements on either side or the top and bottom of the screen are of an equal weight then visual balance is achieved. There are two types of balance: symmetrical and asymmetrical.

With symmetrical balance, the visual elements on each side of an imaginary horizontal or vertical dividing line appear to be exactly the same in terms of visual weight, right down to the proportions and shading. A form with all the fields running down the centre of the page will achieve symmetrical balance. Most forms designed for mobile devices will achieve horizontal symmetrical balance because the fields normally span the width of the screen in equal proportion (Figures 1.19 and 1.20 b and c).

Asymmetrical balance occurs where a visually matched weighting occurs through a combination of objects of differing sizes, shapes and colours. A data collection form may have asymmetrical balance where it has fields down one side, but images or text on the other side of the screen (Figure 1.24).

Melbourne Car Rentals

For all your car rental needs, book with Melbourne Car Rentals

Picking Up Location
City, Airport or Address

Dropping Off Location
Same as Pick up

Pick up Date
dd/mm/yy 13:00

Drop off Date
dd/mm/yy 13:00

Car Type
No preference
Economy
Compact
Midsize
Standard
Fullsize
Premium
Luxury

Special Equipment (tick)
☐ Infant Seat
☐ Toddler Seat
☐ Ski Rack
☐ Snow Chains
☐ Navigation System

BOOK NOW!

Annotations:

- HTTPS with TLS protocols used (see bottom of page) to inform of site security
- Melbourne Car Rentals Logo to identify organisation (branding)
- GPS or other location services can be used for faster and more accurate location finding
- Dates can be picked from calendar. This also helps to ensure that only valid dates are entered
- Drop down list displays choices for 'Car Type'. This makes choices more accurate
- Labels for what data should be entered are placed inside the input boxes
- Drop-down list to choose pickup or dropoff time for car. It defaults to 1pm (1300hrs) and is formatted in 24-hr time
- Tick boxes to choose extra options
- Large, easy to locate button to allow users to proceed with booking
- Consistent Sans Serif typeface will be used to aid onscreen readability

FIGURE 1.23 Mockup of the first page of a transaction screen, created in Balsamiq, showing the key input boxes. Annotation shows how its features allow for efficient, effective data collection. Security methods that protect data during transmission are shown.

DEVICE CLOUD

Create your free Developer Edition account

Account Information

First Name *
Last Name *
Email Address *
Confirm Email Address *
User Name (Minimum of 8 Alphanumeric characters) *
Password *
Confirm Password *
Title *
Company Name *
Company Address *

Device Cloud Developer Edition
Register up to 5 devices for free with a Developer Account and enjoy access to Device Cloud's suite of features to design and develop your Internet of Things solution.

FIGURE 1.24 Asymmetrical balance is achieved on this form by using input fields on the left and an image and text on the right. Contrast is also used with the blue bar across the top of the screen, compared with the general white background of the page body.

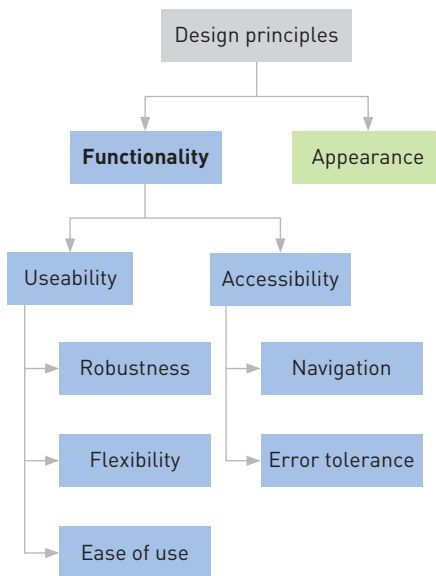


FIGURE 1.25
Design principles for
functionality

Functionality

It is important to develop an understanding of the design principles related to the *functionality* of a solution, not just those related to the *appearance* of a solution. Broadly speaking, there are two key principles we need to consider: useability (including robustness, flexibility and ease of use); and accessibility (including navigation and error tolerance). Each is discussed in the context of an RDBMS-based solution.

Useability

Useability refers to a solution's capability of being used and its availability to a user. It is important that any solution incorporates maximum compatibility with a wide range of devices, including future technologies, so that users are not tied to a particular platform or device. This is one reason why websites and apps are designed to work on desktop PCs as well as tablets and smartphones.

In addition, the solution must provide feedback to a user. The feedback might be an error message, but in most cases it is straightforward information that the system says it understands what it has been asked to do, says when it is doing it and then when it has finished. This feedback also provides a context for where the user is in the system. For example, when a user follows the steps of an online transaction they should be able to see how many steps they have completed and how many steps remain. In addition, the solution explains what it is doing or will do when the NEXT or SUBMIT button is clicked. Error messages, usually to do with validation, need to explain to a user in the clearest, most succinct language what has gone wrong and how to correct it to continue.

Linked with feedback is support for recovery, also known as robustness. If there is a problem or a user needs to correct data or undo an action, then the system should be able to step back then forwards again without the user having to restart the entire transaction. For example, a user might add the products they want to the shopping cart, fill in the shipping details, and then, when they reach the payment stage, realise they need to go back and change the delivery address. They should be able to do this and then continue to the payment screen without any effect on the rest of the transaction. This also applies where validation is triggered in a form. For instance, if an invalid credit card number is entered, the form should hold all other details so the user can simply correct the card number and proceed with the transaction.

Flexibility and ease of use

A solution should be as user friendly as possible for the widest audience possible, and provide support for the various ways that users may perform tasks. Flexibility and ease of use can be achieved in several different ways.

Having a user pre-emptive solution where the user initiates and controls the actions is considered more flexible than one where the solution does all the prompting and the user merely replies. This might include having the user choose the medium of interaction (e.g. email, text message or push notifications via app), and allowing the user to perform certain actions in different ways, allowing them to specify or configure data, and offering different ways of presenting information. This provides a degree of customisation of the solution to the user.

If a user can control their interaction then the solution must have the ability to adapt the user interface to different requirements and needs. In addition, the system must respond within the time that a user considers reasonable, regardless of how it has been configured.

The solution must also be able to support all of the tasks a user wants to perform in the way they expect – this is called **task completeness**. In addition to being able to perform all tasks

expected of it, the solution must also make it easy for the user to understand how to perform those tasks and even allow for the user to define new tasks and ways of performing them.

Finally, a system should support task migration. This is the process of moving a task to an entity who can do it better. For example, the user is the best entity for entering their own personal details for a shipping address, but the system is better at calculating the shipping costs and final totals for payment.

Accessibility

Accessibility means that a system is open (accessible) to the widest range of users. This includes users in different locations, with different cultural backgrounds, abilities and genders. The two important aspects of accessibility are navigation and error tolerance.

In terms of navigation, these are some general rules to follow.

- Help users navigate and find content. The navigation system needs to be clear, simple and intuitive.
- Ensure that navigation can also be done on keyboard-only systems.
- Give users enough time to read and use content. This is especially true of information to users about the consequences of them clicking buttons that will finalise a transaction and/or process payment.
- Use text in place of images for users who cannot perceive the images (e.g. they may be blind or do not understand the meaning of the images).

Error tolerance also has a list of simple considerations.

- Help users avoid and correct mistakes with clear instructions and the ability to undo errors by going back to a previous screen. (This is also connected to robustness.)
- Make it hard for a user to make errors and avoid allowing actions that lead to errors. For example, grey out non-selectable options and ask for confirmation of major actions, such as a purchase, or a deletion of an account.

The World Wide Web Consortium (W3C) is an international community that develops standards to ensure the long-term growth of the Web. The W3C mission is to lead the World Wide Web to its full potential by developing protocols and guidelines that ensure the long-term growth of the Web. The purpose of the Web Accessibility Initiative is to develop strategies, guidelines and resources to make the Web accessible to people with disabilities, and it includes the categories of functionality covered in the Study Design.



World Wide Web Consortium (W3C)

DATA ACQUISITION SOFTWARE

The main software used for acquiring data on websites are **PHP** and **ASP**, which are languages used mainly at server level. Other technologies used by organisations include back-end tools such as Google Analytics and **cookies**.

PHP and ASP

PHP is a general purpose scripting language that is used on web servers. PHP stands for hypertext pre-processor. It is also free of charge. It is used to create **dynamic webpage** content such as Google Earth or Google Maps. Dynamic webpage content requires PHP code to be run and executed by the PHP runtime. PHP is very versatile because it can operate on many web servers, operating systems and platforms, and can be used with any relational database management system.

PHP allows the user to be in control of the information viewed. In essence, dynamic websites allow users to enter data to access text or images. The webpage server accesses the database of information and searches for the information requested by the user. Once located, the information is brought forward to the webpage and displayed. Other examples of dynamic webpages include search engines, online banking, the Bureau of Meteorology and any other websites that return pages based on data the user provides.

Creating a dynamic webpage requires a certain level of programming experience. Having a basic understanding of PHP or ASP is useful. Both PHP and ASP (particularly **ASP.Net**) are used to make webpages dynamic and are used primarily on the server side of things. Programs such as JavaScript are mainly used on the client side rather than the server side.



THINK ABOUT INFORMATICS 1.3

- 1 How is privacy jeopardised by the web?
- 2 Who can capture the data?
- 3 How can users keep their data secure?

For a web server to accept HTTPS connections, a public key certificate needs to be set up for the web server. The certificate needs to be signed by a trusted certificate authority. If this does not occur, then the web browser is unable to accept the certificate. The certificate holder is verified by the certificate authority. Some certificates are free to obtain and others have an annual fee. Therefore certificates have expiration dates. There have been occasions when certificates have expired before their expiration date because details of the private key had been compromised.

Back-end tools

Occasionally, organisations want statistics on who visits their website. They want to know which browsers were used to visit their site, what information was commonly searched for, the location of the visitors (local or international), the length of their visits and the sites they came from. Programs and services such as cPanel and Google Analytics provide organisations with this type of information.

Cookies and spyware

A cookie is a small file that a web server stores on the user's computer. Cookies typically contain data about the user, such as their email address and browsing preferences. The cookie is sent to the computer when a website is browsed and stored on the computer's hard disk. The next time the website is visited, the browser retrieves the cookie from the hard disk and sends the data in the cookie to the website. Cookies are not viruses because they cannot be executed or run, and they cannot replicate themselves; however, they can be misused as **spyware**. Cookies can be used to track people, which leads to privacy issues. (See Chapter 6 for more information about privacy.)

Protection of rights

Many organisations have guidelines in place to protect data that is supplied by individuals and organisations. These guidelines can be in the form of a **privacy policy**, a **shipping and returns policy** and terms and conditions in the case of an online retailer. In addition, security protocols such as **secure sockets layer (SSL)** or **transport layer security (TLS)** are in place to help protect the exchange of data via websites.

Security protocols

TLS, and its predecessor SSL, are protocols used to provide security for communications on the internet. Several versions of the protocols are popular with users of the World Wide Web and email.

HTTPS

HTTP stands for hypertext transfer protocol. Some website URLs start with **HTTPS** rather than HTTP. HTTPS stands for hypertext transfer protocol *secure*, which is a combination of HTTP and the SSL/TLS protocol. HTTPS provides encryption and security in terms of identification of the server. HTTPS is often used when a financial transaction takes place or sensitive data is exchanged via websites. It is widely acknowledged that the internet is not a secure network. HTTPS was created to provide a secure medium through the internet. In very simple terms, HTTPS relies on certificates that are issued by authorities and installed on the browser software. URLs using HTTPS begin with https:// and use port 443, while URLs using HTTP begin with http:// and use port 80. HTTP is *not* secure, thus allowing unauthorised people to access the data that is exchanged. HTTPS is designed to overcome this issue and is considered to be secure.

Privacy policies

An organisation that collects data on individuals or other organisations usually has a privacy policy about how it uses the data collected and to whom it will disclose it (Figure 1.27). By law, privacy policies must be located on a company's website so that they are easy to find. The policy must include details about what data is gathered and how it is used. As a consumer, it is important for you to read privacy policies because some organisations may use your personal details in ways you might not expect.



FIGURE 1.26 The Rebel website issues a security certificate to permit a secure online transaction. Certificates have dates of validity and expiration. Note that this connection uses the TLS security protocol.

rebel

Search Products

Footwear Clothing Sports Gym & Fitness Games & Leisure Fangear Gift Cards On Sale

Home > Privacy Policy

About rebel

- About Us
- Careers
- Charities & Corporate Partners

Customer Service

- FAQ's
- My Account
- Online Orders
- Payment & Pricing
- Products & Services
- Refunds & Returns
- Shipping & Delivery

The Legal Stuff

- Privacy Policy
- Terms & Conditions

Contact Us

Rebel Sport Limited Privacy Policy

Rebel has developed a Privacy Policy that complies with the requirements of the Privacy Act. When you provide Rebel with your name and contact details, the information will be used as permitted by you and only for legitimate business reasons. View our Privacy Policy for further details.

To view Rebel Privacy Policy in full, please click on the link provided, or contact the Privacy Officer on +61 7 3482 7751

[View the Privacy Policy in full.](#)
(PDF Format - 41kb)

Information Provided

Every effort has been made to ensure the accuracy and currency of information provided on this site. However Rebel cannot accept any responsibility for errors or omissions. As the ultimate end use of product is outside of our control, users should check the information supplied or seek detail from us regarding suitability or product requirements. To the extent permitted by law, Rebel is not liable for loss, damage or liability arising from errors, omissions, inaccuracies or any misleading or out of date information whether contained on this site or any link attached to this site.

Intellectual Property Rights

This site contains material in which the copyright is owned by Rebel. This material including wording, depictions and logos may be downloaded for personal use only and may not be used for any other purpose or further reprinted without prior written permission from Rebel.

FIGURE 1.27 The privacy policy section on the Rebel website. A brief version is displayed online, but a full version can be downloaded in PDF format.

Shipping and returns policies

Many organisations that sell goods and services online have shipping and returns policies to assist customers with how to proceed should the order not meet with their satisfaction (Figure 1.28). For example, if you ordered a t-shirt, but when the t-shirt arrived it was smaller than expected, the policy would provide guidance on how to return the item and the necessary data required to identify the order.

Contact Us Track My Order Help Desk rebel active Find a Store Login / Register View Cart

Footwear Clothing Sports Gym & Fitness Games & Leisure Fangear Gift Cards On Sale

Home > Help Desk > Customer Service > Refunds & Returns

About rebel

- About Us
- Careers
- Charities & Corporate Partners

Customer Service

- FAQ's
- My Account
- Online Orders
- Payment & Pricing
- Products & Services
- Refunds & Returns
- Shipping & Delivery

The Legal Stuff

- Privacy Policy
- Terms & Conditions

Contact Us

We want you to be totally satisfied with our products and services.

There may be times when you will need to return a product purchased from us. Our Returns Policy is designed to assist you with your return. The benefits given by our Returns Policy are in addition to other rights and remedies you may have as a consumer under the Australian Consumer Law (ACL).

Help us, help you

When to return a product

You are entitled to a replacement, credit or refund where products have a major failure, such as:

- When it is faulty
- Wrongly described
- Different from a sample shown to you, or
- Do not do what they are supposed to do.

We are not required to replace your product if you simply change your mind.

However, if you are not happy with your product, our Store Managers may, at their discretion, decide to exchange it or provide you with a gift card, providing the product is returned to us in its original condition and with proof of purchase. Please talk to us if you are unsatisfied with a product or the service you have received from us.

Where to return a product:

Take your product to the front desk of your nearest rebel store. This includes products that carry a manufacturer's warranty. Note that our clearance stores may not carry all product lines of a normal store, which can make exchange difficult. [Click Here](#) to access our store locator for all store locations.

FIGURE 1.28 Part of the 'Refunds and Returns' policy for Rebel. Notice there is also a 'Shipping and Delivery' policy listed in the Customer Service menu.

Privacy policies are required under the *Privacy Act 1988*. See Chapter 6 for more details about Australian privacy laws.

THINK ABOUT INFORMATICS 1.4

Gather privacy policies from three different organisations. Examine each policy and make a list of the common points made in each one.

Characteristics of data types

We have discussed a range of methods that organisations use to acquire data from users. In most cases, the input occurs via an online form populated with boxes into which the data is entered. It is important to understand the data types that are used because ultimately any data collected will be stored in a database (discussed later in the chapter) and the data itself needs to be reused for ongoing services or to finalise transactions.

Data types

Data types used in the Study Design include:

- text
- numeric
- date
- character
- Boolean.

Text (string)

The majority of fields can be set as a text data type. Text may also be referred to as a 'string', particularly in connection with programming. This type of field holds a mix of characters (letters, numbers, special characters), also referred to as alphanumeric, to a limit of 255. Names and addresses are considered text data. Postcodes and telephone numbers are normally formatted as text because they may contain spaces and are not intended to be used the same way as a numeric value. It is also more efficient to store the values as text rather than as a large numeric value.

Numeric – integer, floating point

Another key type of data format used in databases is numeric. These fields will only allow numbers to be entered. They are often used when the value is to be used in a calculation of some kind. For example, the quantity of an item purchased might need to be multiplied by its price to calculate a total amount payable. This calculation cannot be performed on a field formatted as text.

Numeric fields can also be categorised into different variations. 'Integer' refers to whole numbers, including negative numbers. Where decimal numbers are required, when dealing with financial transactions or percentages for example, then the 'floating point' data type is used. Depending on the database package you use, the specific names of these numeric data types may vary, but their function will be the same.

Date

Strictly speaking, a date format is another variation of a numeric data type. The value is normally based on the number of days since the 'zero' day built into the operating system or DBMS. For example, day '1' might be displayed as 01 January 1900, while '42673' would be displayed as 30 October 2016. Calculations can be performed on dates, which can be handy when comparing the difference between them. Dates can be formatted to show a combination of years, months, days, hours, minutes and seconds, depending on the needs of the user. In terms of time, they can also display 12- and 24-hour clocks.

Character

This is a text field that will only accept a single alphanumeric character. It is used where there are multiple options for a value, but they can be represented with a single character to make data entry easier and to save storage space. For example, the sizes of a wooden box that comes in small, medium and large sizes might be entered as 'S', 'M' or 'L' respectively. These options might be selected from a radio button group on a form.

Boolean

In cases where the data to be entered falls into the categories of Yes/No, True/False, or even On/Off, the Boolean data type is used. It is often represented as a tick box on forms.

TABLE 1.1 The most common data types used in databases

Data type	Characteristics or uses	Examples
Text	Alphanumeric; up to 255 characters; searchable	Name, address, postcode and telephone number
Numeric	Numbers only (see below for different formats)	Any number that will be used in a calculation
Date	A variation of numbers, but formatted to represent a date and/or time	Any date; can be used in calculations
Character	A variation of a text field that can only hold a single alphanumeric character	Where space is to be saved, but there are more than two options to choose from
Boolean	Represents one of two states, such as True/False	Also represented as Yes/No and On/Off

TABLE 1.2 The various number formats used in Microsoft Access

Field size setting	Description
Byte	A 1-byte integer containing values from 0 to 255
Integer	A 2-byte integer containing values from -32 768 to +32 767
Long integer	A 4-byte integer containing values from -2 147 483 648 to 2 147 483 647
Single	A 4-byte integer containing values from -3.4×10^{38} to $+3.4 \times 10^{38}$ and up to seven significant digits
Double	An 8-byte floating point number containing values from -1.797×10^{308} to $+1.797 \times 10^{308}$ and up to 15 significant digits
Replication ID	A 16-byte globally unique identifier (GUID). Randomly generated GUIDs are long enough that they are not likely to overlap. You use them for a variety of applications, such as tracking goods.
Decimal	A 12-byte integer with a defined decimal precision that can contain values from -1028 to +1028. The default precision is 0. The default scale (the number of decimal places displayed) is 18. You can set the scale up to 28.

THINK ABOUT INFORMATICS 1.5

Take a screenshot of an online data collection form and annotate each of the data input boxes to indicate its likely data type.

Databases

So far in this chapter, we have discussed the variety of ways that data can be collected online, and how it can be secured when transmitted.

The next stage is to understand what happens to the data an organisation acquires via its website. The software tool used to hold and retrieve data is called a database.

Database terminology

Databases are used to handle a range of data, from personal details to school results and financial transactions. Whenever you search on the internet, your keywords are used to query a database, which then displays a list of results. A **flat file** database stores data in tables consisting of rows and columns. Each column contains a **field** with a specific piece of information. Each row in a table holds a record. Table 1.3 shows how a field and record form the row and column elements of a table. A **database record** is a set of information about one entity (a person, event or object). A field relates to a specific piece of information. A **primary key** uniquely identifies each record in a database table. A spreadsheet can often be used as a simple, flat file database.

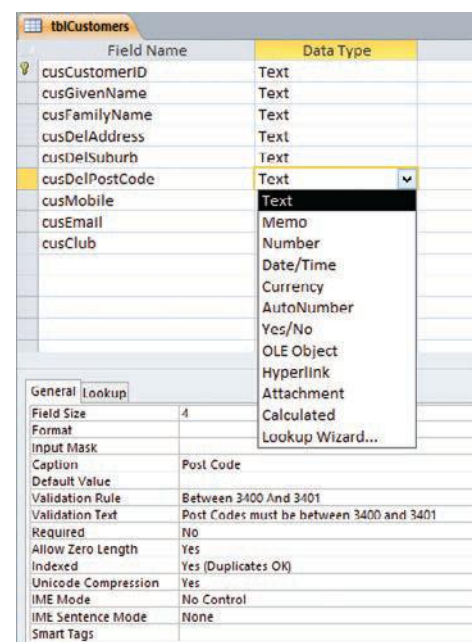


FIGURE 1.29 In a table's Design View, you can choose from several types of data to assign to a field

TABLE 1.3 A simple flat file database consists of a table with fields organised in columns and records in the rows

Customer ID	Family name	Given name	Suburb
1031	White	Tom	Kew
1032	Wu	Harry	Hawthorn
1033	Kozniak	Elijah	Balwyn
1034	Kaur	Sandeep	Balwyn

Although the table holds data, it is not a particularly efficient input tool. A form allows an input screen to be formatted and linked to an underlying table. The first part of this chapter has discussed online forms and their features.

When you need to select a set of data, you use a **query** (see below). The result of a query is usually turned into usable information by putting it into a report. A **report** formats the query data and allows you to add summary statistics, such as totals, as well as headings, to make the information easier to read and understand. It is important to note that, while tables can only hold data, queries, forms and reports can all have formula fields added to perform calculations.

Procedures in a database can be automated to some extent by the use of macros. When run, macros will carry out a set of predetermined tasks, such as printing a report.

For the purposes of this assessment task, you are only required to create tables and queries.

Queries

Database tables hold data, but the real power lies in the ability to query that data to answer questions and make links between data. A query allows you to select a set of specific data based on a series of criteria. The query criteria are the result of the questions we might have of the data. For example, 'How many of our customers in Fairfield are between the ages of 20 and 60?' (Figure 1.30a) or 'Can I have a list of names and numbers for all our customers in Fairfield or Northcote?' (Figure 1.31a).

Modern DBMSs use a structured query language or SQL as a way of standardising how data is managed in databases. Even DBMSs with GUI-based query creation, such as Microsoft Access, use SQL in the background (Figure 1.32).

In Figures 1.30 and 1.31, each query has selected only the requested data. A properly structured query helps to ensure any information we create is both relevant and complete.

Formulas are placed on the input form, **query** or **report**, but they are not linked to a field in the table. A table cannot store the result of a calculation.

You query a database of websites when you type words in the search box on the Google main page.

a

Field:	cusCustomerID	cusCustomerTitle	cusCustomerFirstName	cusCustomerLastName	cusCustomerAge	cusCustomerStreet	cusCustomerSuburb	cusCustomerPostCode
Table:	tbiCustomers	tbiCustomers	tbiCustomers	tbiCustomers	tbiCustomers	tbiCustomers	tbiCustomers	tbiCustomers
Sort:								
Show:	☒	☒	☒	☒	☒	☐	☐	☐
Criteria:					Between 20 And 60		Fairfield	
or:								

b

Customer ID	Title	First Name	Last Name	Age (years)
5825Y	Ms	Zoe	Robertson	52
4820M	Mrs	Maria	Johnson	26
2026X	Mrs	Sierra	Jarvis	41

FIGURE 1.30a The design and **b** results for a query to select all customers living in Fairfield who are between the ages of 20 and 60

a

Field:	cusCustomerTitle	cusCustomerFirstNam	cusCustomerLastNam	cusCustomerSuburb	cusCustomerMobile
Table:	tblCustomers	tblCustomers	tblCustomers	tblCustomers	tblCustomers
Sort:					
Show:	☒	☒	☒	☐	☒
Criteria:				"Fairfield" Or "Northcote"	

b

Title	First Name	Last Name	Mobile Number
Ms	Zoe	Robertson	0442 205 3
Ms	Fiona	Santiago	0409 280 2
Mrs	Maria	Johnson	0419 123 4
Mr	Jeremy	Mcperson	0418 128 6
Ms	Fiona	Holmes	0454 236 1
Dr	Elijah	Valentine	0454 154 7
Mrs	Sierra	Jarvis	0412 862 7
Total			7

FIGURE 1.31a The design and **b** results for a query to create a list of names and mobile numbers of all customers living in Fairfield or Northcote

```

SELECT tblCustomers.cusCustomerID, tblCustomers.cusCustomerTitle, tblCustomers.cusCustomerFirstName, tblCustomers.cusCustomerLastName, tblCustomers.cusCustomerAge
FROM tblCustomers
WHERE (((tblCustomers.cusCustomerAge) Between 20 And 60) AND ((tblCustomers.cusCustomerSuburb)="Fairfield"));

```

FIGURE 1.32 The SQL view for the query in Figure 1.30

Relational databases

As useful as a flat file database may be, it has limitations. The main limitation lies in the creation of redundant data. Often this takes the form of data that is repeated in each transaction, such as the mailing address of a customer or the details of a product purchased. To increase the efficiency and effectiveness of their data, organisations typically make use of relational databases instead of flat file databases. A software package written specifically to create these databases is called a **relational database management system** or **RDBMS**.

To overcome the limitation of flat file databases, a RDBMS is made of multiple tables and also stores data relationships between tables. A relationship is a connection between the data. For a relationship to be established between two or more tables, they must have a common field. The primary key in a table usually acts as the field that joins the tables. When a primary key is used in another table it is known as a **foreign key**.

A **one-to-one relationship** is used when a record in one table is connected to only one record in a second table. For example, an airline's passenger details table will contain records for many passengers, while a seat allocation table may hold records related to the seats on a particular flight. A one-to-one relationship exists between a passenger and their seat allocation. Each passenger has only one seat and each seat can be assigned to only one passenger.

A **one-to-many relationship** indicates that one record in the first table can be connected to more than one record in a second table (Figure 1.33). For example, several workers in an office may share a single telephone extension. Each extension record is related to several employee records.

Data in a **one-to-one relationship** can normally be accommodated on one table. Separate tables might be used if they have already been established for another purpose.

How many is many?
Consider a **one-to-many relationship** between individuals and cars. Each car is owned by only one person, but each individual may own one, five, 20 or no cars. Thus 'many' can mean any number, including zero.

A **many-to-many relationship** is used when each record in the first table can be connected to a number of records in the second table. At the same time, each record in the second table may be related to many records in the first table. For example, a student detail table and a subject detail table may have a many-to-many relationship. Each student studies many subjects, and each subject is studied by many students. Another example of a many-to-many relationship is one in which an order from a customer for goods is related to a range of products. Each order can contain a number of products, while each product can appear on a number of different customer orders.

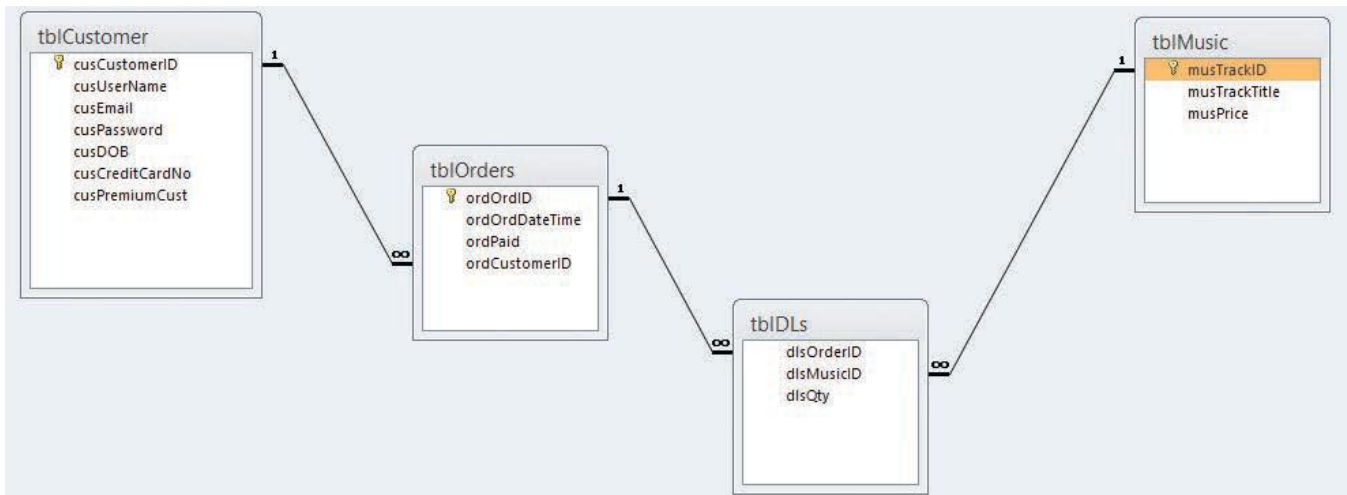


FIGURE 1.33 Tables linked by one-to-many relationships

Creating an RDBMS structure

Once you have identified that a relational database is required to solve an information problem, you will need to establish how to break up the data that will be supplied by the client into fields and tables. It is important to plan this carefully in order to maximise the efficiency of a relational database structure.

Determine the entities that exist in the problem. An **entity** is a single person, place or thing about which data can be stored. Then the characteristics of each entity must be established. For example, these may include data elements like ID numbers, names, dates, addresses and prices. An entity–relationship diagram (ERD) will assist with this (ERDs are discussed below). Each of the entity’s characteristics would then be defined as a field. All the related fields of each entity would then be assigned to a table.

At this point the arrangement of fields in tables may not be as efficient as possible. For this reason we would apply table **normalisation** (explained below) to reduce the redundant fields and maintain data integrity.

Within each table the primary key would be identified to make the links with other tables where they will become foreign keys. Remember: the primary key must be a value that is unique for each record, such as a customer ID number. In terms of the relational aspect of the database, foreign keys (known as referential integrity in Access) are used to ensure that if you are entering data in one table, it already has a corresponding value in another table.

When you are happy with the table structures, establish the most appropriate data type and size (the number of characters to be stored in the field) for each field. To do this, examine the sample data supplied to you. This usually serves as a good indication of the general content for each field. At this point you would be ready to create the RDBMS. The sections below follow this process and show you some of the design tools that will assist you in determining the structure of a RDBMS.

The process of removing redundant data and arranging it into appropriate tables is called normalisation. **Normalisation** is discussed later in this chapter.

**CASE
STUDY****PIZZAS ON THE GO – PART 1**

Pizzas on the Go (POTG) is a pizza shop in Horsham, a town in western Victoria. The owner, Matt, strives to make high-quality, fresh pizzas for an increasingly discerning local clientele. POTG offers both a takeaway and a delivery service to the local community. Over the past five years, Matt has built up a successful business and his pizzas are very popular. While POTG has many customers order directly in-store, the company also takes phone orders. These are entered in a spreadsheet for tracking and later financial reconciliation. This process is time-consuming, especially when it is a busy night and Matt has to juggle both the constant calls plus making the actual pizzas. On several occasions, orders have been recorded incorrectly, leading to embarrassment and upset customers as well as lost revenue for Matt. Some customers cannot remember what they have ordered in previous transactions, but have often said it would be helpful when reconciling their home budget each month.

As a result of this, Matt has turned to a local ICT professional, Nick, for advice on how to improve the ordering process with an online solution. Nick has conducted an analysis of the problem and suggested the creation of an online form and smartphone app that customers could use to make orders via the internet. Orders would feed into a back-end database for processing. The database would have multiple tables and queries that would also allow Matt to compare sales of various types of pizzas, as well as volume of sales at different times each day and between different times of the year. In addition to making online pizza orders, customers would also be able to review their own order history. Nick has established a number of requirements.

Functional requirements

- Customer details must be recordable.
- Existing customer details must be editable.
- Only postcodes in the surrounding area are accepted (3400 and 3401).
- Pizza details must be recordable.
- Existing pizza details must be editable.
- Lists of sales of various pizzas can be generated for comparison.
- Customer ordering history must be reviewable by customers.
- Each order must allow for multiple types of pizzas to be purchased.
- Order totals must be accurate.

Non-functional requirements

- Data must be stored as efficiently as possible.
- Queries must be complete and accurate, and contain relevant data.

Constraints

- Customers will be using a variety of electronic devices to enter order data.
- Matt is confident with PC use, but unfamiliar with database software.
- The solution must be in use before April.
- The solution must be able to accommodate all users regardless of experience or ability.

Scope of the solution***The solution will:***

- allow POTG customers to order pizzas online
- add, remove and edit records of POTG customers and the pizzas they order
- add new orders
- select data relating to sales of specific pizzas according to specified criteria
- select data relating to the ordering history of specific customers according to date ranges they specify.

Case study continues ►

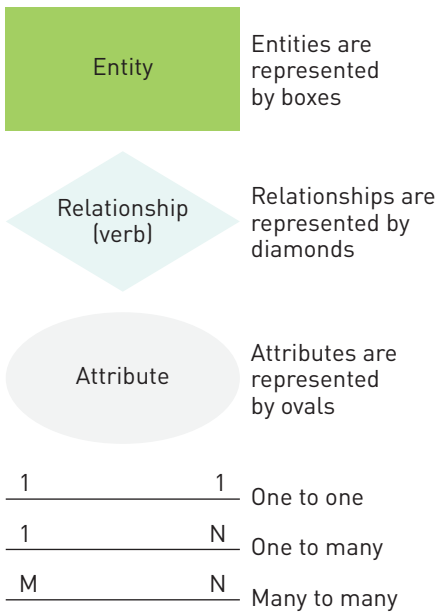
◀ Case study continued

The solution will not:

- record credit card details
- allow for ordering of side orders, such as drinks and pasta.

This solution will be beneficial because it will:

- accurately and quickly record pizza orders, reducing the likelihood of an incorrect order being recorded
- create accurate, complete and relevant lists of pizza sales, allowing Matt to compare sales of different pizzas at different times
- create accurate and timely lists of transactions and their details for customers to check their ordering history.



Entity–relationship diagrams (ERD)

An **entity–relationship diagram (ERD)** is used by database designers to establish the interrelationships between different data elements. Once entities have been determined and their attributes identified, an ERD is created to show how the entities relate to each other. There are different ways of representing ERDs. The Study Design does not specify a preferred style, but there are two commonly used types that are simple to create – the Chen and Bachman models. The examples below are simplified versions of these types. In an ERD, entities are things about which information is sought (such as books or films) and attributes are the elements of data we collect about those entities (such as title and author of a book). Relationships show the linking to draw related data from different entities.

To create an ERD, first identify the entities, then define the relationships and finally add the attributes to each entity. In the Chen model, ERDs use a simple set of symbols, much like a flowchart (Figure 1.34), while the Bachman style will show the attributes in a table for each entity (Figure 1.37). Both types of ERDs also show relationships which exist between the entities in different ways.

FIGURE 1.34 In Chen-style ERDs, boxes are commonly used to represent entities. Diamonds are normally used to represent relationships (usually a verb) and ovals are used to represent attributes. Relationships use '1' for 'one' and 'N' for 'many'. 'M' represents 'many' in place of the '1'.

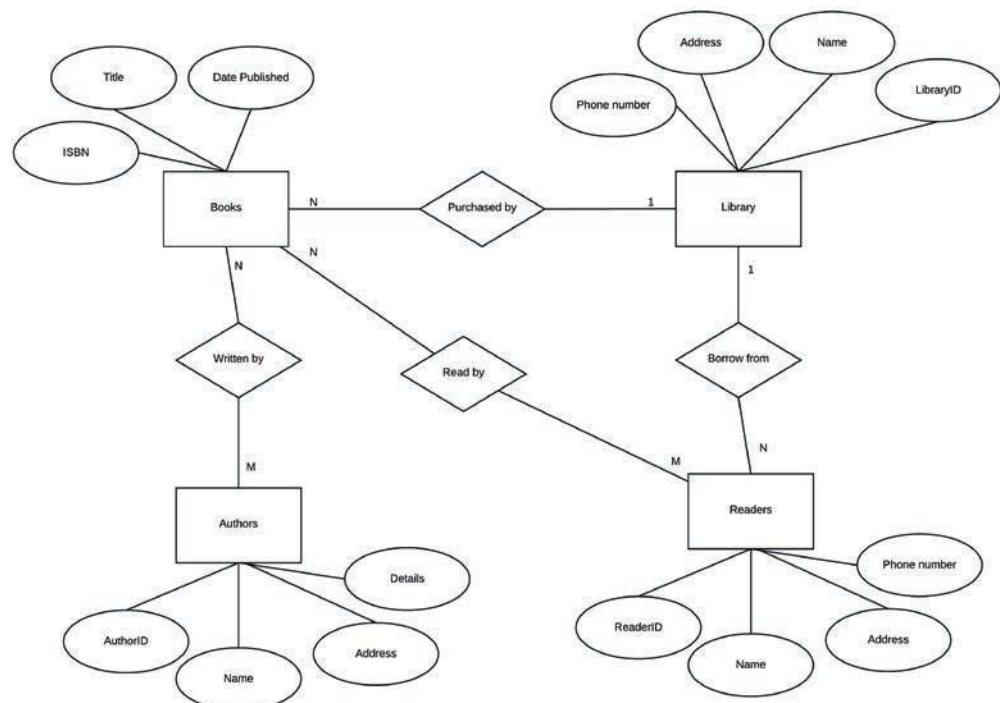


FIGURE 1.35 This Chen-style ERD example shows the relationships that may exist in a library database

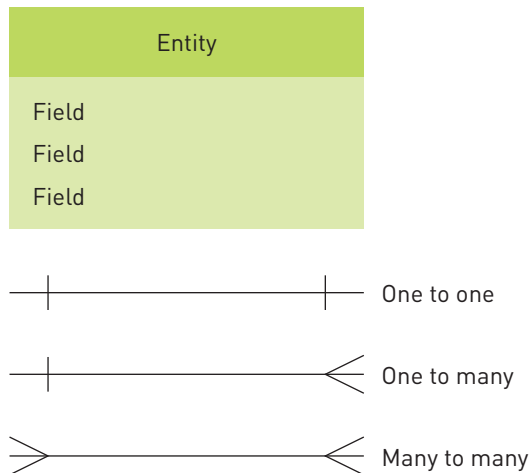


FIGURE 1.36 In Bachman-style ERDs, the entities and their attributes are displayed together in a box. Relationships are indicated with lines. The three-pointed line end that represents ‘many’ is also referred to as a ‘crow’s foot’.

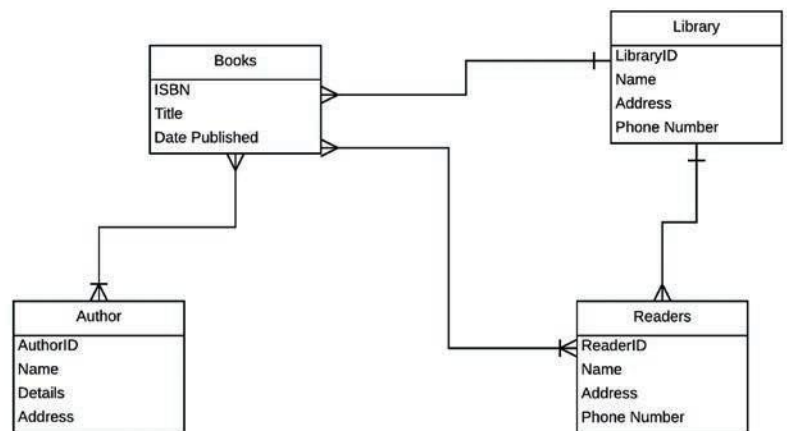


FIGURE 1.37 This is the same library ERD as in Figure 1.35, but represented with the Bachman style

CASE STUDY

PIZZAS ON THE GO – PART 2

Nick has examined some of the existing records and interviewed Matt. From this research he has constructed the ERD in Figure 1.38 to represent the main entities and attributes and their relationships with one another.

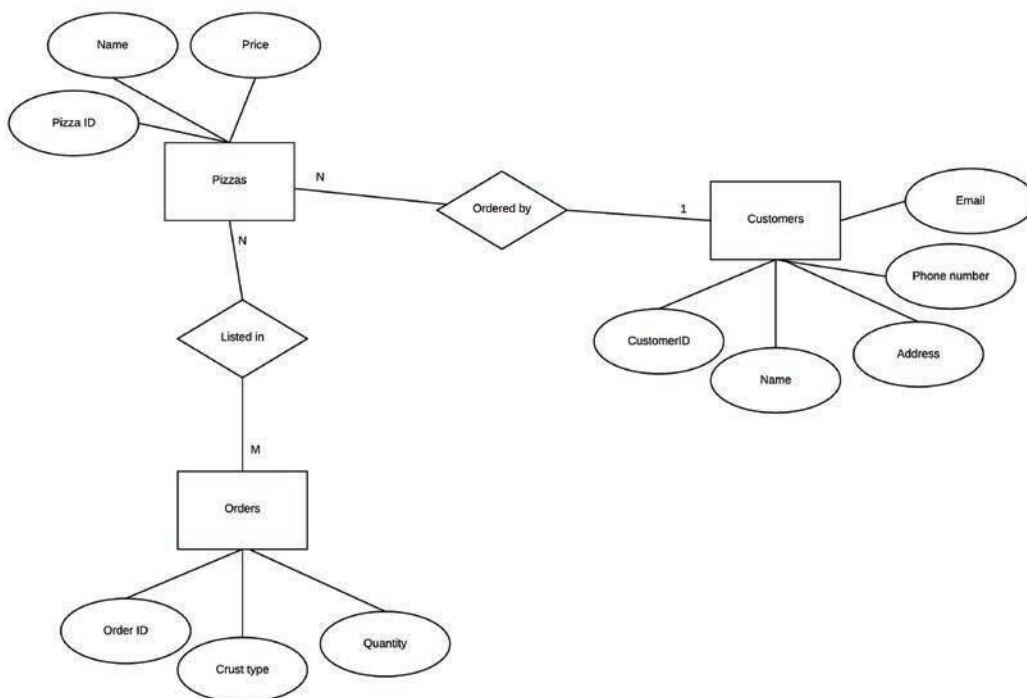


FIGURE 1.38 Nick has identified three key entities for this information solution: pizzas, orders and customers.

Phone numbers, addresses and email addresses are partly redacted throughout this chapter for privacy reasons.

Table normalisation

Once the ERD is created, the elements are broken down 'on paper' into proper fields and tables. But the proposed table structure may not be as efficient as it potentially could be. So, at this point, table normalisation is employed. Normalisation is the process of ensuring that a database conforms to a set of normal forms. Its primary purpose is to remove redundancies that create threats to **data integrity** such as update anomalies. It also plays a role in making querying more efficient. There are six 'normal forms'. Each rule is applied successively from the **first normal form (1NF)** to the sixth normal form (6NF), but for the purpose of this study, you only need to apply the first three normal forms – 1NF, **second normal form (2NF)** and **third normal form (3NF)**.

Normalisation works by providing a set of rules and a systematic procedure to check for various anomalies or deviations in data structure that would make the database less efficient, by ensuring that your fields are in the correct tables. It will not help if you have not chosen the correct fields in the first place.

A well-planned database will automatically comply with the normalisation rules. If the structure of the database does not fit into the normalisation rules, make adjustments to your designs and try the rules again.

a

	A	B	C	D	E	F	G	H	I
1	CustomerID	Name	Address	Suburb	Post Code	Contact Person 1	Telephone 1	Contact Person 2	Telephone 2
2	123	Widgets Inc	35 Colins Street	Melbourne	3000	Dawson, Bryn	0492 400 1	Nieves, Helen	0433 529 4
3	388	Poplets Pty Ltd	13a Mavis Road	Hawthorn	3122	Carter, Martin	0431 246 4	Clarke, Kareem	0489 440 5
4	50	Gears and Sprockets	Lot 2 Murray Road	Preston	3072	Haynes, Nina	0495 339 1	Fleming, Carolyn	0461 577 1

b

	A	B	C	D	E	F	G
1	CustomerID	Name	Address	Suburb	Post Code	Contact Person	Telephone
2	123	Widgets Inc	35 Colins Street	Melbourne	3000	Dawson, Bryn	0492 400 1
3						Nieves, Helen	0433 529 4
4	388	Poplets Pty Ltd	13a Mavis Road	Hawthorn	3122	Carter, Martin	0431 246 4
5						Clarke, Kareem	0489 440 5
6	50	Gears and Sprockets	Lot 2 Murray Road	Preston	3072	Haynes, Nina	0495 339 1
7						Fleming, Carolyn	0461 577 1

FIGURE 1.39 Examples of unnormalised tables stored in a spreadsheet. **a** Notice the repeating groups for Contact Person and Telephone. **b** The same data is displayed and, even though it looks neater, it repeats groups vertically.

Figure 1.40 shows a typical spreadsheet used to record orders. This could also be the single table in a flat file database. Spreadsheets are popular for holding raw data and for calculations because they are easy to construct and use, but their weaknesses become apparent when we start to ask more complicated questions of the data, such as 'How many of our Hawthorn customers bought products last week?', 'What is the value of orders placed by Widgets Inc. in the last financial year?' and 'What are our most popular items in each state?'

To put this data into a form where we can easily and accurately get answers to these types of questions, we need to normalise to get the table and field structures needed to construct a proper RDBMS.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Invoice No.	Date	CustomerID	Name	Address	Suburb	Post Code	Item ID	Item Description	Item Qty	Item Price (\$)	Item Total (\$)	Order Total Price (\$)
2	X239	10-Jan-16	123	Widgets Inc	35 Colins Street	Melbourne	3000	6673	Shovels - Long handled	5	25.95	129.75	
3								7621	Hammer - 2lb	2	39.95	79.90	
4								1523	M10 Screw - 12cm	80	1.25	100.00	
5								492	Bucket - 9Lt	4	2.95	11.80	321.45
6	X240	11-Jan-16	388	Poplets Pty Ltd	13a Mavis Road	Hawthorn	3122	7621	Hammer - 2lb	3	39.95	119.85	
7								6673	Shovels - Long handled	1	25.95	25.95	
8								943	Rivets - bag of 200	2	9.99	19.98	165.78
9	X241	11-Jan-16	50	Gears and Sprockets	Lot 2 Murray Road	Preston	3072	1523	M10 Screw - 12cm	120	1.25	150.00	
10								2884	Drill bits - assorted	8	2.59	20.72	170.72

FIGURE 1.40 A typical, unnormalised spreadsheet file used to store customer order details

First normal form (1NF)

This rule states that there must be no repeating groups in the table. This means that no single row (record) contains more than one value in a field, nor will there be more than one column with the same kind of value. For example, in a table listing products for online purchases, only the current price is recorded in the `Price` field, rather than the original price *and* the sale price. Where multiple telephone numbers exist for a customer, there are no `telephone 1`, `telephone 2`, `telephone 3` fields, only a single `telephone` field (Figures 1.39a and b).

Figure 1.40 is an unnormalised set of data. To get it into 1NF we need to first ‘flatten’ the table. While it looks like we are creating more duplication, this will all be removed by the time we reach 3NF. Figure 1.41 shows the flattened orders table.

A table's primary key is the smallest set of columns needed to uniquely identify a row in the table.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Invoice No.	Date	CustomerID	Name	Address	Suburb	Post Code	Item ID	Item Description	Item Qty	Item Price (\$)	Item Total (\$)	Order Total Price (\$)
2	X239	10-Jan-16	123	Widgets Inc	35 Colins Street	Melbourne	3000	6673	Shovels - Long handled	5	25.95	129.75	321.45
3	X239	10-Jan-16	123	Widgets Inc	35 Colins Street	Melbourne	3000	7621	Hammer - 2lb	2	39.95	79.90	321.45
4	X239	10-Jan-16	123	Widgets Inc	35 Colins Street	Melbourne	3000	1523	M10 Screw - 12cm	80	1.25	100.00	321.45
5	X239	10-Jan-16	123	Widgets Inc	35 Colins Street	Melbourne	3000	492	Bucket - 9Lt	4	2.95	11.80	321.45
6	X240	11-Jan-16	388	Poplets Pty Ltd	13a Mavis Road	Hawthorn	3122	7621	Hammer - 2lb	3	39.95	119.85	165.78
7	X240	11-Jan-16	388	Poplets Pty Ltd	13a Mavis Road	Hawthorn	3122	6673	Shovels - Long handled	1	25.95	25.95	165.78
8	X240	11-Jan-16	388	Poplets Pty Ltd	13a Mavis Road	Hawthorn	3122	943	Rivets - bag of 200	2	9.99	19.98	165.78
9	X241	11-Jan-16	50	Gears and Sprockets	Lot 2 Murray Road	Preston	3072	1523	M10 Screw - 12cm	120	1.25	150.00	170.72
10	X241	11-Jan-16	50	Gears and Sprockets	Lot 2 Murray Road	Preston	3072	2884	Drill bits - assorted	8	2.59	20.72	170.72

FIGURE 1.41 The orders table has now been flattened. This table is now compliant with 1NF.

We now need to identify a primary key (PK) for each record in the table. There is no single column (field) we can use, but if we combine the `Invoice No.` and the `Item ID` then we can create a concatenated PK. For example, we might think of the concatenated PK for Row 1 as X2396673 and for Row 6 as X2407621. This table is now in 1NF.

Second normal form (2NF)

Put simply, this is where a table is in 1NF and any column (field) in that table that is not part of the actual PK must be wholly dependent on the concatenated PK. In other words, in situations where you have more than one PK field in a table, each non-key field must be fully dependent on both keys, not just dependent on one of the keys. To determine if tables comply with 2NF, look at each non-primary key field and determine if the field is wholly related to each of the PKs. If the field can exist independently of any of the PKs (that is, it is not wholly reliant on *all* the keys) then the table is *not* in 2NF. For example, in our orders table from Figure 1.39 we have to see which fields depend on both the `Invoice No.` and `Item ID` columns. If we start with the `Date` field, we can say that it clearly relies on an order being made, so it is dependent on `Invoice No.` But an order date can exist without an `Item ID`, so immediately our table fails the 2NF check!

TABLE 1.4 What happens when the 2NF rule is applied to the other fields in our table from Figure 1.39. Of note are the fields that do not fit with either part of the concatenated PK, such as all the customer details, which will be dealt with in the 3NF process.

Column	Invoice No. (PK)	Item ID (PK)	Meets 2NF?	Comments
Date	✓	✗	✗	Only relates to Invoice No.
CustomerID	✗	✗	?	Will deal with this field in 3NF
Name	✗	✗	?	Will deal with this field in 3NF
Address	✗	✗	?	Will deal with this field in 3NF
Suburb	✗	✗	?	Will deal with this field in 3NF
Post Code	✗	✗	?	Will deal with this field in 3NF
Item Description	✗	✓	✗	It never needs to be ordered!
Item Qty	✓	✓	✓	There is no quantity without an order and the quantity must be of an item.
Item Price	✗	✓	✗	Every item which exists must have a price, but the item doesn't have to be sold.
Item Total	✗	✗	✗	Both of these columns are derived from calculations, so they have no place in our table.
Order Total Price	✗	✗	✗	

The next step is to take all the fields that depend in whole or in part on the second half of the concatenated PK (the `Item ID` field) and put them in their own, new table. The other fields, including the ones we are unsure about, will remain in the existing table. Because this table no longer has a concatenated PK, it is now 2NF compliant. The 2NF checking process will be run again to see how the fields in the new `Order Items` table depend on the concatenated PK it has.

TABLE 1.5 Orders table: What the tables look like at the end of this first part of the 2NF process

Invoice No. (PK)	Date	CustomerID	Name	Address	Suburb	Post Code
X239	10-Jan-16	123	Widgets Inc	35 Colins Street	Melbourne	3000
X240	11-Jan-16	388	Poplets Pty Ltd	13a Mavis Road	Hawthorn	3122
X241	11-Jan-16	50	Gears and Sprockets	Lot 2 Murray Road	Preston	3072

TABLE 1.6 Order Items Table

Invoice No. (PK)	Item ID (PK)	Item Description	Item Qty	Item Price (\$)
X239	492	Bucket – 9L	4	2.95
X239	1523	M10 Screw – 12 cm	80	1.25
X239	6673	Shovels – Long handled	5	25.95
X239	7621	Hammer – 2lb	2	39.95
X240	943	Rivets – bag of 200	2	9.99
X240	6673	Shovels – Long handled	1	25.95
X240	7621	Hammer – 2lb	3	39.95
X241	1523	M10 Screw – 12 cm	120	1.25
X241	2884	Drill bits – assorted	8	2.59

These are the two tables we end up with after the first part of the 2NF process. Invoice No. is brought across to the Order Items table so that we ‘remember’ the order to which each item belongs. The Orders table no longer has a concatenated PK, because it is now based on the single Invoice No. field, but Order Items retains a concatenated PK.

In the second part of the 2NF process, we re-apply the 2NF rules to the Order Items table. Again, we are looking for fields that depend on both parts of the concatenated PK.

TABLE 1.7 The results of re-applying the 2NF rules to the Order Items table

Column	Invoice No. (PK)	Item ID (PK)	Meets 2NF?	Comments
Item Description	✗	✓	✗	It never needs to be ordered!
Item Qty	✓	✓	✓	There is no quantity without an order and it has to be a quantity of an item.
Item Price	✗	✓	✗	Every item that exists must have a price, but the item doesn't have to be sold.

We take out the fields that fail 2NF and create a separate table for them. This new table will be called Items. The Item Qty field stays in the Order Items table (it moved in the previous step). It is now 2NF compliant because there is now a many-to-one relationship between the Order Items and Items tables (in relation to the Invoice table, it has a one-to-many relationship so it had to move).

Table 1.8, Table 1.9 and Table 1.10 are the final versions of the tables for 2NF.

TABLE 1.8 How the tables look at the completion of 2NF: Orders table (no change)

Invoice No. (PK)	Date	CustomerID	Name	Address	Suburb	Post Code
X239	10-Jan-16	123	Widgets Inc	35 Colins Street	Melbourne	3000
X240	11-Jan-16	388	Poplets Pty Ltd	13a Mavis Road	Hawthorn	3122
X241	11-Jan-16	50	Gears and Sprockets	Lot 2 Murray Road	Preston	3072

TABLE 1.9 How the tables look at the completion of 2NF: Items table

Item ID (PK)	Item Description	Item Price (\$)
492	Bucket – 9L	2.95
943	Rivets – bag of 200	9.99
1523	M10 Screw – 12 cm	1.25
2884	Drill bits – assorted	2.59
6673	Shovels – Long handled	25.95
7621	Hammer – 2lb	39.95

TABLE 1.10 How the tables look at the completion of 2NF: Order Items table

Invoice No. (PK)	Item ID (PK)	Item Qty
X239	492	4
X239	1523	80
X239	6673	5
X239	7621	2
X240	943	2
X240	6673	1
X240	7621	3
X241	1523	120
X241	2884	8

Third normal form (3NF)

This is where a table is in 2NF and any column that is not part of the primary key is dependent only on the primary key and no other column. Therefore, to be 3NF compliant, every field in a table must relate directly to the PK. In our Ordering example, having established that the fields in the `Items` and `Order Items` tables are fully dependent on their own respective PKs, we come to the fields related to the customer details in the `Orders` table. We know that an order date cannot exist without an `Invoice No.`, so that will be left in its existing table. But customers and their own details can exist separately to an `Invoice No.` A customer can exist, but does not have to make an order, therefore the customer details fail 3NF because they are not dependent on the PK (`Invoice No.`) in their existing table. They can, however, be put in their own table because all those fields relate directly to the `CustomerID` field as their PK. If we split the `Orders` table to create a `Customers` table, the resulting tables will look like Tables 1.11 and 1.12.

TABLE 1.11 Orders table

Invoice No. (PK)	Date
X239	10-Jan-16
X240	11-Jan-16
X241	11-Jan-16

TABLE 1.12 Customers table

CustomerID (PK)	Name	Address	Suburb	Post Code
123	Widgets Inc	35 Colins Street	Melbourne	3000
388	Poplets Pty Ltd	13a Mavis Road	Hawthorn	3122
50	Gears and Sprockets	Lot 2 Murray Road	Preston	3072

In the `Orders` and newly created `Customers` tables after the first stage of 3NF, our tables are now taking shape, but there is a problem. By splitting them we have also removed their relationship with one another because an order needs a customer connected to it, even though customers do not need to have any orders. To restore this relationship we will place a foreign key (FK) in the `Orders` table. The purpose of the FK is to point to a PK in another table. In this case the field concerned is `CustomerID`. Tables 1.13–1.16 show how our completed tables look once this final step in 3NF is complete. You may also notice that the `Invoice No.` and `ItemID` fields are now also listed as FKs, but the database will concatenate them to form a PK for the `Order Items` field. This is a small refinement, but explains why we have not needed to create a specific PK for the `Order Items` table.

Table 1.13, Table 1.14, Table 1.15 and Table 1.16 are final version of our tables. They are now fully compliant to 3NF.

TABLE 1.13 Orders table

Invoice No. (PK)	Date	CustomerID (FK)
X239	10-Jan-16	123
X240	11-Jan-16	388
X241	11-Jan-16	50

TABLE 1.14 Customers table

CustomerID (PK)	Name	Address	Suburb	Post Code
50	Gears and Sprockets	Lot 2 Murray Road	Preston	3072
123	Widgets Inc	35 Colins Street	Melbourne	3000
388	Poplets Pty Ltd	13a Mavis Road	Hawthorn	3122

TABLE 1.15 Items table

Item ID (PK)	Item Description	Item Price (\$)
492	Bucket – 9 L	2.95
943	Rivets – bag of 200	9.99
1523	M10 Screw – 12 cm	1.25
2884	Drill bits – assorted	2.59
6673	Shovels – Long handled	25.95
7621	Hammer – 2 lb	39.95

TABLE 1.16 Order Items table

These become a concatenated PK		
Invoice No. (FK)	Item ID (FK)	Item Qty
X239	492	4
X239	1523	80
X239	6673	5
X239	7621	2
X240	943	2
X240	6673	1
X240	7621	3
X241	1523	120
X241	2884	8

CASE STUDY

PIZZAS ON THE GO – PART 3

Matt has supplied Nick with the existing spreadsheet he uses to record telephone orders. (Figure 1.42)

Nick takes this data and looks for any repetition of fields to achieve 1NF. Then he looks at the table to see which fields are wholly dependent on the primary keys (OrderID, PizzaID). He creates new tables for fields that depend on their own key to make them compliant with 2NF. Finally, to move the database into 3NF, Nick looks for any fields in the 'Pizza Orders' table that do not depend on its PK. The resulting tables Nick creates at the end of this process are shown in Tables 1.17–1.20.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
Order ID	Order Date	Order Time	Customer ID	Given Name	Family Name	Address	Suburb	Post Code	Mobile	Email	Club Member?	Pizza ID	Pizza Name	Crust type	Qty	Price (\$)
00548	26-Feb-16	1845	2669	Jacqueline	Mercer	55 [REDACTED] Road	McKenzie Creek	3401	0421 974 2 [REDACTED]	[REDACTED]@fastermail.net.au	TRUE	HAW	Hawaiian	D	4	4.95
												MAG	Margherita	D	4	4.95
												VEG	Vegetarian	G	2	7.95
												PEP	Pepperoni	D	1	4.95
12345	20-Jan-16	1823	7913	Thaddeus	Bowen	24 [REDACTED] Avenue	Horsham	3400	0404 109 9 [REDACTED]	[REDACTED]@tafe.edu.au	TRUE	SUS	Super Supreme	D	2	10.95
												CHI	Chicken	T	1	7.95
12346	11-Feb-16	1801	0479	Isabelle	Lynch	13 [REDACTED] Avenue	Horsham	3400	0450 544 4 [REDACTED]	[REDACTED]@yahoo.com	TRUE	MAG	Margherita	D	1	4.95

FIGURE 1.42 The table of order transactions that Matt supplies to Nick**TABLE 1.17** tblPizzaOrders

Order ID (PK)	Order Date	Order Time	Customer ID (FK)
00548	26-Feb-16	1845	2669
12345	20-Jan-16	1823	7913
12346	11-Feb-16	1801	0479

TABLE 1.18 tblCustomers

Customer ID (PK)	Given Name	Family Name	Address	Suburb	Post Code	Mobile	Email	Club Member?
0479	Isabelle	Lynch	13 [REDACTED] Avenue	Horsham	3400	0450 544 4 [REDACTED]	[REDACTED]@yahoo.com	TRUE
2669	Jacqueline	Mercer	35 [REDACTED] Road	McKenzie Creek	3401	0421 974 2 [REDACTED]	[REDACTED]@fastermail.net.au	TRUE
7913	Thaddeus	Bowen	24 [REDACTED] Avenue	Horsham	3400	0404 109 9 [REDACTED]	[REDACTED]@tafe.edu.au	TRUE

Case study continues ►

◀ Case study continued

TABLE 1.19 tblPizzas

Pizza ID (PK)	Pizza Name	Price (\$)
CHI	Chicken	7.95
HAW	Hawaiian	4.95
MAG	Margherita	4.95
PEP	Pepperoni	4.95
SUS	Super Supreme	10.95
VEG	Vegetarian	7.95

TABLE 1.20 tblOrderItems

Order ID (FK)	Pizza ID (FK)	Crust type	Qty
00548	HAW	D	4
00548	MAG	D	4
00548	VEG	G	2
00548	PEP	D	1
12345	SUS	D	2
12345	CHI	T	1
12346	MAG	D	1

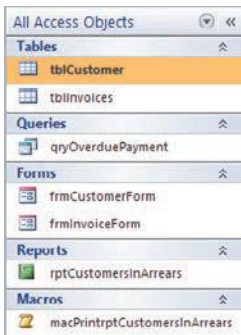


FIGURE 1.43
Examples of named
objects within a
Microsoft Access
database

In Microsoft Access, object names can be up to 64 characters in length, but they cannot begin with leading spaces. They can include any combination of letters, numbers, spaces and special characters except a period (.), an exclamation point (!), an accent grave (`), or square brackets ([]).

Database design tools

Database structure and appearance can be planned using a variety of common planning tools. Data dictionaries, data structure diagrams, flowcharts, **input–process–output (IPO) charts** and query diagrams are useful for representing structure. Relationships between database objects can be represented by ERDs and structure charts. Layout diagrams and mockups can be used to plan the appearance of input forms and reports.

Several of these design tools will be further explained below and in later chapters.

Naming conventions

It is important that every field and object within a database is named properly so that it can easily be identified by a developer or a user. For example, you might prefix tables with **tbl** (for example, **tblCustomer**, **tblProducts**) and fields within the Customer table might be prefixed with **cus** (for example, **cusCustomerID**, **cusAddress**). Queries might be prefixed with **qry**, forms with **frm** and reports with **rpt**. Users will interact with forms and reports often, so ensure that the rest of the name is descriptive (for example, **frmInputCustomerDetails** or **rptOverdueInvoices**). Avoid spaces and underscores in object names. To aid readability, a mix of uppercase and lowercase characters is used in each name. In most databases, a user interacts with a menu system with properly labelled buttons, rather than the individual objects in the main database screen.

Data dictionary

Previously, you will have worked out RDBMS table structures through normalisation. To formalise the fields and their data type, format and size for a new database a common design tool to use is a **data dictionary**. This gives a developer a precise explanation of how each field should be set up in each table. It will also provide details of built-in validation. Note that database tables cannot hold formulas. Calculations can only be performed in objects such

as queries, forms and reports. See Tables 1.21–1.24 for examples of data dictionaries and the key headings to use. You can create data dictionaries entirely by hand or construct a template electronically, then fill out the specific details either by hand or electronically.

TABLE 1.21 The `tblCustomers` table for POTG

Field	Data type	Field size	Input mask	Caption	Description	Validation rule	Validation text
<code>cusCustomerID</code> (PK)	Text	4	----	Customer ID	Customer's individual assigned code	Between 200 and 9999	Customer ID number must be between 0200 and 9999 inclusive
<code>cusFamilyName</code>	Text	20		Family Name	Customer's family name		
<code>cusGivenName</code>	Text	20		Given Name	Customer's given name		
<code>cusDelAddress</code>	Text	40		Address	Customer's delivery address		
<code>cusDelSuburb</code>	Text	20		Suburb	Customer's delivery suburb		
<code>cusDelPostCode</code>	Text	4	----	Post Code	Customer's delivery postcode	Between 3400 and 3401	Values must be between 3400 and 3401 inclusive
<code>cusMobile</code>	Text	12	9999 999 999	Mobile	Customer's mobile number		
<code>cusEmail</code>	Text	40		Email	Customer's email address		
<code>cusClub</code>	Boolean	1		Club Member?	Is customer a member of the Pizza Club?		

TABLE 1.22 The `tblPizzas` table for POTG

Field	Data type	Field size	Input mask	Caption	Description	Validation rule	Validation text
<code>pizPizzaID</code> (PK)	Text	3	-----	Pizza ID		Between 'AAA' and 'ZZZ'	All Pizza IDs are alphabetical between AAA and ZZZ inclusive
<code>pizPizzaName</code>	Text	20		Pizza Name			
<code>pizPizzaPrice</code>	Numeric	99.99		Price (\$)		Between 4.95 and 10.95	All pizza prices are between \$4.95 and \$10.95 inclusive

TABLE 1.23 The `tblPizzaOrders` table for POTG

Field	Data type	Field size	Input mask	Caption	Description	Validation rule	Validation text
ordOrderID (PK)	Text	5	-----	Order ID			
ordCustomerID (FK)	Text	4		Customer ID			
ordOrderDate	Date	Medium Date	dd- MMM- yy	Order Date			
ordOrdTime	Text	4		Order Time	Time of order in 24hr time		

TABLE 1.24 The `tblOrderItems` table for POTG

Field	Data type	Field size	Input mask	Caption	Description	Validation rule	Validation text
oitOrderID (FK)	Text	5	-----	Order ID	Order ID		
oitPizzaID (FK)	Text	3		Pizza ID	Pizza ID		
oitCrust	Text	1		Crust type	D = deep, T = thin, G = gluten free	Like "D" or "T" or "G"	Must be D (deep), T (thin) or G (gluten free)
oitQty	Numeric	2		Qty		Between 1 and 10	Between 1 and 10 (inclusive) pizzas can be ordered

CASE STUDY**PIZZAS ON THE GO – PART 4****THINK ABOUT INFORMATICS 1.6**

What other numeric data types are allowed in numeric database fields? When might each one be used?

Nick is happy with his table normalisation process and so he has moved on to work out the plans for each table by way of data dictionaries. He takes the fields and tables from the normalisation process, then works out the details for each of those fields, using the sample data he was working with earlier. The data dictionaries for the four tables he identified are shown in Tables 1.21–1.24.

Data structure diagram

The purpose of a data structure diagram (DSD) is to indicate the relationships that exist between the specific tables of a planned database so that the developer will know how the tables ought to be linked (Figure 1.44). It is important to indicate the type of relationship in the diagram. Again, the ERD and table normalisation procedure you carried out earlier will assist you in determining how to connect each table. The DSD can be done electronically, but it is often faster and easier to sketch it by hand.

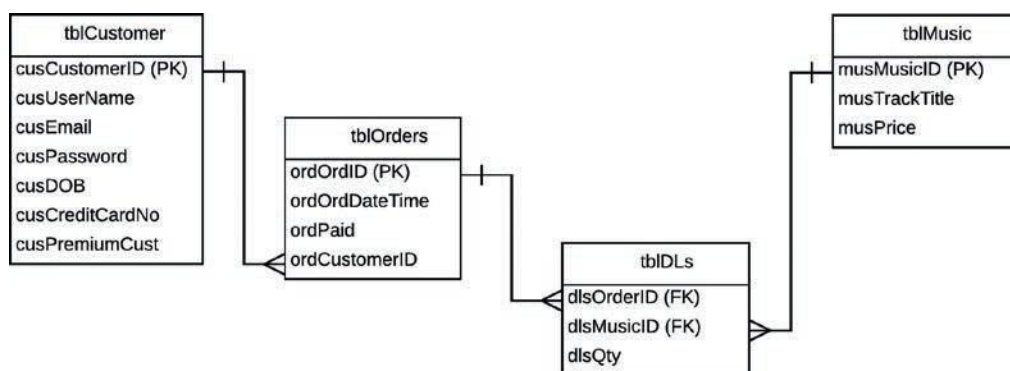


FIGURE 1.44 An example of a data structure diagram with one-to-many relationships. This DSD represents the database table relationships shown in Figure 1.45.

CASE STUDY

PIZZAS ON THE GO – PART 5

Nick uses the data dictionaries and ERD to determine how the tables in the RDBMS will connect with each other and constructs the DSD in Figure 1.45.

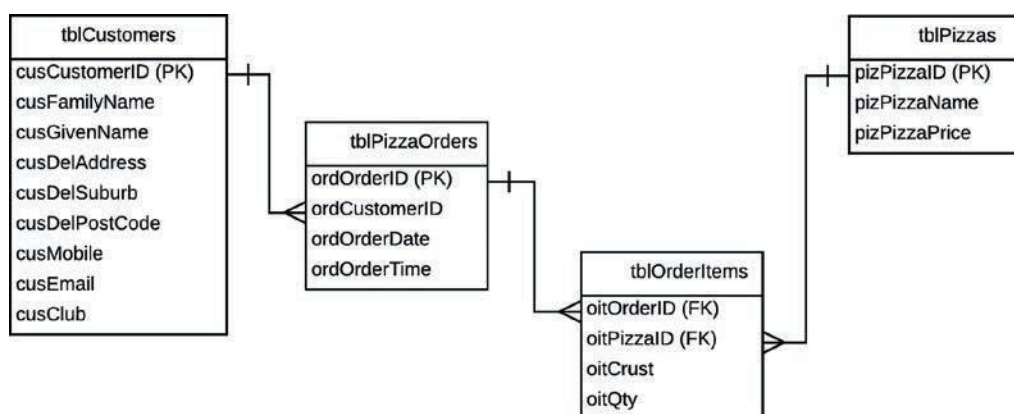


FIGURE 1.45 Nick's DSD for POTG

A field size set to 'integer' allows only whole numbers to be entered. If the field were to include decimal numbers, the field size could be set to 'single', which would allow up to seven decimal places. Integer field size requires less storage space (2 bytes) than a number specified as single (4 bytes).

The query criteria can use <, >, =, <=, >= and <> symbols when dealing with numbers or text. Alternatively, plain language such as 'Between 0 and 20000' or 'Less than 500' can be used. The * and ? symbols are also valid wildcards. Text used in query criteria must be placed inside single inverted commas. Other expressions such as 'is null' or 'is not null' can be used to show records where a field is either empty or contains some data. A query can designate a date by placing # marks at each end of the date. If more than one field has a query criterion assigned to it, the criteria are applied in an 'and' relationship. That is, both the criteria must be met before data is returned by the query.

Query design

Queries were introduced earlier in this chapter. They are used in databases to select only those records that meet the query criteria. They can also be used as a control structure to hold formulas used on forms and on other queries. The result of a query is a list of selected records organised by specified fields. This list is then used as the basis for a printed or electronic report.

The design of a query should specify the fields to be included and the tables to which they belong. The design will also include the criteria for the query and how the resulting records are to be ordered. The query design is usually hand-drawn, although you may set up a basic table as a template. In your solution, you should use at least two criteria and both a primary and a secondary sort (sorting is explained in 'Setting up the queries' later in this chapter). Also, be aware that not all fields used for selecting data need to be shown in the final list.

If formulas are created in the query, you should also ensure that they have the correct data type (e.g. text or numeric) and a proper name applied.

TABLE 1.25 A query design for selecting all the sales of vegetarian pizzas since the start of February. It is sorted by `Crust type` (primary sort key), then by `Order Date` (secondary sort key) and `Order Time` (tertiary sort key). There is a formula to calculate the total price for each transaction. All the pizzas are vegetarian so the `PizzaName` field will not be displayed in the final list because the data would be redundant. Note that this query uses fields drawn from more than one table.

Field	oitCrust	oitQty	tesTotal Cost: [oitQty] *[pizPizzaPrice]	ordOrder ID	ordCustomer ID	ordOrder Date	ordOrder Time	pizPizza Name
Table	tblOrder Items	tblOrder Items		tblPizza Orders	tblPizza Orders	tblPizza Orders	tblPizza Orders	tblPizzas
Sort	Ascending					Ascending	Ascending	
Show	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Criteria						>=#1/02/20 16#		'Vegetarian'
Query name: qryVegetarianSoldByCrust Sources: tblOrderItems, tblPizzaOrders, tblPizzas								

TABLE 1.26 A query design for selecting all the sales of pizzas to a specific customer (ID 7913) during the month of February. It is sorted by `Order Date` (primary sort key) and then by `Order Time` (secondary sort key). There is a formula to calculate the total price for each transaction. All the transactions are from the one customer, so the `CustomerID` field will not be displayed in the final list because the data would be redundant.

Field	ordOrder Date	ordOrder Time	oitOrder ID	pizPizza Name	pizPizza Price	oitCrust	oitQty	tesTotal Cost: [oitQty] *[piz Pizza Price]	cusCusto merID
Table	tblPizza Orders	tblPizza Orders	tblOrder Items	tblPizzas	tblPizzas	tblOrder Items	tblOrder Items		tblCusto mers
Sort	Ascending	Ascending							
Show	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Criteria	Between #1/02/2016 # and #28/02/2016#								'7913'
Query name: qryBowenOrdersInFebruary Sources: tblOrderItems, tblPizzaOrders, tblPizzas, tblCustomers									

CASE STUDY

PIZZAS ON THE GO – PART 6

Nick has designed two queries for POTG. One of the queries will select all of the `Vegetarian` pizzas sold since a new gluten free crust was introduced at the start of February and sort them primarily by `Crust Type`. This is to help Matt determine which crust is the most popular. Matt is only interested in the sales, so there is no customer data linked to the query. The other query will select all of the transactions for a particular customer, in this case ID 7913 – Thaddeus Bowen, for the month of February. This query will be sorted by `Order Date` and `Order Time`. Its purpose is for Matt to review the transactions of any customer and ultimately to allow the customers to check their own ordering history.

Test data

A set of test data that will be used during development to ensure that the solution is functioning correctly is prepared in the **design stage**. The test data should be chosen to test all aspects of the solution, including identification and handling of unreasonable, missing or incorrect data (validation), formulas and summary totals like 'Counts' and 'Sums'. Once the solution has been shown to be functioning correctly, the test data is removed and the 'real' data relevant to the organisation is added. A **test plan** (developed during the design stage) is used to show all the functions to be tested, the sample test data and how the function is expected to handle the data. An example of a section from a test plan is shown in Table 1.27. The actual testing process is covered in detail later in this chapter.

TABLE 1.27 An example of a test plan for a range of common database elements

Formula/function to test	Location	Test data	Expected/required result	Why?
Validation on field oitQty (Existence check)	tblOrderItems table	Null value	Null value: An error message about missing data must be displayed.	There is no value entered and the field requires one.
		1	1: No error message will be displayed.	There is data in the field, as it requires.
Validation on field oitQty (Data Type check)	tblOrderItems table	1	1: No error message will be displayed.	The field requires a numeric value to be entered and 1 is a numeric value.
		A	A: An error message will be displayed to indicate that the data type is incorrect because the field requires a numeric value and 'A' is not numeric.	The field requires a numeric value to be entered, but A is a non-numeric value.
Validation on field oitQty (Range check)	tblOrderItems table	Order Qty: 0 1 2 9 10 11	1 2 9 10 } No message will be displayed.	All values are within the range of 1 to 10 inclusive.
			0 11 } The message 'Between 1 and 10 (inclusive) pizzas can be ordered.' will be displayed.	Both values are outside the range of 1 to 10 inclusive.
Formula: tesTotalCost: [oitQty] * [pizPizzaPrice]	qryBowenOrders InFebruary query	Pepperoni = 4.95 Qty = 2	4.95 * 2 = 9.90	Worked out on a calculator.
Selection: ordOrderDate is >=#1/02/2016# AND pizPizzaName is 'Vegetarian'	qryVegetarianSold ByCrust query	30 records in database	5 records will be displayed.	Because only 5 records match both the date and pizza name criteria.
Primary sort on oitCrust field. (Ascending)	qryVegetarianSold ByCrust query	Crust Type: D G T	The primary sort order will be: D G T	This is because it follows alphabetical order.

TABLE 1.27 (continued)

Formula/function to test	Location	Test data	Expected/required result	Why?
Secondary sort on ordOrderDate field (Ascending)	qryVegetarianSold ByCrust query	Order Date: 8-Feb-16 11-Feb-16 19-Feb-16 26-Feb-16	The secondary sort order will be: 8-Feb-16 11-Feb-16 19-Feb-16 26-Feb-16	This is because it follows calendar dating.
Function: =Count ([ordOrderDate]) in 'Totals' footer	qryVegetarianSold ByCrust query	5 records are selected when the query is run.	The count should be 5	Because there are 5 records selected when the query is run.
Function: =Sum ([tesTotalCost]) in 'Totals' footer	qryVegetarianSold ByCrust query	Order ID: 52311 – 23.85 40991 – 15.90 12347 – 23.85 84779 – 15.90 00548 – 15.90	23.85 + 15.90 + 23.85 + 15.90 + 15.90 = 95.40	Worked out on a calculator.

Manual validation techniques are important, but not foolproof. Some data can get past us when we proofread because it may look correct, but will fail when electronic validation techniques are applied. For example, the date 31 June seems reasonable to many of us and would get by many proofreaders. A field, however, set to display only 'date' type data would reject 31 June as an impossible date.

Ensuring that data fits between two predetermined values is called range checking.

An **IF** statement can be used to set the upper and lower limits of valid data for a particular field in a database. **IF** stands for Inline IF. The **IF** statement can be prepared to display an error message if data falls outside the allowed range, or proceed with calculations if the data is valid.

CASE STUDY

PIZZAS ON THE GO – PART 7

Nick writes up a test plan, in the form of a table, so that when he builds the database, he knows what to look for when actually carrying out the testing in the **development stage**. He has put a sample of the testing plan in Table 1.27.

Validation techniques

Although data is manually validated prior to input, it is still necessary to use electronic validation techniques after the data has been input. Validation is the process by which data that has been input to a computer is assessed to ensure that it is correct and complete. Sometimes errors occur during input, resulting in inaccurate or unreasonable data. Both manual and electronic validation techniques are necessary to ensure that data is free of errors.

Manual validation requires the data entry operator to proofread the data they have entered to ensure its accuracy, completeness and reasonableness before inputting it into the system. Electronic validation relies on software functions to perform checks on accuracy, completeness and reasonableness. Available functions in software include range checks, spell checking, grammar checking, predefined lists, data type checks, **input masks**, alignment and **IF** statements. The key validation checks you need to be able to carry out are existence (is a value actually entered in the field?), data type checking (is the data the correct data type for the field?) and range checking (does the entered value fit within the range of values we expect?).

Databases such as Microsoft Access allow **validation rules** to be established within the tables. A validation rule allows data to be entered into a field only if it obeys the given rule

(Figure 1.46). If there is no validation rule, and the data is the correct format specified for the field, it is entered directly into the table. If a validation rule has been set up, the data is checked against the rule and only accepted if it obeys the rule. If the data is rejected, a message appears on the screen (Figure 1.47). Validation rules should be added to tables prior to the entry of any data in order to avoid data integrity issues.

tblCustomers		
Field Name	Data Type	
cusCustomerID	Text	Customer's individual assigned code
cusFamilyName	Text	Customer's family name
cusGivenName	Text	Customer's given name
cusDelAddress	Text	Customer's delivery address
cusDelSuburb	Text	Customer's delivery suburb
cusDelPostCode	Text	Customer's delivery postcode
cusMobile	Text	Customer's mobile number
cusEmail	Text	Customer's email address
cusClub	Yes/No	Is customer a member of the Pizza Club?
Field Properties		
General		
Field Size	4	
Format		
Input Mask		
Caption	Customer ID	
Default Value		
Validation Rule	Between 200 And 9999	
Validation Text	Customer ID number must be between 200 and 9999 inclusive.	

FIGURE 1.46 A table design screen with a validation rule applied to one of the fields



FIGURE 1.47 The validation error message when incorrect data is entered

An input mask can also be used to reduce the chance that invalid data is entered. An input mask can be set for any field that holds text or date-type data and controls how the data can be entered. Display characters are used to indicate where data may be entered. A '9' set as an input mask allows any single-digit number to be entered, and an 'L' allows any single-letter (character) to be entered.

Using a list tool on a form to direct data entry into a field where there are limited possible responses ensures that only valid data is added. For example, if the product used to replace a deteriorated roof comes in only four colours, a list box can be set up that contains only the allowed choices. This forces the user to make a selection from the list, thus eliminating invalid data or incorrect spelling.



FIGURE 1.48 An example of an `IIF()` statement **a** in Design View and **b** the resulting message. Note that this `IIF()` statement is being used to flag data in a field, rather than to validate it.

An error message should indicate what is wrong with the data and how to correct it, rather than simply stating that there is an error. For example, the error message 'Age must be between 10 and 70 years of age' is far more useful than 'Age is incorrect'.

An input mask that uses an 'L' to force the entry of a character can be useful in situations such as the entry of a student's form in a school database. For example, an input mask of 99L set for a form field would allow 11D, 10A, 12B or even 09F to be entered. Alternatively, if only Year 12 forms are to be entered, a mask of '12'L would only permit entries that begin with 12.

Developing a RDBMS

Once the database designs are complete, they are passed on to a developer who creates the actual database, although in this task you will be performing both roles. The next section outlines this process.

Creating the tables and relationships

A developer creates a database file and then begins building each of the tables (Figures 1.47–1.50) from the data dictionaries. This is the start of the **development stage** of the PSM. The correct names and data types are assigned and validation rules applied. Then the developer will create the relationships as identified in the data structure diagram (Figure 1.51). They will need to ensure that referential integrity (presence of data in foreign keys) is applied. In GUI-based database programs, these processes are often simplified by drag-and-drop field placement and linking.

Refer to the Microsoft Access online help for more information about table relationships and referential integrity.

tblCustomers			
	Field Name	Data Type	Description
V	cusCustomerID	Text	Customer's individual assigned code
	cusFamilyName	Text	Customer's family name
	cusGivenName	Text	Customer's given name
	cusDelAddress	Text	Customer's delivery address
	cusDelSuburb	Text	Customer's delivery suburb
	cusDelPostCode	Text	Customer's delivery postcode
	cusMobile	Text	Customer's mobile number
	cusEmail	Text	Customer's email address
	cusClub	Yes/No	Is customer a member of the Pizza Club?
Field Properties			
General			
Field Size	4		
Format			
Input Mask			
Caption	Post Code		
Default Value			
Validation Rule	Between 3400 And 3401		
Validation Text	Post Codes must be between 3400 and 3401 inclusive		
Required	No		
Allow Zero Length	Yes		
Indexed	Yes (Duplicates OK)		
Unicode Compression	Yes		
IMF Mode	No Control		
IME Sentence Mode	None		
Smart Tags			

FIGURE 1.49 The Design View of the Customers table for POTG, with a validation rule selected

tblOrderItems			
	Field Name	Data Type	
	oitOrderID	Text	Order ID
	oitPizzaID	Text	Pizza ID
	oitCrust	Text	D = deep, T = thin, G = gluten free
	oitQty	Number	
Field Properties			
General			
Field Size	1		
Format			
Input Mask			
Caption	Crust type		
Default Value	"D"		
Validation Rule	Like "D" Or "T" Or "G"		
Validation Text	Must be D (deep), T (thin) or G (gluten free)		

FIGURE 1.50 The Design View of the Order Items table for POTG, with a validation rule for the pizza Crust Type selected. Note that there is no individual PK for this table because it has two FKs.

tblPizzaOrders		
Field Name	Data Type	
ordOrderID	Text	
ordCustomerID	Text	
ordOrderDate	Date/Time	
ordOrderTime	Text	Time of order in 24hr time
General Lookup		
Format	Medium Date	
Input Mask		
Caption	Order Date	

FIGURE 1.51 The Design View of the Pizza Orders table for POTG

tblPizzas		
Field Name	Data Type	
pizPizzaID	Text	
pizPizzaName	Text	
pizPizzaPrice	Number	
General Lookup		
Field Size	Double	
Format	Fixed	
Decimal Places	2	
Input Mask		
Caption	Price (\$)	
Default Value		
Validation Rule	Between 4.95 And 10.95	
Validation Text	All pizza prices are between \$4.95 and \$10.95 inclusive	

FIGURE 1.52 The Design View of the Pizza table for POTG, with a validation rule selected

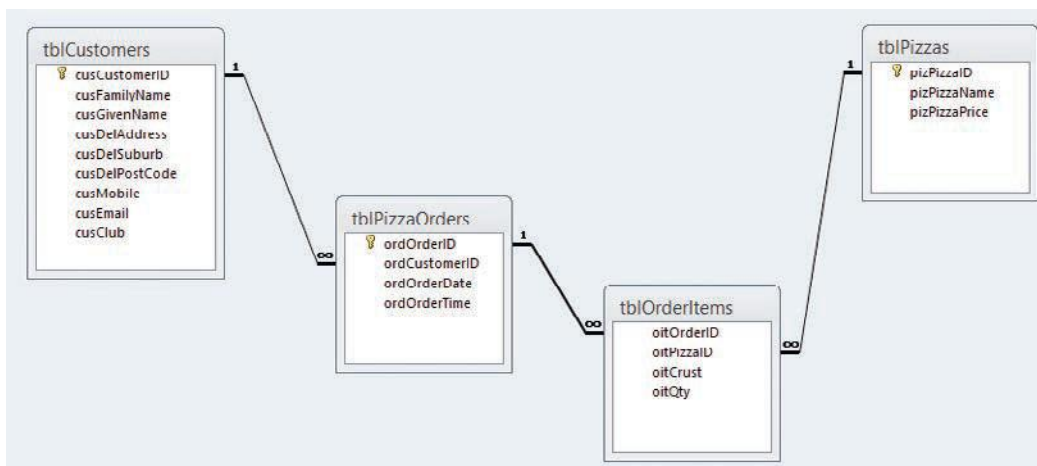


FIGURE 1.53 The relationships view showing how tables are linked within the POTG database

CASE STUDY

PIZZAS ON THE GO – PART 8

Nick has shown the designs of the database to Matt. Matt is happy with the work, so Nick starts the **development stage**. He creates each of the tables (Figures 1.49 to 1.52) and then creates the relationships between them (Figure 1.53). Figures 1.54 to 1.57 show each of the tables with data entered.

tblCustomers									
Customer ID	Family Name	Given Name	Address	Suburb	Post Code	Mobile	Email	Club Member?	
0479	Lynch	Isabelle	13 Avenue	Horsham	3400	0450 544 4	@yahoo.com	☒	
1932	Evans	Justine	7 Street	Horsham	3400	0439 052 6	@gmail.com	☐	
2201	Tran	Sebastian	Lot 1 Highway	Dooen	3401	0406 154 0	@gmail.com	☐	
2609	Mercer	Jacqueline	35 Road	McKenzie Cree	3401	0421 974 2	@fastermail.net.au	☒	
3006	Duncan	Whitney	1 Court	Haven	3401	0424 218 8	@telstra.org	☒	
4821	Hogan	Frances	4 Court	Haven	3401	0427 471 4	@iprimus.net.au	☒	
7913	Bowen	Thaddeus	24 Avenue	Horsham	3400	0404 109 9	@tafe.edu.au	☒	
8345	Wade	Zachery	833 Highway	Dooen	3401	0451 169 7	@westnet.com	☐	
8860	Shepherd	Brenden	47 Road	Bungalally	3401	0426 376 7	@gmail.com	☒	

FIGURE 1.54 The Datasheet View of the Customers table for POTG with data entered

◀ Case study continued

tblOrderItems			
Order ID	Pizza ID	Crust type	Qty
00548	HAW	D	4
00548	MAG	D	4
00548	VEG	G	2
00548	PEP	D	1
12345	SUS	D	2
12345	CHI	T	1
12346	MAG	D	1
12347	MAI	D	5
12347	BBQ	D	2
12347	VEG	G	3
12348	AUS	D	2
12348	AUS	T	1
12348	HAW	T	3
12348	PEP	D	4
22698	MAG	D	2
32216	AUS	D	1
32216	BBQ	G	1
32216	SUS	T	1
40991	CHI	D	2
40991	VEG	G	2
40991	PEP	T	2
52311	VEG	D	3
52311	BBQ	T	2
52311	MAG	D	1
59420	SUS	D	1

FIGURE 1.55 A section of the Datasheet View of the Order Items table for POTG, with data entered. Note how the two FKs make up a PK.

tblPizzaOrders			
Order ID	Customer ID	Order Date	Order Time
00548	2669	26-Feb-16	1845
Pizza ID	Crust type	Qty	
HAW	D	4	
MAG	D	4	
VEG	G	2	
PEP	D	1	
*	D		
12345	7913	20-Jan-16	1823
12346	0479	11-Feb-16	1801
12347	1932	11-Feb-16	1234
12348	8860	12-Feb-16	1915
22698	0479	15-Feb-16	2003
32216	8345	24-Feb-16	1758
40991	7913	08-Feb-16	1901
52311	0479	05-Feb-16	1835
59420	2201	03-Mar-16	1632
83245	7913	18-Feb-16	1845
84779	3006	19-Feb-16	1812
97645	7913	18-Feb-16	1924

FIGURE 1.56 The Datasheet View of the Pizza Orders table for POTG. One of the records has been expanded so you can see how it is linked to the underlying Order Items table.

tblPizzas			
Pizza ID	Pizza Name	Price (\$)	
AUS	Aussie	4.95	
BBQ	BBQ Meatlovers	10.95	
CHI	Chicken	7.95	
HAW	Hawaiian	4.95	
MAG	Margherita	4.95	
MAI	Marinara	10.95	
PEP	Pepperoni	4.95	
SUS	Super Supreme	10.95	
VEG	Vegetarian	7.95	

FIGURE 1.57 The Datasheet View of the Pizza table for POTG

Setting up the queries

Once the tables in a database have been set up and linked, the queries are created. Because most queries are based on more than one table, the tables need to have been added to the query and already joined with one another (as explained above). In the query, the tables will automatically show the links between the field lists.

It is important to follow the query plan closely because misplacing filtering criteria will result in inaccurate information being produced.

The left-most field in the query Design View that is assigned a sort specification is the primary sort key. The field next most left is the secondary sort key, and so on. The fields can be dragged to different positions to change the sort order.

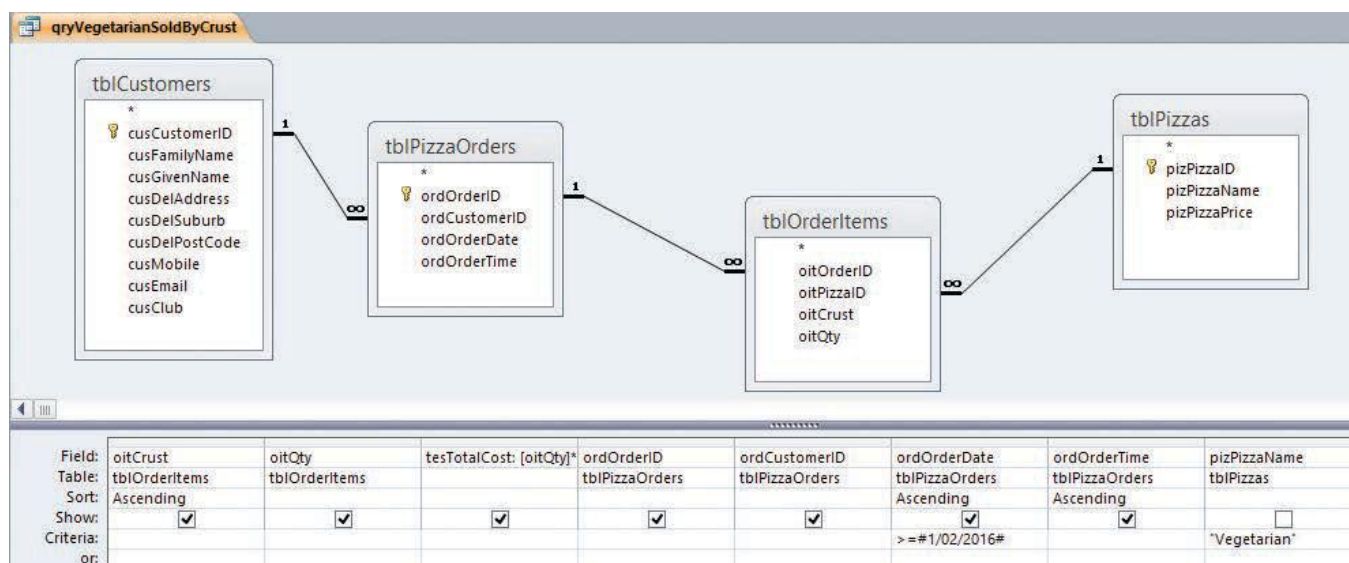


FIGURE 1.58 This query will select all vegetarian pizzas sold since the start of February and sort them by crust type, then order date and order time

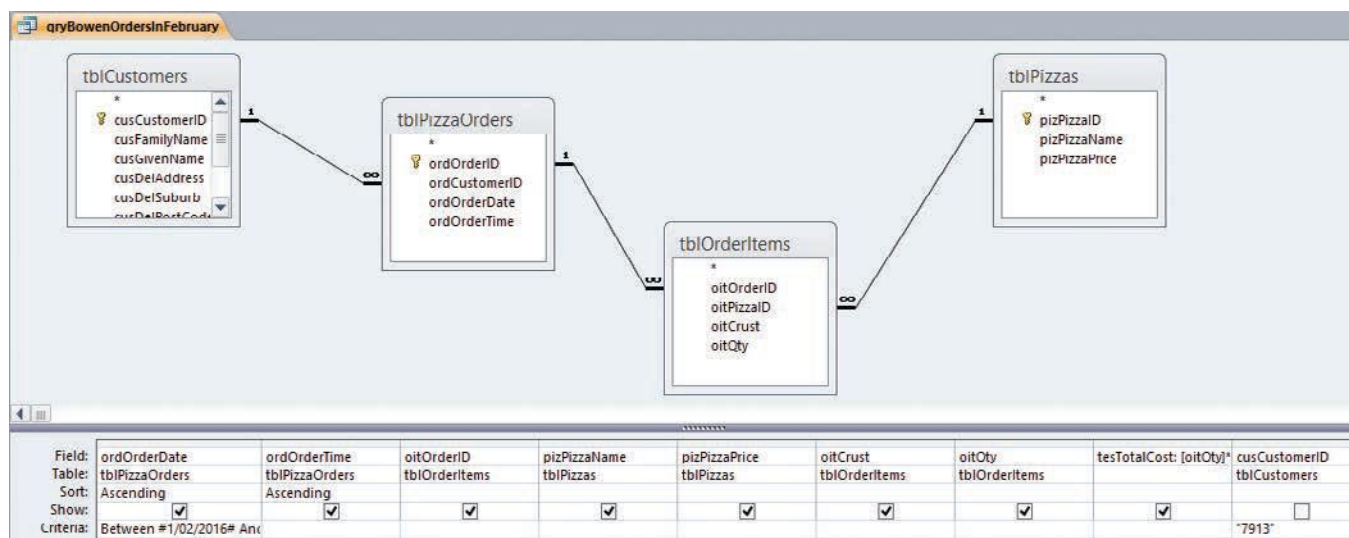


FIGURE 1.59 This query will select all pizzas sold during February to customerID 7913 and sort them by order date and order time

```

qryVegetarianSoldByCrust
SELECT tblOrderItems.oiCrust, tblOrderItems.oiQty, [oiQty]*[piPizzaPrice] AS tesTotalCost, tblPizzaOrders.ordOrderID, tblPizzaOrders.ordCustomerID, tblPizzaOrders.ordOrderDate, tblPizzaOrders.ordOrderTime
FROM tblPizzas INNER JOIN ((tblCustomers INNER JOIN tblPizzaOrders ON tblCustomers.cusCustomerID = tblPizzaOrders.ordCustomerID) INNER JOIN tblOrderItems ON tblPizzaOrders.ordOrderID =
tblOrderItems.oiOrderID) ON tblPizzas.pizPizzaID = tblOrderItems.oiPizzaID
WHERE (((tblPizzaOrders.ordOrderDate) >=#2/1/2016#) AND ((tblPizzas.pizPizzaName)='Vegetarian'))
ORDER BY tblOrderItems.oiCrust, tblPizzaOrders.ordOrderDate, tblPizzaOrders.ordOrderTime;

```

FIGURE 1.60 The SQL view of the vegetarian pizzas query shown in Figure 1.58

```

qryBowenOrdersInFebruary
SELECT tblPizzaOrders.ordOrderDate, tblPizzaOrders.ordOrderTime, tblOrderItems.oiOrderID, tblPizzas.pizPizzaName, tblOrderItems.oiCrust, tblPizzas.pizPizzaPrice, tblOrderItems.oiQty, [oiQty]*[piPizzaPrice] AS
tesTotalCost
FROM tblPizzas INNER JOIN ((tblCustomers INNER JOIN tblPizzaOrders ON tblCustomers.cusCustomerID = tblPizzaOrders.ordCustomerID) INNER JOIN tblOrderItems ON tblPizzaOrders.ordOrderID =
tblOrderItems.oiOrderID) ON tblPizzas.pizPizzaID = tblOrderItems.oiPizzaID
WHERE (((tblPizzaOrders.ordOrderDate) Between #2/1/2016# And #2/28/2016#) AND ((tblCustomers.cusCustomerID)='7913'))
ORDER BY tblPizzaOrders.ordOrderDate, tblPizzaOrders.ordOrderTime;

```

FIGURE 1.61 The SQL view of the customerID 7913 query from Figure 1.59

CASE STUDY

PIZZAS ON THE GO – PART 9

Using his designs, Nick creates the two queries for POTG. He drags and drops the fields he needs from different tables and creates formulas to calculate totals. He then puts in selection criteria. The Design Views for each query are shown in Figures 1.58 and 1.59. The SQL views are shown in Figures 1.60 and 1.62. The data selected by each query can be seen in Figures 1.62 and 1.63.

qryVegetarianSoldByCrust						
Crust type	Qty	Total Cost (\$)	Order ID	Customer ID	Order Date	Order Time
D	3	23.85	52311	0479	05-Feb-16	1835
G	2	15.90	40991	7913	08-Feb-16	1901
G	3	23.85	12347	1932	11-Feb-16	1234
G	2	15.90	84779	3006	19-Feb-16	1812
G	2	15.90	00548	2669	26-Feb-16	1845
Total	12	95.40				5

FIGURE 1.62 The result of running the vegetarian pizzas query

qryBowenOrdersInFebruary							
Order Date	Order Time	Order ID	Pizza Name	Crust type	Price (\$)	Qty	Total Cost (\$)
08-Feb-16	1901	40991	Pepperoni	T	4.95	2	9.90
08-Feb-16	1901	40991	Vegetarian	G	7.95	2	15.90
08-Feb-16	1901	40991	Chicken	D	7.95	2	15.90
18-Feb-16	1845	83245	Pepperoni	G	4.95	1	4.95
18-Feb-16	1924	97645	Marinara	G	10.95	1	10.95
Total		5				8	57.60

FIGURE 1.63 The result of running the customerID 7913 ordering history query

Testing the solution

Testing is a necessary step in problem solving. Testing the product is done to ensure that all features and functions work correctly and that the information produced is free of errors.

For a large project, testing the solution is conducted both formally and informally.

- Informal testing occurs during the **development stage**. As software developers create a solution, they are constantly testing to see if the solution behaves according to expectations.
- Formal testing involves testing the overall solution after it has been completed to see if it solves the problem. This type of testing is formalised with comprehensive documentation of each test and each test outcome.

A **bench test** forms part of the formal testing process in which sample data that has been developed during the **design stage** is used to determine how the solution behaves with a range of data. Data is chosen to see whether calculations are performed correctly and how erroneous data is handled by validation techniques. For each piece of sample test data, the expected outcome is determined and compared with the observed response of the solution. Table 1.27 showed an example of a test plan created during the design stage. This test plan is now used to allow testers to check each part of the solution to ensure its proper operation. The results of testing are normally written by hand on printouts from the solution.

As well as testing that the formulas and functions work in the solution, you must ensure that the solution meets the actual needs of the intended users. This process is called **user acceptance testing** and usually involves asking users to follow a series of steps to complete a task in the solution or interpreting the information contained in the output, then providing feedback to the developers.

The testing procedure involves six steps. The first three steps will have been built into the test plan during the design stage.

Testing procedure

- 1 Establish the tests that will be conducted: State the formula or function to be tested.
- 2 Determine the test data that will be used: Specify the test data.
- 3 Determine the expected results: Specify what the expected or required result is and explain why you expect this result.
- 4 Record the actual result that is generated by the solution.
- 5 Correct any identified errors.
- 6 Provide a concluding statement about how the formula or function works correctly.

Features to test

- Tables: Validation, input masks.
- Query: Selection criteria, sorting, formulas, summary statistics such as totals and counts.
- User acceptance: Following a set of procedures (functionality), interpreting the output (presentation and clarity of message).

Testing table validation

The first key feature to test in a database is table validation. Testing validation first ensures that the data is accurate before it is entered via forms or queries are run. It is important in the testing process to choose data that will test the boundaries of the validation rule as well as the data type and even the existence of the data. For example, if the validation rule was set to check for a value between 5 and 10 inclusive, then the test data values would be 4, 5, 6 and 9, 10, 11. A simple rule of thumb is to test a value before the value itself and one value greater than the value in the validation rule. Figures 1.66–1.69 show examples of how to annotate testing for the three parts of validation.

If you rely only on informal testing, you run the risk of not revealing errors that would occur when various modules or components of the solution are put together.

Functional and non-functional requirements were discussed earlier in this chapter.

tblOrderItems			
Order ID	Pizza ID	Crust type	Qty
52311	VEG	D	3
52311	BBQ	T	2
52311	MAG	D	1
59420	SUS	D	1
59420	CHI	D	2
83245	PEP	G	1
84779	VEG	G	2
84779	AUS	D	1
97645	MAI	G	1
12345	D		1
12345	D		
*	D		

Microsoft Access

You must enter a value in the 'tblOrderItems.oiQty' field.

OK

Help

[Was this information helpful?](#)

FIGURE 1.64 Results of existence check on the Qty (quantity) field in the Order Items table

TABLE 1.28 Validation testing: existence check on the Qty (quantity) field in the Order Items table from Figure 1.64

Function to be tested	Validation (existence check) on field oiQty in tblOrderItems table
Test data	Qty Null value (blank) 1
Required result and reason	Null value: An error message about missing data must be displayed because no value has been entered and the field requires one. 1: No error message will be displayed because a value has been entered in the field, as required.
Actual result	Null value: An error message about missing data is displayed 1: No error message is displayed Therefore, actual result = required result ✓ Validation (existence) OK

Access Objects

Customers

OrderItems

PizzaOrders

Pizzas

BowenOrdersInFebruary

Everything

VegetarianSoldByCrust

tblOrderItems			
Order ID	Pizza ID	Crust type	Qty
52311	VEG	D	3
52311	BBQ	T	2
52311	MAG	D	1
59420	SUS	D	1
59420	CHI	D	2
83245	PEP	G	1
84779	VEG	G	2
84779	AUS	D	1
97645	MAI	G	1
12345	D		1
12345	D		

The value you entered does not match the Number data type in this column.

Enter new value.

Help with data types and formats.

FIGURE 1.65 A data type check on the Qty field in the Order Items table. The green highlighted area indicates a validation error – a non-numeric value has been entered. The yellow highlighted area indicates a numeric value has been correctly entered.

tblOrderItems		
Field Name	Data Type	
oitOrderID	Text	Order ID
oitPizzaID	Text	Pizza ID
oitCrust	Text	D = deep, T = thin, G = gluten free
oitQty	Number	

FIGURE 1.66 The Data Type column shows how the Qty field's data type has been set up to only accept numeric values

TABLE 1.29 Validation testing: data type check on the Qty field in the Order Items table from Figure 1.65

Function to be tested	Validation (data type check) on field oitQty in tblOrderItems table
Test data	Order Qty 1 A
Required result and reason	1: No error message will be displayed because the field requires a numeric value, and 1 is a numeric value. A: An error message will be displayed to indicate that the data type is incorrect because the field requires a numeric value, and 'A' is not numeric.
Actual result	1: No error message is displayed A: An appropriate error message is displayed Therefore, actual result = required result ✓ Validation (data type) OK

tblOrderItems			
Order ID	Pizza ID	Crust type	Qty
52311	VEG	D	3
52311	BBQ	T	2
52311	MAG	D	1
59420	SUS	D	1
59420	CHI	D	2
83245	PEP	G	1
84779	VEG	G	2
84779	AUS	D	1
97645	MAI	G	1
12345		D	1
12345		D	2
12345		D	9
12345		D	10
12345		D	0
		D	



FIGURE 1.67 A range check on the Qty field in the Order Items table. The green highlighted area indicates a validation error – a quantity outside the acceptable range has been entered, being 0. The yellow highlighted area indicates a correct range of values.

12345		D	1
12345		D	2
12345		D	9
12345		D	10
12345		D	11
*		D	



FIGURE 1.68 A range check on the Qty field in the Order Items table. The green highlighted area indicates a validation error – a quantity outside the acceptable range has been entered, being 11. The error message is the same as for Figure 1.67.

TABLE 1.30 Validation testing: data type check on the Qty field in the Order Items table from Figure 1.64. Note the importance of explaining what happens for each value.

Function to be tested	Validation (range check) on field oitQty in tblOrderItems table		
Test data	Order Qty		
	0		
	1		
	2		
	9		
	10		
	11		
Required result and reason	1 2 9 10	No error message will be displayed because all values are within the range of 1 to 10 inclusive.	
	0 11	The message 'Between 1 and 10 (inclusive) pizzas can be ordered.' will be displayed because both values are outside the range of 1 to 10 inclusive.	
Actual result	1 2 9 10	No error message is displayed.	
	0 11	The message 'Between 1 and 10 (inclusive) pizzas can be ordered.' is displayed.	
Therefore, actual result = required result			
✓ Validation (range check) OK			

Testing the query selection criteria

Queries need to be tested to ensure that they select the correct data. Once you have created a set of test data, manually check the test data to identify the records that meet the criteria specified in the query. The test data must be chosen carefully to ensure that there are some records that should be returned by the query, some that should be rejected for failing to meet the first criterion even though they meet the second criterion, some that fail to meet the second criterion after meeting the first criterion and finally some that miss both criteria. You should note the records in the test data that should be returned and compare them against the records listed on the printout from the query. The records on both lists should match exactly.

TABLE 1.31 The vegetarian pizzas query using 30 records from database, with the five results bolded

Crust type	Qty	Total Cost (\$)	Order ID	Customer ID	Order Date	Order Time	Pizza Name
D	2	21.90	12345	7913	20-Jan-16	1823	Super Supreme
D	1	4.95	52311	0479	05-Feb-16	1835	Margherita
D	3	23.85	52311	0479	05-Feb-16	1835	Vegetarian
D	2	15.90	40991	7913	08-Feb-16	1901	Chicken
D	5	54.75	12347	1932	11-Feb-16	1234	Marinara
D	2	21.90	12347	1932	11-Feb-16	1234	BBQ Meatlovers
D	1	4.95	12346	0479	11-Feb-16	1801	Margherita

TABLE 1.31 (continued)

Crust type	Qty	Total Cost (\$)	Order ID	Customer ID	Order Date	Order Time	Pizza Name
D	4	19.80	12348	8860	12-Feb-16	1915	Pepperoni
D	2	9.90	12348	8860	12-Feb-16	1915	Aussie
D	2	9.90	22698	0479	15-Feb-16	2003	Margherita
D	1	4.95	84779	3006	19-Feb-16	1812	Aussie
D	1	4.95	32216	8345	24-Feb-16	1758	Aussie
D	4	19.80	00548	2669	26-Feb-16	1845	Margherita
D	1	4.95	00548	2669	26-Feb-16	1845	Pepperoni
D	4	19.80	00548	2669	26-Feb-16	1845	Hawaiian
D	1	10.95	59420	2201	03-Mar-16	1632	Super Supreme
D	2	15.90	59420	2201	03-Mar-16	1632	Chicken
G	2	15.90	40991	7913	08-Feb-16	1901	Vegetarian
G	3	23.85	12347	1932	11-Feb-16	1234	Vegetarian
G	1	4.95	83245	7913	18-Feb-16	1845	Pepperoni
G	1	10.95	97645	7913	18-Feb-16	1924	Marinara
G	2	15.90	84779	3006	19-Feb-16	1812	Vegetarian
G	1	10.95	32216	8345	24-Feb-16	1758	BBQ Meatlovers
G	2	15.90	00548	2669	26-Feb-16	1845	Vegetarian
T	1	7.95	12345	7913	20-Jan-16	1823	Chicken
T	2	21.90	52311	0479	05-Feb-16	1835	BBQ Meatlovers
T	2	9.90	40991	7913	08-Feb-16	1901	Pepperoni
T	3	14.85	12348	8860	12-Feb-16	1915	Hawaiian
T	1	4.95	12348	8860	12-Feb-16	1915	Aussie
T	1	10.95	32216	8345	24-Feb-16	1758	Super Supreme

qryVegetarianSoldByCrust						
Crust type	Qty	Total Cost (\$)	Order ID	Customer ID	Order Date	Order Time
D	3	23.85	52311	0479	05-Feb-16	1835
G	2	15.90	40991	7913	08-Feb-16	1901
G	3	23.85	12347	1932	11-Feb-16	1234
G	2	15.90	84779	3006	19-Feb-16	1812
G	2	15.90	00548	2669	26-Feb-16	1845
Total						
	12	95.40				5

FIGURE 1.69 The five results from the vegetarian pizzas query, highlighted by date

Function to be tested	Selection: ordOrderDate is >=#1/02/2016# AND pizPizzaName is 'Vegetarian' in qryDebtors query
Test data	30 records in database
Expected result and reason	5 results will be displayed, because only 5 records match both date and type of pizza criteria (refer to Table 1.31).
Actual result	The 5 expected records are displayed (refer to Figure 1.69). Therefore, actual result = expected result ✓ Query selection criteria OK

Testing formulas and summary statistics

It is also critical to test formulas and summary statistics (such as 'counts' and 'sums') that occur in your queries. Choose appropriate test data so that it is easy to determine if the final result is accurate. The examples shown in Figures 1.70–1.72 can be easily adapted to demonstrate testing for a wide range of similar database elements.

qryBowenOrdersInFebruary							
Order Date	Order Time	Order ID	Pizza Name	Crust type	Price (\$)	Qty	Total Cost (\$)
08-Feb-16	1901	40991	Pepperoni	T	4.95	2	9.90
08-Feb-16	1901	40991	Vegetarian	G	7.95	2	15.90
08-Feb-16	1901	40991	Chicken	D	7.95	2	15.90
18-Feb-16	1845	83245	Pepperoni	G	4.95	1	4.95
18-Feb-16	1924	97645	Marinara	G	10.95	1	10.95
Total							57.60

FIGURE 1.70 Testing the Total Cost formula in the vegetarian pizza query

Function to be tested	Formula: <code>tesTotalCost: [oitQty]*[pizPizzaPrice]</code> in <code>qryBowenOrdersInFebruary</code> query
Test data	Pepperoni = 4.95 Qty = 2
Expected result and reason	4.95 * 2 = 9.90 (Worked out on a calculator)
Actual result	9.90 is displayed (refer to top row of Figure 1.70) Therefore, actual result = expected result ✓ Formula OK

qryVegetarianSoldByCrust							
Crust type	Qty	Total Cost (\$)	Order ID	Customer ID	Order Date	Order Time	
D	3	23.85	52311	0479	05-Feb-16	1835	
G	2	15.90	40991	7913	08-Feb-16	1901	
G	3	23.85	12347	1932	11-Feb-16	1234	
G	2	15.90	84779	3006	19-Feb-16	1812	
G	2	15.90	00548	2669	26-Feb-16	1845	
Total							5

FIGURE 1.71 Testing the Count function in the vegetarian pizza query

Function to be tested	Count on <code>ordOrderDate</code> field in <code>qryVegetarianSoldByCrust</code> query
Test data	5 records are selected when the query is run
Expected result and reason	The count should be 5 because 5 records are selected when the query is run (refer to highlighted section of Figure 1.71).
Actual result	The count is 5. Therefore, actual result = expected result ✓ Count OK

qryVegetarianSoldByCrust						
Crust type	Qty	Total Cost (\$)	Order ID	Customer ID	Order Date	Order Time
D	3	23.85	52311	0479	05-Feb-16	1835
G	2	15.90	40991	7913	08-Feb-16	1901
G	3	23.85	12347	1932	11-Feb-16	1234
G	2	15.90	84779	3006	19-Feb-16	1812
G	2	15.90	00548	2669	26-Feb-16	1845
*						
Total	12	95.40				5

FIGURE 1.72 Testing the Sum function in the vegetarian pizza query

Function to be tested	Sum on <i>tesTotalCost</i> field in <i>qryVegetarianSoldByCrust</i> query
Test data	Order ID and Total Cost (\$) 52311: 23.85 40991: 15.90 12347: 23.85 84779: 15.90 00548: 15.90
Expected result and reason	$23.85 + 15.90 + 23.85 + 15.90 + 15.90 = 95.40$ (worked out on calculator)
Actual result	The sum is 95.40 (refer to bottom row of Figure 1.72). Therefore, actual result = expected result ✓ Sum OK

Testing sorting

A key feature of the query is being able to display its results in an organised and usable manner (i.e. as information). These primary and secondary sorts have to be tested. The key is to demonstrate that a sort works alphabetically or numerically or even alphanumerically. You will need at least three test values to demonstrate this adequately (Figure 1.73).

qryVegetarianSoldByCrust						
Crust type	Qty	Total Cost (\$)	Order ID	Customer ID	Order Date	Order Time
D	3	23.85	52311	0479	05-Feb-16	1835
G	2	15.90	40991	7913	08-Feb-16	1901
G	3	23.85	12347	1932	11-Feb-16	1234
G	2	15.90	84779	3006	19-Feb-16	1812
G	2	15.90	00548	2669	26-Feb-16	1845
T	1	7.95	11111	8345	19-Feb-16	2012

FIGURE 1.73 Testing the primary sort function in the vegetarian pizza query

Function to be tested	Primary sort (ascending) on oitCrust field in qryVegetarianSoldByCrust query
Test data	Crust Type D G T
Expected result and reason	The primary sort order will be: D G T This is because it follows alphabetical order.
Actual result	The primary sort order is: D G T (refer to the highlighted area of Figure 1.73) Therefore, actual result = expected result ✓ Primary sort OK

qryVegetarianSoldByCrust						
Crust type	Qty	Total Cost (\$)	Order ID	Customer ID	Order Date	Order Time
D	3	23.85	52311	0479	05-Feb-16	1835
G	2	15.90	40991	7913	08-Feb-16	1901
G	3	23.85	12347	1932	11-Feb-16	1234
G	2	15.90	84779	3006	19-Feb-16	1812
G	2	15.90	00548	2669	26-Feb-16	1845
T	1	7.95	11111	8345	19-Feb-16	2012

FIGURE 1.74 Testing the secondary sort function on the vegetarian pizza query

Function to be tested	Secondary sort (ascending) on ordOrderDate field in qryVegetarianSoldByCrust query
Test data	Order Date 8-Feb-16 11-Feb-16 19-Feb-16 26-Feb-16
Expected result and reason	The secondary sort order will be: 8-Feb-16 11-Feb-16 19-Feb-16 26-Feb-16 This is because it follows calendar dating.
Actual result	The actual secondary sort order is: 8-Feb-16 11-Feb-16 19-Feb-16 26-Feb-16 (refer to the highlighted area of Figure 1.74) Therefore, actual result = expected result ✓ Secondary sort OK

**CASE
STUDY****PIZZAS ON THE GO – PART 10**

Nick finished creating the POTG database and then completed formal testing. To begin, he tested the electronic validation in each of the tables. Examples appear in Figures 1.64–1.68.

Next, Nick tested the selection criteria for each of the two queries – refer to Figure 1.69.

He then went on to test the formulas and other summary statistics on each query, as seen in Figures 1.70–1.72.

Finally, Nick tested that the primary and secondary sorts on each query were working properly, as shown in Figures 1.73 and 1.74.

During the testing process, if Nick encountered any errors he corrected them and ran the testing procedure again until the result was accurate. Once he was sure that the solution was working as expected, he showed the results of the formal testing to Matt. This gave Matt confidence in the reliability of the solution Nick had created.

Protecting valuable data

Any data in a database is valuable. This means that its integrity needs to be protected so that it always produces the most accurate information possible. A database may hold confidential data that must be kept private from anyone except the authorised users. Most government databases fall into this category, but any data gathered via online sources must also be protected because it usually contains personal contact details, usernames, passwords and even credit card details.

There are, therefore, several physical and software-based controls that can protect data, both when stored and when transmitted. These measures help to reduce or eliminate the possible damage caused by deliberate or accidental threats, as well as event-based threats such as natural disasters. The most common measures include backing up, electrical protection, usernames and passwords, system security software and encryption.

Backing up

Databases are generally ‘live’, but they need the ability to have data ‘rolled back’ to an earlier time in case of data corruption or malicious activity. Full copies of databases can be made onto any storage device. This means the entire database can be restored to a previous point in time if necessary. Alternatively, many databases support version control of data, so they actually store changes to each data element in any record so that if necessary only that particular field is restored. A wiki site such as Wikipedia is a good example of an implementation of this type of database protection.

Electrical protection

A common way of protecting against power events is to use a well-regulated power supply that can intercept power surges or spikes. However, a database may need to remain accessible even under adverse power conditions. For this reason, backup power supplies and batteries are often used. An uninterruptable power supply (UPS) can regulate electricity coming into a computer and supply backup electricity for a short time if a blackout occurs.

Article Talk

Read Edit View history Search

Database design: Revision history

View logs for this page

Browse history

From year (and earlier): 2015 From month (and earlier): all Tag filter: Go

For any version listed below, click on its date to view it. For more help, see [Help:Page history](#) and [Help:Edit summary](#).

External tools: [Revision history statistics](#) • [Revision history search](#) • [Edits by user](#) • [Number of watchers](#) • [Page view statistics](#)

(cur) = difference from current version, (prev) = difference from preceding version, m = minor edit, → = section edit, ← = automatic edit summary

(newest | oldest) View (newer 50 | older 50) (20 | 50 | 100 | 250 | 500)

Compare selected revisions

- (cur | prev) 07:26, 13 May 2015 58.65.176.86 (talk) .. (11,678 bytes) (-5) .. (→Conceptual schema) (undo)
- (cur | prev) 01:01, 20 April 2015 ClueBot NG (talk | contribs) m .. (11,683 bytes) (+8,169) .. (Reverting possible vandalism by Bbllggppligg to version by 46.29.126.242. False positive? Report it. Thanks, ClueBot NG. (2201209) (Bot)) (undo)
- (cur | prev) 01:01, 20 April 2015 Bbllggppligg (talk | contribs) .. (3,514 bytes) (-8,169) .. (Restructuring the whole article.) (undo)
- (cur | prev) 15:12, 24 March 2015 46.29.126.242 (talk) .. (11,683 bytes) (-26) .. (→External links: updated Efficient Database Design ref with the right page) (undo)
- (cur | prev) 09:37, 3 March 2015 DivineAlpha (talk | contribs) m .. (11,709 bytes) (-53) .. (Reverted edit(s) by 112.196.3.86 (talk): editing tests (using WikiPatroller)) (undo)
- (cur | prev) 09:36, 3 March 2015 112.196.3.86 (talk) .. (11,762 bytes) (+47) .. (undo)
- (cur | prev) 09:11, 3 March 2015 112.196.3.86 (talk) .. (11,715 bytes) (+6) .. (→Conceptual schema) (undo)
- (cur | prev) 20:00, 14 January 2015 117.199.118.32 (talk) .. (11,709 bytes) (+6) .. (→Design processDatabase design basics. (n.d.). Database design basics. Retrieved May 1, 2010, from <http://office.microsoft.com/en-us/access/HA012242471033.aspx>) (undo)
- (cur | prev) 11:36, 8 January 2015 Widr (talk | contribs) m .. (11,703 bytes) (-21) .. (Reverted 2 edits by 164.100.222.244 using STiki) (undo)
- (cur | prev) 11:16, 8 January 2015 164.100.222.244 (talk) .. (11,724 bytes) (-5) .. (→ER diagram (entity-relationship model)) (undo)
- (cur | prev) 11:10, 8 January 2015 164.100.222.244 (talk) .. (11,729 bytes) (+26) .. (→ER diagram (entity-relationship model)) (undo)
- (cur | prev) 12:22, 23 December 2014 202.83.18.51 (talk) .. (11,703 bytes) (-1) .. (undo)

FIGURE 1.75 Wikipedia stores backup copies of articles and changes made to them. This allows editorial staff to ‘roll back’ to a previous version of an article without having to rewrite it.

Username and passwords

Username and passwords can be used to restrict access to an entire database, so that only authorised users can open it. Alternatively, usernames and passwords can be used in a similar manner to that of logging on to networks; that is, they can be set to allow certain users to edit records (similar to read/write

access), while others can only read, but not alter them (read access only). These types of databases will also record who changes what data as an auditing feature.

Systems security software

Various types of **malware** can cause significant damage to data. It is important to use a variety of systems security software in order to prevent malware or other malicious coding from running on your system and either damaging or destroying the contents of a database.



FIGURE 1.76 A password can be used in Microsoft Access to prevent unauthorised users from opening a database file

Encryption involves coding data into a form that only authorised users can read and decrypt. The encryption and decryption process both require a ‘key’. When the plain text is encrypted it is called ciphertext. When SSL or TLS protocols are used over HTTPS (on port 443) an encrypted and secure ‘pipeline’ is created between the user and the database, so that if the data being transmitted is intercepted it cannot be read without the key. This system is commonly used for internet transactions involving registration or order forms and for credit card details.

Encryption (storage and transmitted)

Like any other data that needs to be secured while being transmitted, encryption is used to prevent interception and theft. If data is entered via a website then it will use SSL or TLS over an HTTPS connection. Encryption was discussed earlier in this chapter. You will recall that online data acquisition forms use this type of encryption.

When stored, databases can also be encrypted. The contents can be encrypted inside fields. This is known as **hashing** and is often used to protect passwords. The entire database file can also be encrypted so that, if stolen, its contents cannot be opened and used by another person.

CASE STUDY

PIZZAS ON THE GO – PART 11

For the POTG database, Nick has limited access to personal and financial details to authorised users, such as Matt, by placing a password and username on the files. Customers can create new orders and change their delivery and other personal details, but they cannot alter any transaction records. Any changes to personal data and new orders are logged.

Anti-malware software will be installed on the server hosting the database while an encrypted connection will be used to transmit data between the server and customer's devices. It is unknown what security customers may implement on their own devices, but limiting what their user account can access in the POTG database and not storing any data on their device will minimise any malware threats originating from their devices.

A cloud storage service also acts as a backup and it keeps multiple versions of the POTG database. A portable external hard drive is also used as a secondary backup of the database each night.

ESSENTIAL TERMS

ASP active server pages; a general purpose scripting language used on web servers

ASP.Net a web application tool that assists in building dynamic webpages

bench test a check that determines how the solution, or a component of the solution, would behave with a range of test data; usually performed on paper, with the sample data being manipulated in the same way the proposed solution would manipulate the data electronically (the intent is to predict how the solution will treat the sample data)

cookie a small file that a web server stores on a user's computer, typically containing data about the user, such as the user's email address and web-viewing preferences

data acquisition software programs designed to create online user forms and then assist in retrieving data from the forms, transmit the data and then store it in a database

data integrity ensuring that all data in a database is complete

database a software tool used to hold and retrieve data

database record a set of data about one entity; for example, a person, event or object

dynamic webpage a page that changes as requested by the user; dynamic webpages provide more flexibility and interactivity than mere text and images

effectiveness completeness, readability, attractiveness, clarity, accuracy, accessibility, timeliness, communication of message, relevance and useability (within the context of a solution)

efficiency improvements in time, cost and effort (within the context of a solution)

entity a single person, place, or thing about which data can be stored

entity-relationship diagram (ERD) a diagram that shows the interrelationships between different entities and their attributes

field contains the same type of data for a series of records; the same field for a series of records will contain the same type of data

first normal form (1NF) a table with no repeating groups; that is, no single row has a column containing more than one value or more than one column with the same kind of value

flat file a database that stores data in tables consisting of rows and columns; also called a single table database

foreign key a common linking field; a key defined in a second table that refers to a primary key in a first table; see *also* primary key

global economy a place where all economies of the world merge and goods and services are sold at an international level

hashing a method of encryption that can be used to encrypt the contents of databases inside fields; a method of protecting passwords by taking a variable-length password and creating a cryptic, fixed-length password from it by generating and using a salt value (if the password field is then viewed by a user, they cannot decrypt the hashed password)

HTTPS hypertext transfer protocol secure, a combination of a secure version of HTTP and the SSL/TLS protocols; HTTPS provides encryption and security in terms of identification of the server

input mask a control applied to how data is entered, which can be set to accept a specified number of digits, dates or characters in a commonly accepted format

input-process-output (IPO) charts diagrams used to identify inputs, outputs and the processing steps required to transform the inputs into the outputs

layout diagram a hand-drawn sketch that shows the elements to be included on an input form; diagram would indicate the placement of fields and labels, the fonts to be used and any graphics or other elements to be included

machine validation a database's ability to detect limited types of errors in data that has been keyed, such as incorrect numbers entered for a postcode or certain boxes being empty, and then notify the data-entry operator of the error

malware any computer program designed to have undesirable or harmful effects

manual validation the process of visually inspecting (proofreading) data to check if any mistakes have been made in copying the data, generally completed by a data-entry operator

many-to-many relationship when each record in the first table can be connected to multiple records in a second table

normalisation removing redundant data and arranging it into appropriate tables to improve data integrity and reduce redundancy; it involves following a systematic set of rules to check for anomalies or deviations in data structure, which ensures fields are in the correct tables

one-to-many relationship when a single record in the first table can be connected to multiple records in a second table; for example, several workers in an office with the same phone extension

one-to-one relationship when a single record in the first table can be connected to only a single record in a second table; for example, a seat allocation table for airline passengers that holds records related to seats on a given flight, with a one-to-one relationship between each passenger and their allocated seat

ongoing services services provided by businesses or organisations after point of sale, such as tracking information that allow you to view the processing and shipping details of your order

online form a webpage with input fields into which a user enters and then submits data for processing; forms are largely a replacement for paper-based forms and have features that include dropdowns, checkboxes, radio buttons, text fields and mandatory fields; common types of forms include email sign-up forms, online purchases, subscriptions and enrolments

PHP hypertext pre-processor; a general purpose scripting language used on web servers

primary key a field attached to each record in a database; the value of this key should be unique for each record in the database; for example, are number plates on cars, tax file numbers and even your own student number; its purpose is to identify data related to each record, which may be stored across multiple tables

privacy policy a document usually found on a website that informs users why the organisation is gathering data, the purposes for which it will be used and how it will be stored, in accordance with the *Privacy Act 1988*

query to select specific data based on a series of criteria in order to answer questions and make links between data; the criteria are the results of questions about the data; for example, 'How many of our customers are female?'

RDBMS relational database management system; a database that stores data tables that are arranged in rows and columns, with tables linked by a common field; relationships may be one-to-one, one-to-many or many-to-many

report the result of a query converted into a format that is usable, easier to read and understand, that also allows you to add headings and summary statistics such as totals

second normal form (2NF) a table that is in 1NF and any column that is not part of the primary key is dependent on the whole primary key

secure sockets layer (SSL) a protocol used to provide security for communications on the internet; several versions are popular with users of the World Wide Web and email; it preceded TLS

shipping and returns policy a document that explains to customers what action they can take and the procedure to follow if they need to return items; shipping details explain the charges a customer might need to pay and how long products may take to ship

spyware a hard-to-detect file, discreetly installed without a user's express permission, that collects small amounts of information about users without their knowledge; it collects information on the user's online activities such as search habits, sites visited and software installation. It may also alter a computer's settings, including slowing down connection speeds and changing the default homepage browser settings

task completeness a solution's ability to support all of the tasks a user wants to perform, in the way they expect

test plan a technique for recording tests to be carried out and the results of the tests; typically, a test plan states the type of test, what test data will be used, what results are expected, and, ultimately, the actual results

third normal form (3NF) a table that is in 2NF and any column that is not part of the primary key is dependent only on the primary key and no other column

transport layer security (TLS) a protocol used to provide security for communications on the internet; several versions are popular with users of the World Wide Web and email; its predecessor was SSL

user flow diagram (UFD) a chart that depicts the steps a user follows to complete a transaction, such as a purchase or sign-up on a website

validation an activity within the development stage of the PSM; its purpose is to check for the reasonableness of data being input; typically, validation is used for existence checking, range checking and data-type checking

validation rule a rule that allows data to be entered into a form or table only if it obeys the specified rule; if the data does not meet the criteria, a message box displays an error message and the data is deleted



IMPORTANT FACTS

- 1 The **problem-solving methodology (PSM)** involves four stages: analysis, design, development and evaluation.
- 2 Organisations acquire data via **online facilities** for several reasons, including:
 - 24-hour customer access
 - improved efficiency through direct data entry by customers
 - improvements in effectiveness
 - access to global markets
 - marketing opportunities
 - ability to provide ongoing services.
- 3 Users **supply data** for online transactions because it is convenient, offers them a variety of choice and reduces costs.
- 4 Networks allow users to **share resources** such as data, information, hardware and software and to transfer money via electronic funds transfer.
- 5 A **UFD** should show the multiple entry points that could get a user to the beginning of a transaction and the key data flows or interactions they must go through to complete the transaction.
- 6 Department stores and other retailers, government sites, schools and social networking sites are a few examples of types of websites that allow individuals and organisations to **supply data**.
- 7 Techniques for **efficient and effective data collection** include short instructions, validation, appropriate types of fields, clear path to completion of a transaction and mobile friendliness.
- 8 Design principles related to **appearance** include alignment, repetition, contrast, space and balance.
- 9 Design principles related to **functionality** include useability, flexibility and ease of use, and accessibility.
- 10 Organisations must **protect the rights** of individuals and organisations who supply data. Methods used include the use of security protocols and stating privacy, shipping and returns policies.
- 11 **Shipping and returns policies** explain to customers what action they can take and the procedure to follow if they need to return items. The shipping details explain what charges a customer might need to pay and how long products may take to ship.
- 12 Common data types used in a **RDBMS** include text, numeric, date, character and Boolean.
- 13 A **field** contains the same type of data for a series of records. The same field for a series of records will contain the same type of data.
- 14 Records that meet specified criteria can be selected from the total number of records by performing a **query**. Each criterion can be a number, a piece of text or an expression.
- 15 The **structure for a RDBMS** can be determined by identifying entities, normalising tables, identifying fields and tables, identifying primary and foreign key fields, and defining data types and field sizes.
- 16 Use a **consistent naming convention** for RDBMS objects to make it clear what they are (for tables and queries) and what they are part of (for fields). Examples are `tblCustomers` for a table, or `qryAllProductsSoldToday` for a query. Fields should begin with letters that indicate the object they come from, such as `cusFamilyName`.
- 17 **Common design tools** used with databases include entity–relationship diagrams, data dictionaries, data structure diagrams, data structure charts, flowcharts, input–process–output charts and layout diagrams.
- 18 A **list tool** can be used to direct the entry of data where there is a limited number of allowed values. A list box on a form will display a predetermined set of values from which the user must select.
- 19 **Informal testing** is performed by the solution developer to ensure that components and functions are operating correctly.
- 20 **Formal testing** involves testing the overall solution after it has been completed.
- 21 **Test procedures** include establishing which tests will be conducted, determining which test data will be used, determining expected results, conducting the test, recording the results and correcting any identified errors.
- 22 Organisations use a range of software and physical controls to **protect their data**. These controls include backing up, providing electrical protection, using usernames and passwords, installing system security software and using encryption.

TEST YOUR KNOWLEDGE

DATA COLLECTION

- 1 Why would organisations prefer to use online forms to gather data than hardcopy forms?
- 2 List and expand on three reasons why individuals would prefer to purchase goods and services online.
- 3 List three specific techniques organisations use to make data collection efficient and effective.
- 4 In relation to design principles, explain the difference between appearance and functionality.
- 5 How do policies protect users (customers)?
- 6 Describe the method used to secure data being transmitted between a user and an online store.



Review quiz

RELATIONAL DATABASE MANAGEMENT SYSTEMS (RDBMS)

- 7 Make a table that summarises each of the key data types and provide an example of the data type intended for each field.
- 8 How does Boolean data differ from 'character'?
- 9 How does a relational database differ from a flat file database?
- 10 Explain how a field and a record work together.
- 11 What is the purpose of a query?
- 12 Explain the purpose of an entity-relationship diagram.
- 13 Why are primary keys important in relational databases?
- 14 What makes a good primary key?
- 15 How does a foreign key differ from a primary key?
- 16 Explain what a 'one-to-many' relationship is.
- 17 How does table normalisation help ensure the integrity of data in a RDBMS?
- 18 List the ways that an ERD assists with RDBMS design.
- 19 List two other RDBMS design tools related to functionality and explain how they are useful to database designers.
- 20 What are some important features of a query design?
- 21 Why is it necessary to select appropriate test data during the design stage?
- 22 Explain the importance of establishing validation rules for RDBMS tables.
- 23 During which stage of the PSM does testing take place?
- 24 Why is testing performed during this stage?
- 25 Pick one software and one physical control for data protection and explain how each is used to protect data.

APPLY YOUR KNOWLEDGE

A1 Software Supplies sells commercially produced software packages over the internet. A1 uses an online order form for customers to order software and have it delivered to their nominated address. Orders are currently stored in a spreadsheet, but this is now proving to be unwieldy and an alternative RDBMS has been suggested. As an employee at A1, you have been asked to create a list of orders received in the past week that have been paid for but have not yet been delivered. The current week to use for the purposes of this task is 03–10 April 2016.

- 1** Table 1.32 shows the unnormalised set of ordering data with which you have been supplied. Normalise this data to 3NF, explaining what you have done at each normal form.
- 2** Using your normalised tables create the following design elements.
 - a** An ERD.
 - b** A data dictionary, including table plans for each of the tables – you may need to add or remove fields in some tables. Also remember to include validation and to choose appropriate data types and formats.
 - c** A data structure diagram – how the tables need to be linked.
 - d** A sorted query that only lists orders that have been received during the week 03–10 April 2016, and that have been paid for but not yet delivered. Be sure to include relevant customer details, such as name and postal address, and appropriate calculations in the query.
 - e** A sorted query that lists all the transactions for Adobe Audition that have been paid for. Be sure to include relevant fields and appropriate calculations in the query.
- 3** Using a RDBMS package, create the A1 Software Supplies database. This will involve constructing tables and the two queries you designed above.
- 4** Test an example of each of the following database elements to ensure they are working as expected.
 - a** Electronic validation in the tables (existence, data type and range check)
 - b** Any formulas in the queries
 - c** The actual queries
 - d** The primary and secondary sort in a query
- 5** Use a graphics package to create a layout diagram of the starting screen of the order form A1 customers may use. Annotate the layout diagram to show how you have incorporated various design principles (both appearance and functionality) and security as well as how the data collection techniques are both efficient and effective.
- 6** Using a graphics package, create a UFD of the sequence of steps an A1 customer would follow in order to make an online software purchase. Annotate the UFD to indicate possible entry points to the ordering screen as well as where and how data protection is used.

TABLE 1.32 The unnormalised table for A1 Software Supplies

ORDER ID	ORDER DATE	PAID?	SENT?	CUSTOMER ID	FAMILY NAME	GIVEN NAME	DELIVERY ADDRESS	SUBURB	POST CODE	PHONE NO.	SOFTWARE ID	PRODUCT NAME	PRICE (\$)	QUANTITY	TOTAL COST (\$)
SM5360	16-Jan-16	TRUE	FALSE	2958	Burns	Mia	P.O. Box 251	Horsham	3400	98235872	581653	Pinnacle Studio	79	56	4424
MK2500	18-Jan-16	TRUE	TRUE	4801	Guerra	Lillith	197 Egestas Avenue	Laverton	3028	98004246	913096	MYOB	599	5	2995
											79389	Adobe Photoshop Elements	130	8	1040
JP7002	29-Jan-16	FALSE	FALSE	3801	Blair	Brooke	5/28 Nullam Avenue	Abbotsford	3067	98926102	684924	Adobe Audition	229	4	916
											485090	Adobe Creative Suite	435	3	1305
PS2357	14-Feb-16	TRUE	FALSE	1496	Sweet	Ralph	7476 Nisi Street	Traralgon	3844	98024988	244622	Norton Security	146	10	1460
											334806	MS Office	249	10	2490
											684924	Adobe Audition	229	10	2290
HH0254	16-Mar-16	TRUE	TRUE	5498	Mcclure	Marvin	97 Meteus Street	Dandenong	3175	98206895	334806	MS Office	249	2	498
											518116	MS Windows	215	5	1075
MJ4626	16-Mar-16	TRUE	FALSE	1496	Sweet	Ralph	7476 Nisi Street	Traralgon	3844	98024988	581653	Pinnacle Studio	79	2	158
											79389	Adobe Photoshop Elements	130	3	390
WD3684	01-Apr-16	TRUE	FALSE	6803	Hayes	Oprah	30 Sed Street	Hawthorn	3122	98949901	913096	MYOB	599	4	2396
ZJ0527	02-Apr-16	TRUE	TRUE	2958	Burns	Mia	P.O. Box 251	Horsham	3400	98235872	518116	MS Windows	215	10	2150
											913096	MYOB	599	10	5990
LF5814	03-Apr-16	FALSE	FALSE	3801	Blair	Brooke	5/28 Nullam Avenue	Abbotsford	3067	98926102	684924	Adobe Audition	229	3	687

TABLE 1.32 (continued)

ORDER ID	ORDER DATE	PAID?	SENT?	CUSTOMER ID	FAMILY NAME	GIVEN NAME	DELIVERY ADDRESS	SUBURB	POST CODE	PHONE NO.	SOFTWARE ID	PRODUCT NAME	PRICE (\$)	QUANTITY	TOTAL COST (\$)
JE0611	04-Apr-16	TRUE	FALSE	2958	Burns	Mia	P.O. Box 251	Horsham	3400	98235872	581653	Pinnacle Studio	79	3	237
											485090	Adobe Creative Suite	435	3	1305
RR3262	04-Apr-16	TRUE	TRUE	1496	Sweet	Ralph	7476 Nisi Street	Traralgon	3844	98024988	953726	Acronis True Image	59	5	295
											448484	Corel Designer	149	4	596
YT6617	05-Apr-16	TRUE	TRUE	2958	Burns	Mia	P.O. Box 251	Horsham	3400	98235872	244622	Norton Security	146	8	1168
											334806	MS Office	249	3	747
											485090	Adobe Creative Suite	435	2	870
QA5745	05-Apr-16	TRUE	FALSE	2028	Livingston	Clayton	64/152 Ute Avenue	Williamstown	3016	98470245	684924	Adobe Audition	229	6	1374
											953726	Acronis True Image	59	2	118
											334806	MS Office	249	5	1245
											518116	MS Windows	215	5	1075
											684924	Adobe Audition	229	5	1145
PY7538	05-Apr-16	TRUE	FALSE	5839	Parrish	Briar	6083 Vulputate Avenue	Footscray	3011	98247044	953726	Acronis True Image	59	12	708
											684924	Adobe Audition	229	56	12824
IM8529	06-Apr-16	TRUE	FALSE	6803	Hayes	Oprah	30 Sed Street	Hawthorn	3122	98949901	953726	Acronis True Image	59	65	3835
											518116	MS Windows	215	22	4730

TABLE 1.32 (continued)

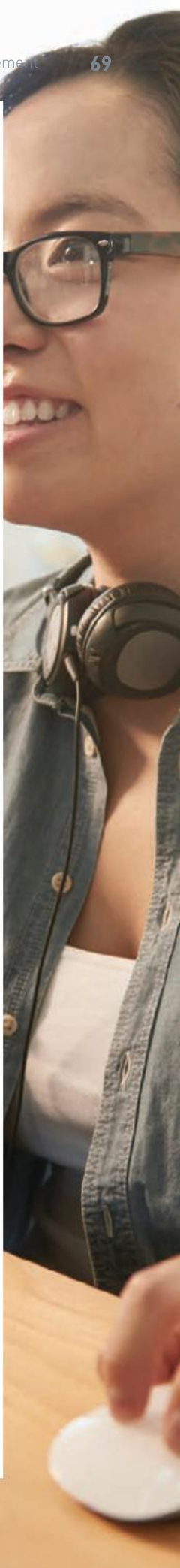
ORDER ID	ORDER DATE	PAID?	SENT?	CUSTOMER ID	FAMILY NAME	GIVEN NAME	DELIVERY ADDRESS	SUBURB	POST CODE	PHONE NO.	SOFTWARE ID	PRODUCT NAME	PRICE (\$)	QUANTITY	TOTAL COST (\$)
EB9801	06-Apr-16	TRUE	FALSE	6803	Hayes	Oprah	30 Sed Street	Hawthorn	3122	98949901	485090	Adobe Creative Suite	435	8	3480
IU2820	07-Apr-16	TRUE	FALSE	2958	Burns	Mia	P.O. Box 251	Horsham	3400	98235872	448484	Corel Designer	149	9	1341
UX3020	07-Apr-16	TRUE	TRUE	6803	Hayes	Oprah	30 Sed Street	Hawthorn	3122	98949901	518116	MS Windows	215	99	21285
CB0814	07-Apr-16	FALSE	FALSE	1496	Sweet	Ralph	7476 Nisi Street	Traralgon	3844	98024988	79389	Adobe Photoshop Elements	130	1	130
DF3357	08-Apr-16	FALSE	FALSE	2491	Carlson	Jaden	478 Pellentesque Road	Yarraville	3013	98566907	448484	Corel Designer	149	10	1490
SU8832	08-Apr-16	TRUE	TRUE	4801	Guerra	Lillith	197 Egestas Avenue	Laverton	3028	98004246	913096	MYOB	599	6	3594
US5970	10-Apr-16	TRUE	FALSE	6803	Hayes	Oprah	30 Sed Street	Hawthorn	3122	98949901	334806	MS Office	249	45	11205
GP5943	17-Apr-16	TRUE	FALSE	5839	Parrish	Briar	6083 Vulputate Avenue	Footscray	3011	98247044	485090	Adobe Creative Suite	435	6	2610
ZH3405	21-Apr-16	TRUE	TRUE	2958	Burns	Mia	P.O. Box 251	Horsham	3400	98235872	448484	Corel Designer	149	87	12963
											684924	Adobe Audition	229	6	1374
											79389	Adobe Photoshop Elements	130	17	2210

TABLE 1.33 The final result of the query for software that had been paid for but not dispatched during the week of 3-10 April 2016

ORDER ID	ORDER DATE	CUSTOMER ID	FAMILY NAME	GIVEN NAME	DELIVERY ADDRESS	SUBURB	POST CODE	PHONE NO.	PRODUCT ID	PRODUCT NAME	PRICE (\$)	QUANTITY	TOTAL PRICE (\$)
JE0611	04-Apr-16	2958	Burns	Mia	P.O. Box 251	Horsham	3400	98235872	485090	Adobe Creative Suite	435	3	1305
JE0611	04-Apr-16	2958	Burns	Mia	P.O. Box 251	Horsham	3400	98235872	581653	Pinnacle Studio	79	3	237
QA5745	05-Apr-16	2028	Livingston	Clayton	64/152 Ute Avenue	Williamstown	3016	98470245	953726	Acronis True Image	59	2	118
QA5745	05-Apr-16	2028	Livingston	Clayton	64/152 Ute Avenue	Williamstown	3016	98470245	684924	Adobe Audition	229	6	1374
PY7538	05-Apr-16	5839	Parrish	Briar	6083 Vulputate Avenue	Footscray	3011	98247044	953726	Acronis True Image	59	12	708
PY7538	05-Apr-16	5839	Parrish	Briar	6083 Vulputate Avenue	Footscray	3011	98247044	684924	Adobe Audition	229	56	12824
EB9801	06-Apr-16	6803	Hayes	Oprah	30 Sed Street	Hawthorn	3122	98949901	448484	Corel Designer	149	9	1341
EB9801	06-Apr-16	6803	Hayes	Oprah	30 Sed Street	Hawthorn	3122	98949901	485090	Adobe Creative Suite	435	8	3480
IM8529	06-Apr-16	6803	Hayes	Oprah	30 Sed Street	Hawthorn	3122	98949901	518116	MS Windows	215	22	4730
IM8529	06-Apr-16	6803	Hayes	Oprah	30 Sed Street	Hawthorn	3122	98949901	953726	Acronis True Image	59	65	3835
IU2820	07-Apr-16	2958	Burns	Mia	P.O. Box 251	Horsham	3400	98235872	448484	Corel Designer	149	5	745
US5970	10-Apr-16	6803	Hayes	Oprah	30 Sed Street	Hawthorn	3122	98949901	334806	MS Office	249	45	11205

TABLE 1.34 The final result of the query about Adobe Audition software that has been sold

ORDER DATE	ORDER ID	CUSTOMER ID	FAMILY NAME	DELIVERY ADDRESS	SUBURB	POST CODE	PHONE NO.	QUANTITY	TOTAL COST (\$)
14-Feb-16	PS2357	1496	Sweet	7476 Nisi Street	Traralgon	3844	98024988	10	2290
05-Apr-16	EK6196	2028	Livingston	64/152 Ute Avenue	Williamstown	3016	98470245	5	1145
05-Apr-16	PY7538	5839	Parrish	6083 Vulputate Avenue	Footscray	3011	98247044	56	12824
05-Apr-16	QA5745	2028	Livingston	64/152 Ute Avenue	Williamstown	3016	98470245	6	1374
21-Apr-16	ZH3405	2958	Burns	P.O. Box 251	Horsham	3400	98235872	6	1374



PREPARING FOR

UNIT 3 OUTCOME 1

Design a solution, develop it using a relational database management system, and diagrammatically represent how users interact with an online solution when supplying data for a transaction

In Unit 3, Outcome 1, you are required to apply the design and development stages of the problem-solving methodology to create a relational database solution to an information problem. The database solution should include tables with relationships and at least two queries. Use a layout diagram to represent the page on which a person begins an online transaction. It may not be the landing page, but it must be the page on which they commence their transaction. Create a user flow diagram to represent the steps involved (and the data to be supplied) in a transaction, culminating perhaps in a confirmation from the organisation.

OUTCOME MILESTONES

You are required to complete the following.

- 1 Read the design briefs (analysis) supplied by your teacher and identify the problem.
- 2 Apply the problem-solving methodology to the case study.
- 3 Create an ERD.
- 4 Normalise the supplied data to the third normal form (3NF).
- 5 Design the solution.
- 6 Use relevant functions of RDBMS software to develop the solution.
 - Validate the data.
 - Test the solution.
 - Manipulate the data.
- 7 Use relevant software functions to demonstrate features of the software.
- 8 Use a graphics package to create the user interface on the page that a user initiates an online transaction, that includes security.
- 9 Use a graphics package to create an annotated UFD to demonstrate your knowledge of how data is acquired and supplied when making online transactions, including how the data is protected.

STEPS TO FOLLOW

- 1 Carefully read the design briefs and identify the solution that needs to be produced.
- 2 Identify the information that needs to be created. The queries that you produce will be based on the information needs of the organisation. What needs to be communicated and to whom?
- 3 Create an ERD to represent your understanding of the entities involved in the solution and their relationships with one another.
- 4 Normalise the data supplied by your teacher to the 3NF. Explain your reasoning at each step.
- 5 Use appropriate design techniques to represent your plans for the solution.

- 6 Use RDBMS software to develop the solution according to your designs. Attempt to use as many of the relevant functions of the software as possible.
- 7 Apply appropriate formats and conventions to produce an effective solution.
- 8 Use appropriate test data to test all features of your solution. Record the results in a test plan and on annotated printouts.
- 9 Correct any errors revealed by the testing.
- 10 Input the data into the solution. You may initially need to manually validate data to rectify any obvious errors.
- 11 Use machine validation to validate the inputted data.
- 12 For security and integrity purposes, regularly back up your files onto a flash drive or other location as indicated by your teacher.
- 13 Consider the questions below. They relate to efficient and effective techniques for acquiring data from websites as well as security controls. Use these questions to assist you in completing the form design and UFD below.
 - Why is data acquired and supplied via the organisation's website?
 - What techniques should the organisation use to acquire data via their website?
 - How should the organisation protect the rights of their data providers?
 - What software and physical security controls might A1 use to protect data that is either stored or transmitted?
- 14 Prepare a layout diagram to represent a form screen that would initiate an online transaction process. Annotate the form screen to indicate design principles and data acquisition techniques. Also annotate to show any aspects from the questions in step 13. You must use graphics software for this diagram.
- 15 Prepare a UFD of the stages involved in an online transaction – this scenario might differ from the one used for the RDBMS. Annotate it to show the various ways that users may link to the transaction process. Clearly indicate the data security measures that are employed and where. Also annotate the UFD to include features that are based on any of the questions in step 13. You must use graphics software for this diagram.

DOCUMENTS REQUIRED FOR ASSESSMENT

- 1 A representation of the design of the solution, including an ERD and table normalisation
- 2 Printouts of the solution
- 3 An annotated solution indicating where testing has taken place
- 4 Annotated layout diagram of the page on which an online transaction commences, including any reference to security and data protection
- 5 Annotated UFD

ASSESSMENT

This task is marked out of 100 and is worth 10% of your total school assessed study score.

- 90 marks are allocated for the RDBMS solution.
- 10 marks are allocated for the annotated UFD and user interface page.

A more detailed set of assessment criteria will be prepared and distributed by your teacher prior to the commencement of the task. Details on the actual criteria for this Outcome and the breakdown of how each criterion will be marked will be available at <http://informatics3and4.nelsonnet.com.au>.

DATA ANALYTICS: DRAWING CONCLUSIONS, PART 1

Key knowledge

After completing this chapter, you will be able to demonstrate knowledge of:

Data and information

- primary and secondary data sources (digital and non digital) and methods of data acquisition, including observation, interview and querying of resources
- suitability of quantitative and qualitative data for manipulation including comparisons (quantitative) and policy formation (qualitative)
- data types and data structures relevant to selected software tools
- one of the following methods for referencing primary and secondary sources: Harvard, American Psychological Association (latest edition), Chicago, Institute of Electrical and Electronics Engineers (IEEE)
- criteria to check the integrity of data including timeliness, authenticity, relevance, accuracy
- techniques for coding qualitative data to support manipulation

Approaches to problem solving

- features of a reasonable hypothesis including a specific statement identifying a prediction and the variables
- solution specifications: requirements, including data to support the prediction of the hypothesis, constraints and scope

For the student

This chapter begins the discussion of the theory and skills required for Unit 3, Outcome 2, which is the first part of the School-assessed Task (SAT) that you will complete in Unit 4, Outcome 1.

You will be introduced to hypotheses, solution requirements, types and sources of data, methods of data acquisition, methods of referencing sources and how to check the quality of data.

For the teacher

This chapter, along with Chapter 3, covers the theory required for Unit 3, Outcome 2, which is the first part of the SAT that is completed in Unit 4, Outcome 1. Chapter 2 focuses on data: its types, acquisition, referencing, integrity, coding and the legislation relevant to storing and communicating personal data.

What is data?

Data is made up of facts and statistics. Raw facts have no context to them, so you cannot make much sense of them, or give them any meaning. To understand and make meaning of your data, you need to process it, converting it into something useful: information.

The potential value of data cannot be fully unlocked without processing it into information. Marked ballot papers after an election hold a great deal of data and thus considerable potential value, but in their raw form, they hold little value. They need to be processed by counting and grouping into their electorates before they become useful information: election results.

In this chapter, you will learn more about hypotheses and data. As part of Unit 3, Outcome 2, you need to formulate a hypothesis, collect data for your hypothesis and come to a conclusion as to whether it has been supported or refuted. The first thing we will discuss is what makes up a hypothesis and how to formulate one for your Outcome.

Next, we will talk about how to determine the specifications for a solution, both in terms of general solutions and your specific solution, to help you understand how to determine factors such as scope and constraints.

Then, we will discuss the data you need to collect, in terms of how it is categorised when it is interpreted by you when you are the main researcher, and how it is categorised when it is interpreted by someone else. Then, we will talk about how data can be categorised in a different way – as either quantitative or qualitative.

The first thing to do after forming a hypothesis is to gather data, so we will cover how you can acquire data through sources such as libraries and data sets, and then through methods such as interviews and observation. We will also discuss how to reference those sources properly.

Once you have gathered all of this data, you need to store it, protect it and understand what type of data it is. We will discuss integrity of data and how to maintain it, through measures such as timeliness, accuracy, authenticity and relevance. This is important because you need to maintain the integrity of the data you collect for your Outcome so that your results and conclusion are considered reliable.

Finally, we will briefly cover data types and structures, because this is relevant to your Outcome but is covered in more detail in Chapter 1.

Refer to Chapter 1 for its discussion of databases and data types.

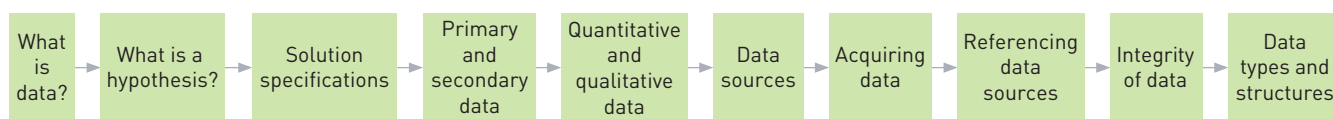


FIGURE 2.1 Chapter map

What is a hypothesis?

A **hypothesis** is an educated guess or a prediction that you can test in a scientific way. A hypothesis tries to explain an observation or unexplained pattern, and is often written in the form of an ‘If ... then ...’ statement, though it can also be a question or a conditional statement.

An ‘If ... then ...’ statement:

If I eat a paleo diet when I am overweight, then I may lose weight because I will no longer eat processed foods, refined sugar, dairy or whole grains, and this may improve my glucose tolerance and cholesterol levels, and reduce my waist circumference.

A hypothetical question:

Does eating a vegan diet have health benefits, because it can be high in fibre and it reduces intake of saturated fats, which could help to maintain a healthy weight?

A conditional statement:

Eating a low-fat diet for six months *may* have health benefits *because* it may improve cardiovascular functioning *and* help to maintain a healthy weight.

In Unit 3, Outcome 2, you must formulate a hypothesis, gather data to support it and come to a conclusion as to whether it has been supported or refuted. A hypothesis cannot be proven; it can only be supported or refuted using evidence.

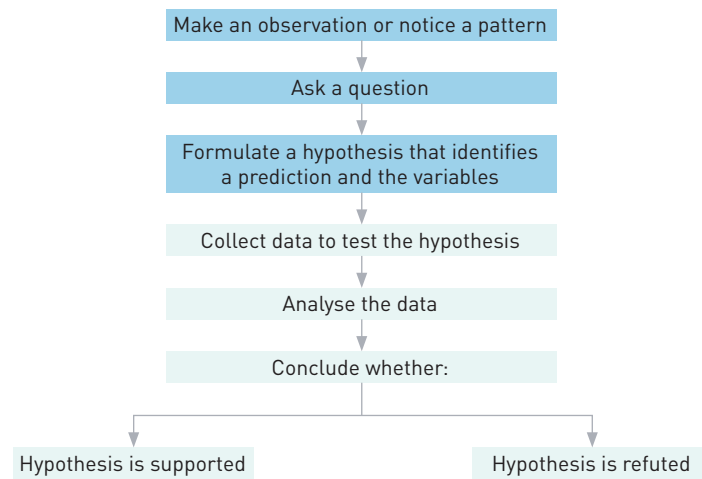


FIGURE 2.2 Formulating and testing a hypothesis for Unit 3, Outcome 2

Your hypothesis could be on a topic drawn from a field such as entertainment, sport, business, education, science, medicine or food. For example, it could reflect an existing or emerging trend, such as confirming or predicting a changing pattern in food culture in a specific area caused by shifting demographics. Similarly, you could investigate changing patterns in sporting culture caused by shifting demographics in a specific area, such as whether a particular sport has been growing in popularity and why.

Note: It is very important that, for your Outcome hypothesis to be considered acceptable, you **must** formulate a hypothesis that identifies **two** variables. For example, if you noticed an unexplained pattern of popularity of people eating a paleo diet and you wanted to find out why, examples of structurally acceptable hypotheses that you could test would be:

Eating a paleo diet is becoming popular in Melbourne because it is being promoted by TV celebrities and famous chefs.

Do increasing numbers of people in Melbourne eat a paleo diet because they believe it has health benefits?

These hypotheses are structurally acceptable because of the inclusion of the reason that explains *why* this hypothesis is being proposed: 'because it is being promoted by TV celebrities and famous chefs', 'because they believe it has health benefits'.

You will find the SAT much more engaging and interesting if you choose a hypothesis that relates to an interest of yours.

These variations would also be *more* structurally acceptable, but they would be too difficult to test for your Outcome:

If I eat paleo 'bone broth', then I will experience health benefits because it contains a lot of calcium, magnesium and other nutrients.

Does eating paleo 'bone broth' have health benefits because it is rich in collagen?

We are including these difficult hypotheses so you gain a better understanding of how to structure a hypothesis that is both acceptable and *testable*. You need to be sure that your hypothesis is testable, and that you are able to sample the relevant population. Even the acceptable hypotheses could be improved, because 'in Melbourne' is a large sample of people to try to test, and does it mean 'living in Melbourne' or 'happens to be in Melbourne at the time of eating'?

Eating a paleo diet is becoming popular among people who live in Richmond because it is being promoted by TV celebrities and famous chefs.

Do increasing numbers of people who live in Richmond eat a paleo diet because they believe it has health benefits?

You must be able to make a prediction from your hypothesis; perhaps the popularity of the paleo diet will die down because it is promoted less by celebrities as the 'next big thing'. Perhaps its popularity will increase if the health benefits become more obvious.

The hypothesis you formulate for your Outcome will have an impact upon the data you gather and your solution. The nature of the hypothesis you write will determine the data types you need. It will also have an impact upon the solution you create for Unit 4, Outcome 1. Ensure you create a reasonable hypothesis that identifies a prediction and the variables.

Characteristics of a reasonable hypothesis

TABLE 2.1 Characteristics of a reasonable hypothesis

Characteristic	Reason	Examples
1 It should be able to generate a testable prediction.	You need to be able to gather data that will allow you to support or refute the claim. If you cannot make a prediction, it is not a workable hypothesis.	As a result of climate change, the average daily temperature in summer in Hobart will rise 1.05° degrees within the next five years. This is testable because you can capture quantitative data about the temperature, which makes it a testable prediction.
2 It should have a specific statement identifying the variables. When a hypothesis says that one thing causes or affects another thing, the independent variable is the cause of the change, and the dependent variable is the factor that is affected. Researchers change the independent variable to observe the effects on the dependent variable.	Both the independent and dependent variables need to be measurable if the hypothesis is to be valid.	• Time (independent variable) affects the number of transistors in a CPU (dependent variable). • Dark clouds (independent variable) indicate imminent rain (dependent variable). • Eating junk food (independent variable) causes obesity (dependent variable). • Immigration (independent variable) leads to economic growth (dependent variable).
3 It must be testable.	A proposition that can never be supported or refuted is not a reasonable hypothesis.	Is it possible that our universe is a single atom within another universe? – not testable 'Our universe is expanding' – is testable

Characteristic	Reason	Examples
4 It needs at least <i>some</i> supporting evidence or observation.	Unsubstantiated supposition is not a hypothesis, regardless of the charisma, status or dedication of the believer. A hypothesis with no evidence is conjecture, speculation or faith.	'Crystals promote healing' – not supported by evidence 'Maggots have beneficial medical applications' – supported by current and historical medical evidence
5 It should be based on a specific observation and be limited to specific circumstances.	It may be hard to identify a single, consistent, independent variable among different observations.	'All diseases are caused by faulty genes' – not based on a specific observation, so not reasonable 'This particular rare disease is caused by a faulty gene' – may be reasonable
6 Evidence should be objective, not subjective.	Subjective evidence is opinion, not fact.	'Cats are better pets than dogs' – opinion 'Pet ownership has beneficial medical effects for people' – supported by evidence
7 It should not be vague or ambiguous.	Ambiguity and vagueness leave room for subjectivity.	'Democracy is the best form of government' – not reasonable, because 'best' is not objectively defined, but is vague and ambiguous 'Democracy promotes economic growth better than other forms of government' – may be reasonable
8 It is not trivial.	Insignificant hypotheses are of little merit.	'Brontosauruses are thin at one end, much thicker at the middle and thin again at the far end' is falsifiable and objective, but insignificant 'Brontosauruses were warm-blooded' would be significant because it would have evolutionary and biological implications
9 It should be able to make testable predictions.	A successful prediction provides strong evidence that the dependent variable is in fact affected by the independent variable.	Moore's law states that the number of transistors in an integrated circuit doubles approximately every two years. A testable prediction is that within 24 months, the number of transistors in a CPU will have doubled.
10 It should be no more complicated than it need be to explain an observation.	Researchers use Occam's Razor, a principle that states that when there are competing hypotheses to explain an observation, the simpler one with fewer assumptions and conditions is preferable.	To explain crop circles is it more likely that (a) aliens travelled thousands of years to interfere with our crops and then fly away, or (b) some bored local farmers had some fun?

Formulating your hypothesis

One way of formulating a hypothesis statement is as follows.

- 1 Choose a topic you find interesting, perhaps from one of those suggested on page 74.
- 2 Think about an interesting question that may be true about that topic, such as:
 - Male professional athletes earn more than female professional athletes.
 - The consumption of genetically modified food is increasing steadily in Australia.
 - The PC is becoming obsolete and will soon disappear.
 - Buddhism is rapidly growing in popularity in Australia.
 - More and more people are becoming vegan in Australia.
- 3 Conduct preliminary research to locate relevant sources of online and printed data on your topic to see if it is worthwhile pursuing for your Outcome.
- 4 Restructure your topic into a testable, reasonable hypothesis statement.

Alternatively, if you find it difficult to choose a topic, consider looking at data sets online until you find one that interests you.

You could also scan news websites until an issue grabs your attention. To make the Outcome interesting, form your hypothesis from something that is relevant to you – local sporting culture or emerging food trends in your community.

Proof, support and refutation

A hypothesis can never be **proven** true with data: it can only ever be **supported**. However, any hypothesis can be **refuted** (proven to be false). A reasonable hypothesis is not like a debate topic with equally valid opinions on both sides. You must be able to objectively support it or refute it.

Captain Cook probably held the following hypothesis:

All swans are white because every swan I have ever seen in my life in England has been white.

Then he sailed to Australia and saw his first black swan. Hypothesis refuted.

A thing can never be proven to be impossible. History is filled with instances of people who declared that many now-commonplace things were impossible:

What can be more palpably absurd than the prospect held out of locomotives travelling twice as fast as stagecoaches?

— *The Quarterly Review*, 1825.

A rocket will never be able to leave the Earth's atmosphere.

— *The New York Times*, 1920.

There is no chance that the iPhone is going to get any significant market share.

— Steve Ballmer, *USA Today*, 2007.

I think there is a world market for maybe five computers

— Thomas Watson, President of IBM, 1943.

There is no reason anyone would want a computer in their home.

— Ken Olsen, founder of Digital Equipment Corporation, 1977.

When you collect your data, make sure that your hypothesis and data are testing what you think you are testing, and nothing else is influencing your result. To do this, you need to conduct a **fair test**. To create a fair test, you need to understand variables. A variable is a factor that can change. In a typical experiment or investigation, several variables may affect your results. As you collect your data, you must make sure that you plan a fair test so that only one variable may change and the others are kept the same, or are **controlled**. (Independent and dependent variables are also described in Table 2.1.)

Controlling your variables will increase the certainty of your data being able to support or refute your hypothesis.

A theory is not the same as a hypothesis. A theory is a broad, unifying explanation of a set of hypotheses. It is well-supported by a wide set of factual evidence, has been repeatedly tested and verified under different circumstances and accepted as true by most experts ... yet could still be proven to be false. For example, we do not call Einstein's theory of relativity (encompassing both his theory of special relativity and general relativity) a 'theory' because we doubt its legitimacy. On the contrary: both have been repeatedly tested and confirmed in the decades since they were proposed. This differs dramatically from a hypothesis, which is, by comparison, a modest proposal with little evidence.

**CASE
STUDY****PAVLOV'S DOGS?**

One day, Pavlov, a hospital worker, notices patients playing with a dog that a visitor has brought in during visiting hours. When he approaches to reprimand the visitor for bringing an animal into the hospital, he stops himself, because he realises that many of the patients are smiling and seem to be free of their usual pain.

Finding this interesting, Pavlov forms a general hypothesis:

Interaction with animals affects the mental health of hospital patients.

While this is interesting, the hypothesis is vague and unfocused. It is not really a hypothesis at all. Pavlov makes it more specific and targeted:

Interaction with animals *improves* the mental health of patients.

Now there is some direction, but the hypothesis is not directly testable and verifiable. Pavlov is curious whether he can verify his hunch. He wonders why animals improve the mental health of patients, and how his hypothesis can be supported. He devises a testable hypothesis that may lead to research:

Interaction with animals improves the mental health of patients *because* it reduces their stress levels.

This has now established variables that can be modified and observed.

- Independent variable: Interaction with animals can be modified by controlling the amount of interaction patients have with animals.
- Dependent variable: Stress levels can be measured by conducting interviews or measuring stress hormones in patients' blood.

Pavlov realises that when he collects data, it is important that other variables that could affect the results are **controlled**. What if, rather than the pets affecting stress levels, other factors were responsible, such as illness, diet, sex, location, age and so on? Data should be filtered so that only comparable data is actually compared. Pavlov realises that it would be futile to compare the data from 90-year-old dementia patients in Germany *with* pets, with children with cancer in China *without* pets. Too many uncontrolled variables such as age, location, diet, and so on, could affect the data.

Once Pavlov stated his hypothesis clearly, he could generate a prediction from it:

A patient's stress levels will be reduced if the patient is given exposure to domesticated animals, such as cats or dogs.

With interest from hospital administrators, experiments were conducted, primary data was collected from interviews, and blood tests and secondary evidence was collected from similar research elsewhere. These data were averaged and compared. A statistical significance test led Pavlov to the conclusion that, on average, there was a statistically significant improvement in the mental health of patients who were given regular exposure to animals compared with those without exposure.

Pavlov's theory was supported, and the hospital administration used the information to create and fund a 'Pets as Therapy' programme.

Solution specifications

The School-assessed Task (SAT) comprises two solutions:

- The solution for Unit 3, Outcome 2 is the **information** you create from the data you collect and analyse. This information is used to support or refute your hypothesis.

- The solution for Unit 4, Outcome 1 is the **multimodal online solution** (MMOS) that communicates your findings about the hypothesis.

The following sections discuss the specifications of the solution you will create for Unit 3, Outcome 2, to help you create a solution that meets the requirements and to prepare you for gathering the appropriate data. When you create your solutions for the SAT, you will be using the problem-methodology (PSM) to guide their development. This discussion will include the analysis stage, which involves defining the specifications of what is needed to produce a satisfactory solution.

- Solution requirements: What your information needs to do, the qualities it should have, and what data is required to create the information
- Constraints: The limits and restrictions under which the information must be produced
- Scope: What the information must achieve, and what it is not required to achieve

The following section discusses each of these specifications in turn.

Solution requirements

Any solution that you submit will have requirements that are functional and non-functional, as well as data requirements.

Functional requirements

Functional requirements describe *the tasks* that your solution should be able to perform. In the simplest terms, these are the things you want your solution to be able to do – the main reason you are creating it. For example, an accounting program's functional requirements may specify that it must be able to:

- print annual summaries of expenses and income
- create expense charts
- calculate wages and deduct tax.

The functional requirement or requirements are usually achieved in a specific and identifiable place in a solution, such as a particular formula, a piece of programming or a webpage.

The main functional requirement of your Outcome's information would be that it lets you reach a valid and substantiated conclusion as to whether your hypothesis is supported or refuted.

Non-functional requirements

Non-functional requirements describe the *attributes* or *qualities* that your solution should have. Using the accounting program as an example again, its non-functional requirements may require it to be the following: accurate, secure, fast and easy to use.

A non-functional requirement will probably not be achieved in one specific place in a solution. It usually requires a combination of factors across an entire solution. For example, achieving ease of use in an accounting program may involve building in simple menus, shortcut keys, context-sensitive help, using clear language and making sure the interfaces are user-friendly in many different places.

Later in the chapter, we will discuss how data needs to be timely, authentic, relevant and accurate. The information you produce for your solution for this Outcome needs to be useable. This could mean it needs to be the following: accurate, authentic, timely, relevant, complete, up-to-date (current), specific, unambiguous and unbiased.

Data requirements

The specifications you write for your solution must include a description of the data that you require to support your hypothesis. You must locate this data for your Outcome. You must find, select, reference, organise, process and interpret it to produce information to enable you

to reach a conclusion. Your data requirements relate to the analysis, design and development stages of the problem-solving methodology (PSM).

For example, if you wanted to use Pavlov's hypothesis from page 78, you would need data on:

- the length of time that patients spend with animals
- levels of stress hormones in patients' blood before and after being with animals
- subjective reports of stress.

If you wanted to use one of the paleo diet examples such as:

Do increasing numbers of people who live in Richmond eat a paleo diet because they believe it has health benefits?

As a bare minimum, you would require data on:

- numbers of people in Richmond who eat a paleo diet, to support the idea that numbers are increasing
- subjective reports of why people choose to eat paleo, to support or refute the idea that they are doing so because they believe it has health benefits.

You would also need to define the 'paleo diet' and 'health benefits' as part of your planning, and your subjects and respondents would respond to this information.

Interpreting your data

Interpreting your data may be a complex process depending on the hypothesis you have chosen. You may find it difficult to be sure whether your hypothesis has been supported or refuted.

When you are interpreting your data, consider the following questions.

- Have you been able to support or refute the hypothesis satisfactorily?
- Do your results meet your expectations?
- Do your results make sense?
- Could your results be interpreted differently?
- Is your supporting data of good quality?
- Is your supporting data current?

You should also consider using tools for organisation and techniques for finding patterns, such as **mindmapping**. Patterns in data can be noticed more easily if you use tools, such as **graphs**, or other visual techniques, such as **frequency distribution tables**, which will show the number of observations in each category.

Chapter 3, Identifying patterns and relationships between data, which begins on page 152, will help you to interpret your data for this solution.

This section on constraints is something you should revisit when you are working on Unit 4, Outcome 1.

Constraints

Constraints are limiting factors or conditions you need to consider when you are designing a solution. A constraint will usually reduce your freedom of design choice. Constraints generally fall into five categories: economic, technical, social, legal and useability (Figure 2.3).

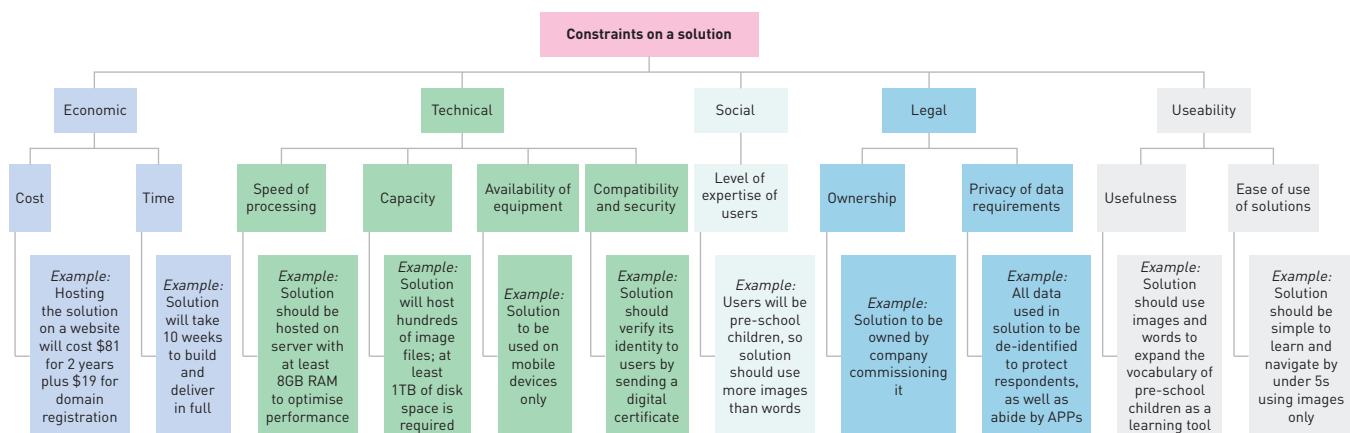


FIGURE 2.3 Constraints on a solution

Many constraints are **economic**. Perhaps your first consideration when designing your solution should be time. Realistically, how much can you achieve in the timeframe you have? Are you assuming you will work more quickly than you can? Consider whether your solution relies on any other parties and whether they have any time constraints as well, such as whether you need to borrow any equipment and for how long. What are the costs of buying, maintaining and upgrading the equipment? What will it cost to pay consultants when you are designing your solution? You also need to factor in the costs of consumables. For example, if you need to do a lot of printing during the design of your solution, consumables might include printer ink or toner.

When you consider **technical** constraints, you need to look at the availability of equipment and what it is capable of achieving. You need to check that the equipment you want to purchase, rent or borrow is available in the early stages of planning. Capacity constraints also apply. Find out how much the system you are using can cope with in terms of disk space, bandwidth, CPU speed and memory. It needs to meet any applicable security requirements that are particular to your solution.

You may have **social** constraints that relate to level of expertise of users – some users may have limitations that have an impact upon how you design your solution, while others may fear changes to existing solutions.

You must also meet any **legal** constraints by ensuring that your solution does not breach any copyright, data privacy or spam laws. You must also not breach any of these laws yourself while developing your solution.

Finally, consider the **useability** aspect of your solution. Is your solution actually useful? Does it serve a purpose? If it serves a purpose, you must still make sure that your intended users will still find it easy enough to use. It is unwise to build a solution for users who will find it unworkable.

Scope

Identify what your solution will be and what you expect it to achieve. This is the scope. Remember that the solution for Unit 3, Outcome 2 is the information you produce from the data you have gathered to support or refute your hypothesis.

Therefore, you need to make sure that your hypothesis clearly states what it includes so that the information you produce does not go beyond the scope. A reasonable hypothesis is very specific so it may be supported or refuted. A hypothesis that is too broad in scope is difficult to support or refute.

The scope of a *solution* is largely defined by its functional and non-functional requirements, and may include lists of functions that are not required. For example, a program to copy files may **explicitly** exclude a user interface from its scope because it will only ever work in the background, unseen by users.

Define the scope precisely so developers can allocate time and resources accurately and know when contractual obligations have been met.



NelsonNet additional resource: Constraints on a solution table template

SCOPE

The following sample hypotheses vary in scope, but all have potential problems. Discuss them in class. For each hypothesis, identify the problem, and then suggest either a replacement hypothesis or how the existing hypothesis could be tweaked to fix the problem.

- 1 Australian native animals are dying out.
- 2 Supermarket fruits and vegetables are having their flavour bred out of them.
- 3 The internet is changing everything.

Once you have settled on a hypothesis and know the scope of your solution, you need to begin collecting data to investigate it. You need to collect both primary and secondary data for your Outcome. The following section of the chapter discusses primary and secondary data.

Primary and secondary data

Data that has not been filtered by interpretation or evaluation is called **primary data**. Often, these are facts that you, the researcher, have collected directly to answer a specific question, but it may also be old data that has never seen proper scrutiny before. The lack of previous interpretation is what categorises something as primary data.

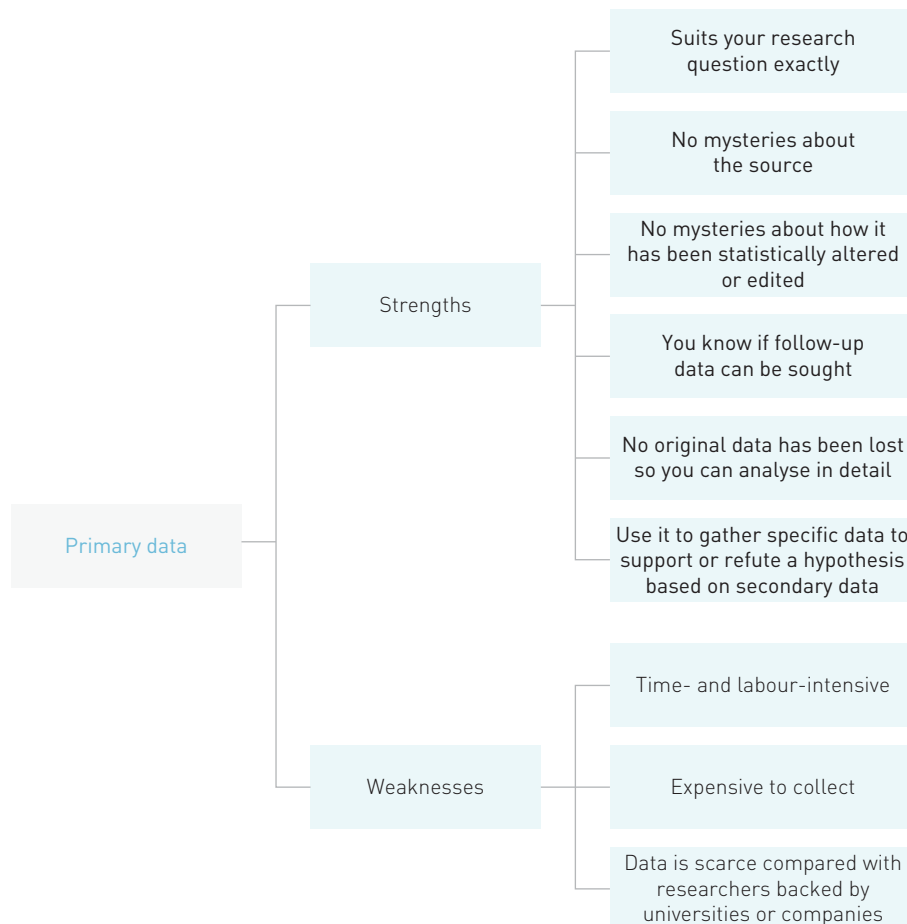


FIGURE 2.4A Strengths and weaknesses of primary data

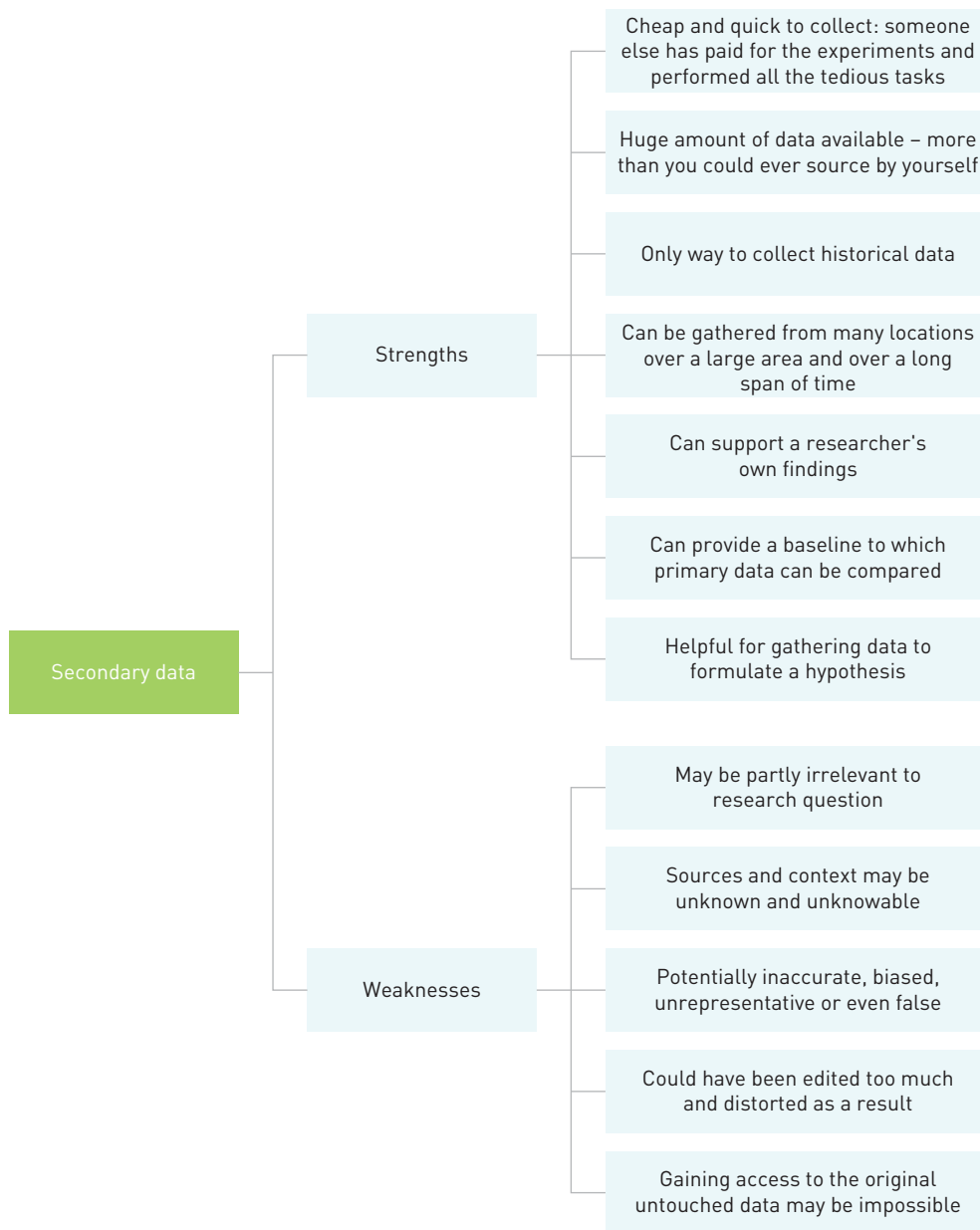


FIGURE 2.4B Strengths and weaknesses of secondary data

Journal articles, results of questionnaires, diaries and letters, emails, internet postings, speeches, audio and video recordings (if unedited), official documents and photos may all be primary data. Therefore, when *you*, as the researcher, need to interpret or evaluate unedited data that you are using in your Outcome, it is primary data.

Researchers also collect (new) primary data using interviews, focus groups, surveys, experiments, observations and measurements.

Secondary data differs from primary data because it has been collected and interpreted by someone *other* than the researcher. The collectors of the data could be other researchers, or government departments, or any of a variety of sources: encyclopaedias and other books, biographers, conductors of polls and surveys, journalists, newspapers and magazines, the Australian Bureau of Statistics, internet posts, databases and so on. When using secondary



NelsonNet additional resource: Potential hypothesis topic and data sources table

data, it is especially important to consider whether both the data and its interpretation are from a reputable source. Therefore, when you collect data for your Outcome that has been interpreted by someone else already, you should consider this data to be secondary data.

POTENTIAL HYPOTHESIS TOPIC AND DATA SOURCES

If have not decided on a hypothesis yet, use this activity to help you.

- 1 List three general topics that you think you may be interested in using for your hypothesis in the table below. You will only end up using one of them, but it is good idea to have a few in mind just in case one or more are not suitable.
- 2 Brainstorm the potential primary and secondary data types for each of your topics and list them in the table. They do not need to be correct or exhaustive. You are just trying to imagine the possible types of data you could gather for the topic if you end up choosing to formulate a hypothesis around it, such as interviews, photos, journals, newspaper articles, interviews and so on.

Potential hypothesis topic 1		Potential hypothesis topic 2		Potential hypothesis topic 3	
Primary data sources	Secondary data sources	Primary data sources	Secondary data sources	Primary data sources	Secondary data sources

When collecting data, the origin of the interpretation tells you whether it is primary or secondary, but this is not the only type of categorisation that applies to data. Data is also categorised in other important ways: one of them is as quantitative or qualitative data.

Quantitative and qualitative data

When you collect primary and secondary data to support your hypothesis, it will be made up of quantitative or qualitative data, or perhaps both. Both quantitative and qualitative data serve a purpose. It will be up to you to choose the correct type of data for each occasion.

If you do not know much about something, but you want it to be the subject of your Outcome, you may consider starting with some small-scale qualitative research. It will help you to hone in on what interests you and what other people generally think and feel about that subject, and it will give you ideas for the quantitative research you need to conduct later.

The easiest way to identify quantitative data and qualitative data is to think of them as *quantity* and *quality*. Quantitative data is concerned with numbers and measurement. It uses an objective approach, closed questions and can be easily scored.

Qualitative data is expressed in words because it is concerned with feelings, personal views and experiences, and opinions. Qualitative data is more difficult to analyse statistically because it has to be coded to be scored. It is subjective.

For example, if you used the following hypothesis:

Do increasing numbers of people who live in Richmond eat a paleo diet because they believe it has health benefits?

You would need to collect both quantitative and qualitative data for this hypothesis: quantitative data to confirm whether increasing **numbers** of people are, in fact, eating a paleo diet in Richmond, and qualitative data to learn if it is because they **believe** it has health benefits.

Quantitative data	Qualitative data
Expressed numerically	Expressed in words
Acquired by measurement	Concerned with opinions, feelings, motives and preferences
Uses scientific approach	Holistic (total picture) approach
Can be easily analysed statistically	Needs to be encoded to be analysed statistically
Collectable in large quantities	Collected on a smaller scale
Objective facts, such as, '8% of RubyMede students play netball for at least two hours per week'	Subjective facts, such as 'I play netball for two hours each week because it helps me to socialise and de-stress, and I don't have to think about schoolwork or or exams'
Can be compared with historical data	Can inform policy development
Can be used to check current status of an ongoing issue	Can be used to delve into ethical or moral issues

FIGURE 2.5 The differences between quantitative and qualitative data

THINK ABOUT
INFORMATICS 2.1

Categorise the scoring of these Olympic events as either qualitative or quantitative, and explain the reason for each choice.

1 Figure skating

2 Speed skating

3 100 m sprint

4 Judo

5 Team show jumping (equestrian)

THINK ABOUT INFORMATICS 2.2

After what you deem to be a particularly horrifying experience with some baked kale chips, you hypothesise that people only like to eat 'superfoods' such as kale because they are fashionable and that no one really likes to eat them.

Describe the sorts of quantitative and qualitative data you would need to find to generate data to support your hypothesis.

The following section discusses coding qualitative data. You will find this useful for your Outcome because the issue you discuss and choose to frame your hypothesis around may relate to emerging trends and shifting patterns. Trends and shifting patterns reflect opinions, which are subjective (qualitative) in nature.

Coding qualitative data

Quantitative data is collected using techniques such as online questionnaires that feature Likert scales, multiple-choice dropdowns and/or radio buttons. It takes a highly structured, digital and numeric form. This makes it easy to score, process and thus convert into information. It is not very difficult to unlock the potential value of quantitative data.

The same cannot be said for qualitative data. Even if qualitative data is gathered online, when you answer a qualitative question, it will often be the type of question that gives the respondent a textbox to write his or her answer. Qualitative data is non-numerical. Answers are expected to be freeform. The idea is that when you ask qualitative questions you want respondents to tell the *complete* truth, so you do not limit the length of answers or the form, patterns or vocabulary they will take. The result is that you cannot score it easily.

After collecting qualitative data for your Outcome, you will most likely need to digitise and code it so you can process and use it as information.

How do you turn this data with all of its potential value into useful information when it cannot be scored very easily? You must transform the data by interpreting and coding it into a summarised form that will help you to analyse it appropriately. The summarised form could be a label, category or simply a number.

Do not try to invent rules for how each and every answer can be transformed into a simple code. This is unwise and would be a very time-consuming task, because qualitative answers can be massively unpredictable in the forms they may take. You could also unwittingly introduce bias and error that would skew the information.

Applying categories or labels may be more helpful than applying numbers to certain types of data. It is important to consider the data output and also the question asked when deciding how to process the data.

QUALITATIVE DATA CODING

Consider this question asked of politicians.

What are your views on same-sex marriage?

- Each response to the question above is listed in the table below. Code each response in the right-hand column as a number from 0 to 5, with 0 = strongly approve of same-sex marriage to 5 = strongly disapprove of same-sex marriage.

Response	Coded as
Look, it's really not for me, but I don't mind if other people do it.	
I have to vote with my party on this one, so I really can't say.	
I think marriage is an outdated institution and no one should do it – same-sex couples or otherwise.	
The Bible's pretty clear about this one – marriage is between two people. A man and a woman. That's it.	
I'm pretty disgusted by the fact Australia is so far behind the rest of the world at this point. I'm embarrassed for us that we're still not giving same-sex couples the right to live, love and marry as they want.	





This is a slippery slope. Allow marriage between same-sex couples and suddenly you'll be allowing marriage between a man and a child, or between people and animals. It's just not right!	
I'm annoyed that this has become an issue and more important things have been put on the backburner. Of course same-sex marriage should be allowed. But it shouldn't be the massive issue it has become!	
Marriage is a religious institution that just happens to have had legal trappings attached to it. It would be better if we permitted civil unions for same-sex couples instead.	

- 2 What criteria did you use to turn a thought into a single number? Did you find it very difficult to apply the numbers?
- 3 The detail and nuance in the responses are stripped back by coding each response as a number. There are other ways of coding qualitative data, such as categorising and labelling. How might you have coded the responses to better represent their content?

CASE STUDY

READY SET GO! GYM

A group of 28 people who all belong to the Ready Set Go! gym indicate to management that they do not want to renew their membership within the same month. The manager of the gym, Kym Carr, is alarmed by this news.

While it is normal policy to try to upsell a new membership, or send out a (quantitative) questionnaire, this has not been very successful. The upselling is typically met with aggression or avoidance and the questionnaires do not get returned. Kym has been taking some psychology courses that have taught her a bit about statistics and interview techniques, so decides to do something different – she schedules an interview with each member, in which she collects qualitative data to find out why they are leaving the gym. When scheduling the interview, she indicates that she wants to fix the problems with the gym and will not try to sell them anything unless it seems like there is something feasible that they may want.

She asks each member the following two questions only, and lets each member talk as much as they want.

- 1 Why are you leaving the gym?
- 2 How could we convince you to stay?

Each interview lasts for approximately half an hour and, to Kym, it feels very private and different to the appointments with members that she usually has. She feels as though the members are opening up to her in ways they normally would not. At certain points, she feels almost as if she is hearing their secrets. She takes notes, but does not interrupt them unless she needs them to further explain something she feels she may have misunderstood.

When I first started I had a lot of enthusiasm and energy, you know, but as the weeks went on, I was coming by myself and it just got harder. And then I weighed myself and nothing was really happening, and that made it worse. I feel pretty isolated when I do come in, and I'm sorry, but you've got that one really rude guy who works here and makes fun of clients within earshot of other people. I find myself worried I'm going to come in on a night when he's working and he's going to see me and make fun. I mean, you realise that's just not on, right? The only machines I really like are the cross trainers and the rowing and they're not looked after properly. I just feel a bit let down. I feel like crap when I come here so I don't want to come here. I'm meant to come here so I can start to feel better about myself, not worse. And it's not like I have anyone to talk to about it.

Case study continues ►

◀ Case study continued

	A	B	C
1	Reason	Mentioned	Notes
2	The cost of membership	5	Maybe do 20% discount as incentive?
3	Poor state of equipment	10	Which equipment? X-trainers, bikes in spin-cycle rooms, some of the barbells, rowing machine, stepper
4	Unpleasant smell from weight room	13	Apparently it smells a bit like mould but it's pretty potent
5	Too much chlorine in the spa and pool	4	Member actually said it hurts his eyes
6	Lack of appealing classes	15	Shortage of good yoga, pilates etc... Not enough variety in the aerobics classes and also we just don't do enough in the aquatics classes full stop
7	Unprofessional behaviour from staff	12	Eric is mentioned 9 times – he needs to go. Notify HR.
8	Low motivation	7	
9	Depression	2	
10	Lack of results	15	Not enough increased muscle mass or significant weight loss within expected timeframe. Why aren't my staff monitoring this? Link to depression and low motivation maybe?

FIGURE 2.6 Reasons for leaving the gym

After all of the interviews, Kym makes a spreadsheet list of the basic reasons why the members want to leave the gym.

She notes that during the interviews, three of the people who wanted to cancel their memberships had expressed deep disappointment over the employee who has been unprofessional and linked their lack of results to him. One of the members who mentioned depression also mentioned lack of results, cost of membership and unprofessional behaviour of gym staff.

Kym decides to apply categories to the data to make it easier to understand the issues.

13	Reason	Mentioned	Categories
14	The cost of membership	5	economic concerns
15	Poor state of equipment	10	physical environment; safety (high priority)
16	Unpleasant smell from weight room	13	physical environment
17	Too much chlorine in the spa and pool	4	physical environment; safety (high priority)
18	Lack of appealing classes	15	physical environment; lack of support; lack of variety
19	Unprofessional behaviour from staff	12	staffing concerns; safety (high priority)
20	Low motivation	7	personal concerns
21	Depression	2	personal concerns
22	Lack of results	15	personal concerns

FIGURE 2.7 Reasons for leaving with categories applied

25	Reason to stay	Mentioned	Categories
26	Free membership	9	economic concerns
27	Fire Eric	9	staffing concerns; safety (high priority)
28	Fix the equipment	10	physical environment; safety (high priority)
29	Fix the pool and spa	3	physical environment; safety (high priority)
30	Get rid of the smell	12	physical environment
31	More/better classes	12	variety
32	Provide more support	12	physical environment; support
33	Nothing you can do?	15	personal concerns

FIGURE 2.8 Reasons to stay at the gym with categories applied

◀ Case study continued

Kym soon realises that the ‘physical environment’ and safety of the gym is the most significant problem. As a manager in a health-related industry, Kym considers safety to be its own category and automatically views anything mentioned within it as a high priority. Applying categories makes a huge difference.

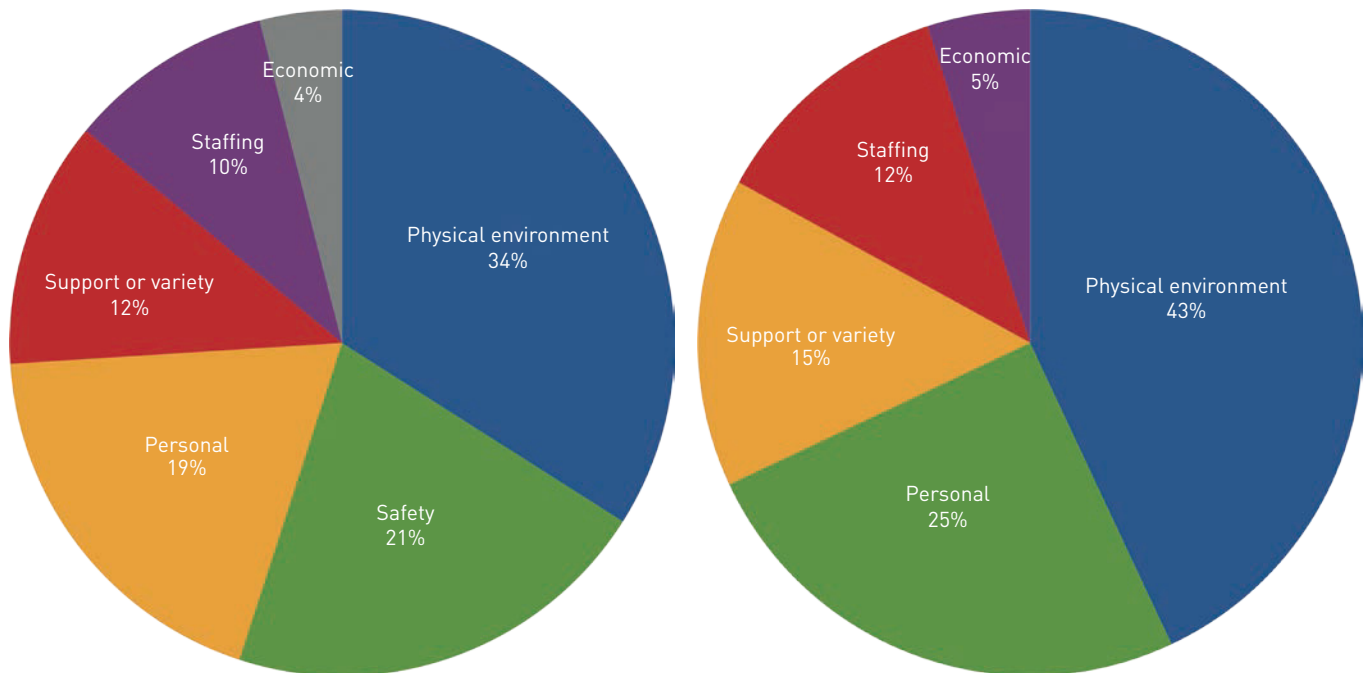


FIGURE 2.9 The pie chart on the left shows safety as its own category. It is not really correct, because the staffing issues are all a safety issue, and most of the physical environment issues also overlap with safety. The pie chart on the right does not show safety at all. Rather, it groups items by category and allows Kym to assign safety as a priority to them as needed. How else could you graph this data to show safety without providing a misleading tally?

Kym is able to bring in maintenance staff to look at the equipment, investigate the smell in the weight room and deal with the levels of chlorine in the spa and pool.

Another significant category is ‘staffing concerns’, which also overlaps with safety. Kym decides to fire the employee who has been mentioned repeatedly as problematic because the complaints are very serious and he has been warned twice in the past. She offers a free renewal of their memberships to those members who complained about that staff member.

Kym feels a bit despondent that she has so many people falling under the ‘nothing you can do’ category – personal concerns. She notes that all of the people who said they had low motivation also said they came to the gym by themselves to work out.

After some brainstorming with her staff, Kym contacts all the members with ‘personal concerns’ (depression, low motivation and lack of results) and asks them if they would consider opting in to her new buddy system program. This pairs members with similar schedules together and also assigns them to a gym instructor as a mentor to monitor their progress and provide them with a more hands-on approach so they do not feel alone.

Kym also contacts a few local TAFEs and vocational schools to find recently qualified yoga and Pilates instructors to add new classes to the gym, and volunteers to run two more aquatics classes herself each week.

In all, Kym is only able to convince 17 of the 28 members to keep their memberships, but this is a substantial improvement over the usual percentage. It has also allowed her to weed out an employee who made members uncomfortable, fix broken equipment and help struggling members, which would not have been possible with a standard checkbox questionnaire in the mail.

THINK ABOUT INFORMATICS 2.3

Qualitative coding is difficult, but so is qualitative scoring.

In an examination consisting of multiple-choice and short-answer questions, how is each type of question marked? How do examiners ensure fair marking of short-answer questions?

Consider an English examination's essay on a novel: you may wonder how the many qualities of an essay can be converted into a single number that can fairly be compared with thousands of essays by other students. In short, how can one essay marker's performance be detected as excellent, competent or incompetent?

When essays or short answers are assessed in VCE examinations, the assessors are provided with very clear marking guidelines, and these apply to both holistic and analytical questions. All questions are marked twice and the level of discrepancy between scores on any one question is very small. There are also strict procedures to ensure that assessors are within an acceptable range on designated questions. For example, if a mark is outside an acceptable range the assessor must discuss their scoring with the Chief Assessor to identify the marking differences.

◀ Case study continued

Further, Kym is excited by the prospect of her new buddy system and mentor program, which will help gym members who usually work out alone to get to know someone with similar concerns while also helping them monitor their progress. Kym is pleased that she conducted the qualitative interviews, even if they did take much longer than quantitative multiple-choice questionnaires.

The following month, five more members drop their memberships, but there is also an upswing in new memberships, and Kym's efforts with the buddy program earn Ready Set Go! a small mention in the local paper.

Qualitative coding can be valuable to *create* categories. Until a survey is conducted, researchers may have little idea of the general categories of problems.

Quantitative research can then apply pre-determined categories. It usually does not allow respondents to add new categories.

Coding can reduce the original wordiness to a more manageable form using freely chosen summary terms, which is sometimes called **descriptive coding**.

Coding can also replace entire ideas with symbols from a fixed set of summary numbers, codes or labels, which is known as **analytic coding** or **theoretical coding**.

Encoded qualitative data allows responses to be grouped, averaged, totalled and compared. Once labels and categories have been assigned, it is easy to find individuals who meet certain criteria, and to determine patterns. For example, to find out whether student trust is related to the teacher's experience (Figure 2.10), run a query to find all respondents with the 'Trust' code and get an average of how long their teachers have been teaching. Look up counts of whether individual teachers have fewer than five years of classroom experience, more than five years of classroom experience and more than 10 years of classroom experience, to see how often 'Experienced' and 'Trust' are commonly combined.

- 1 She cares about my education, so I trust her.
 - 2 Plus, she has loads of experience – she's been around here for years and years. She totally knows what she is doing when it comes to helping me get good grades on my Outcomes.
 - 3 She always makes time to answer my questions when I get in a panic, which is often!
- 1 = Trust 2 = Experience 3 = Availability

FIGURE 2.10 Descriptive coding of qualitative data about teachers

Rubric

While a single movie reviewer will have personal, internal criteria for assigning stars, teams of reviewers need to use a shared assessment scheme so their ratings are consistent and comparable. When coding is done by many people, the coded results could be unreliable if they have differing interpretations of answers. Groups of coders, such as VCE exam markers, need to have a clear and shared understanding of how to interpret, evaluate and encode responses. One way to give meaning to codes is to use a rubric: a detailed list of descriptive grading criteria that correspond with a code (the mark). Teachers assess VCE Outcomes in this way.

TABLE 2.2 Grading criteria rubric

33–40 marks	Thorough and insightful analysis correctly identifies all the website requirements of an online community and acknowledges all relevant technical and non-technical constraints (8–10 marks).	All selected design tools are appropriate (8–10 marks).	Correct design techniques are consistently applied to fully represent the functionality and appearance of a website that is feasible and accordant with the analysis (8–10 marks).	All manual and electronic validation techniques effectively check the reasonableness of data (8–10 marks).
Descriptions are given for 25–32 marks, 17–24 marks, 9–16 marks and finally ...				
1–8 marks	Identifies few website requirements of an online community. Few technical or non-technical constraints are stated (0–2 marks).	Limited relevant design tools are selected and limited design techniques are applied to outline minimal functionality and appearance features of the website (0–2 marks).	Few links exist between the design and the analysis (0–2 marks).	Limited application of manual or electronic validation techniques affects the quality of critical data and hence the solution (0–2 marks).

Now that you know whether the data is primary or secondary, and quantitative or qualitative, the question becomes how best to gather it for your Outcome. This chapter deals with two main ways of acquiring data: through experimentation in the form of methods such as observation and interviews, and through data sources such as libraries and data sets. You should try to use both of these for your Outcome if possible. The next section deals with data sources such as libraries and data sets.

Data sources

Deciding on a hypothesis and finding relevant data sources takes thought, judgement and care. A great deal of data is available online. The trick is to find the data that is worth using and to use it correctly. While Chapter 3 will deal with the forming of your hypothesis, this chapter is about the nature of data itself, which will ultimately support or refute it.

It is important that you investigate both primary and secondary sources. It is also important that you choose the best data you can because in Unit 4, Outcome 1, your findings about your hypothesis for this Outcome will be presented in a multimodal online solution (MMOS) suitable for a global audience.

The source of data is very important. Treat data without an identified source with care and take necessary steps to resolve matters of authenticity. When data has provenance it means the **chain of custody** is known and therefore it is more reliable. Without provenance, your trust would be blind because you would not be able to contact the data's creators – the data could even be a work of fiction. Putting complete trust in anonymous data is unwise.

However, if you know how data has been collected, statistically manipulated, edited and/or abridged, you can interpret the information more wisely. A reputable data source with **authority** is more likely to provide high-quality data. Be wary of data given by people who *are* experts, but *not* necessarily experts in the relevant field. The opinion of a champion footballer on politics, or an actor on climate change, are no better than any other person's.

There are hundreds of data sites in Victoria and Australia on many topics. Choose a topic that interests you. Search online for key words relevant to your topic and add 'data set'.

You will probably find data that will inspire many reasonable and interesting hypotheses.

Libraries

On a local level, **local and school libraries** are useful data sources. They will have many printed, **multimodal** and online resources, including encyclopaedias, textbooks and other reference books that are both in print and online. They also keep collections of government reports and publications, as well as autobiographies, video documentaries, oral histories, newspapers, magazines and brochures.

Trove is the National Library of Australia's searchable online newspaper archive. It has more than 270 million Australian and online resources.



Trove

FIGURE 2.11 Trove, the National Library of Australia's online data store



Hansard



Annual reports: government and parliament



Data.gov.au



Melbourne Data

Government

For local government sources, you could start with **Melbourne Data**. This is where the City of Melbourne shares its data, from parking bay sensor recordings to tree canopy data and pedestrian sensor counts.

You can also look at **Hansard** for transcripts of parliamentary debates. Annual reports are produced by all government bodies and these may be useful to you as well. **GovPond** is a portal of open government data that searches over 3400 Australian data sets. **Data.gov.au** offers public data sets from many locations, groups and topics, including a national public toilet map, overseas arrivals and departures, taxation statistics and even Victorian ice skating centres.

Government agencies

The **Australian Bureau of Statistics (ABS)** is tailored to students with census data that is available in Excel format and tables. The **Australian National Data Service** provides data collections for Australian researchers.

CSIRO's Data Access Portal is a very large, detailed and technical science data store. The **Atlas of Living Australia** shares biodiversity knowledge and has a large number of data sets, which is useful if you have a particular hypothesis; for example, how housing developments might affect Brushtail possums.



CSIRO



ABS



Atlas of Living Australia

The screenshot shows the 'Atlas of Living Australia' website. The browser address bar displays 'collections.ala.org.au/datasets#filters=status%3AdataAvailable'. The website header includes the logo, navigation links (ALA Apps, ALA Info), a search bar, and a 'Search' button. Below the header, the breadcrumb 'Home → Datasets' is visible. The main heading is 'Atlas of Living Australia data sets'. A paragraph explains that content comes from various data sets provided by institutions, collectors, and community groups. Below this, instructions guide the user on how to refine results using filters and how to view descriptions or full metadata. The main content area shows 'Showing 718 data sets.' with search, reset, and download buttons. It includes dropdowns for 'Results per page' (set to 20), 'Sort by' (set to name), and 'Sort order' (set to ascending). A left sidebar titled 'Refine Results' shows 'Current Filters' with 'Status: Data available [X]' and 'Resource type' with options like Records (510), Species-list (170), Website (27), Document (8), and Uploads (3). The 'License type' section shows Custom license (473), No information (128), and CC BY (55). The main list displays three data sets: '2WEL-01' (species-list, other license), 'ABBBS Bird Banding records from the Australian Antarctic Territory and Heard Island.' (records, other license, with a 'View records' link), and 'ABRS Flora of Australia Online'.

FIGURE 2.12 The data sets page on the Atlas of Living Australia website, with a filter applied just to show those data sets with data available on the website

Other options

Local historical societies and newspapers also make good data sources. You might consider conducting interviews with local people for primary data if your hypothesis relates to your local area. You could try looking at **online forums or discussion groups** on special interests, such as Yahoo! groups, Usenet and Google groups, but it is important to be very discriminating about these.



Most newspapers now offer free or paid access to their online archives, including Online Newspapers, which links to many local and international news sites.

Once you have located the correct data sets, you will need to reference the data sources. Refer to the ‘Referencing data sources’ section of this chapter for more information. The following section of this chapter moves onto the other major method of gathering data for your Outcome: acquiring it directly.

Acquiring data

Primary and secondary data can both be acquired using a variety of methods including observation, interviews, surveys and questionnaires. When you need to acquire data, you have to start by asking questions. You need to understand how questions work before you begin to formulate them for your Outcome so that you can gather the right data to support or refute your hypothesis.

Asking questions

Questions take two basic forms: closed and open. **Closed questions** are usually found in surveys, exams and forms. They force you to answer in a particular way because they allow only pre-determined responses, such as yes or no. For example, ‘Would you prefer Chinese or Mexican food for lunch?’ only offers you the choice of those two options, so it is a closed question. Closed questions are good for collecting facts and numbers – what you already know as quantitative data. **Open questions** are often used in interviews. Open questions do not limit the answer; there is no pre-determined set of options, nor a limit on length, type or content. Instead of being asked which of two things you would like for lunch, an open question would ask, ‘What would you like for lunch?’ Open questions are better for collecting data that relates to opinions and subjective experiences – or qualitative data.

Closed questions

Having a limited choice of responses may prevent respondents from giving detail or reasons in their answers, but it also makes it quicker and easier for them to respond and participate. It will also be easier for you to collate and analyse a set of answers drawn from closed questions than wade through the answers elicited from open questions. It is relatively easy to set up a questionnaire based on closed questions – you can do it yourself online.

Closed questions can be written and formatted in a number of ways; for example, they could use radio buttons, multiple-choice or Likert scales (Table 2.3).

TABLE 2.3 Formats for closed questions

Format	Example
Radio button	☐ Yes ☐ No
Multiple choice	Which form of public transport do you prefer? A Train B Tram C Bus D Teleportation
Likert scale	Strongly agree Agree Undecided Disagree Strongly disagree How often do you exercise? Daily Weekly Monthly Never On a scale of 1 to 10, with 1 being ‘Not at all’ and 10 being ‘Very strongly’, how strongly do you believe that crepes are the best breakfast food?
Ranked priorities	Place numbers 1–10 to indicate which forms of transport you prefer from most favourite to least favourite.

A particular problem with closed questions is that sometimes none of the available answers apply to a respondent. If this happens, the respondent may quit without completing the survey, or choose to give you inaccurate answers. For example, consider a closed question such as:

How often do you attend church? Frequently | Sometimes | Rarely

By not providing the respondent with a 'Never' option, the question builds bias into the information derived from the answers – suggesting specifically that all of the respondents attend *church* from time to time, and therefore that all of the respondents are to some degree religious and are Christians.

To formulate a good closed question, you do need to have a bit of foresight to predict all of the applicable options for answers. Given that no one can be expected to be psychic, a good way to come up with these is to conduct preliminary research using open questions to gain knowledge about the topic. If you do not understand your topic, you cannot formulate successful questions and answers about it, so you need to explore it first. Once you understand a topic better with open questions, you can create closed questions with a range of all-encompassing answers.

As discussed in Chapter 1, databases allow data fields to be made mandatory, meaning that a record will not be accepted while those fields are left empty. This is often the case with closed question fields. Only make fields mandatory if further processing would be pointless if those fields were missing; for example, a missing email address when registering for a mailing list.

Open questions

When you ask someone an open question, you impose no restriction on their answer. When you have conversations with friends, or go for a job interview, you are often asked open questions where you give detailed, individual answers. When you apply for your first job, it will not be unusual to hear an open question, such as:

Can you tell me about a time when you had to solve a difficult problem with minimal resources and no outside help?

Answers to open questions can be lengthy, rich, vivid and unique. Answers may even critique the question itself. When you write an open question or answer one, you are given an opportunity for greater accuracy than with a closed question, but you should keep the pitfalls in mind.

Open questions become less useful as the number of respondents increases. Analysing the answers to hundreds of unrestricted responses is a time-consuming task. It requires a great deal of personal judgement and processing effort to convert the answers into a form that can be statistically counted. Researchers use **coding** to convert the answers to open questions into more manageable codes that are easier to process statistically. For instance, how would you store an answer like the following?

Well, I'm not entirely sure I'm comfortable with – well, I mean, I guess I'm okay with it. I don't disagree with it. I suppose I'm neutral. Actually, look ... it's a personal bugbear for me and mine, but I don't care about it for other people.

A closed question that offered only yes and no as options would have forced the respondent to commit to a definite answer that was easier to count and average. However, the ambiguity, delay and depth of thought shown in the original answer would have been lost, and that in itself may have been informative to the researchers.

IDENTIFYING QUESTIONS

The following questions are a mix of different types of questions. Not all of the questions are suitable for research, but understanding whether they are quantitative or qualitative, and closed or open will help you to understand how to structure your own research questions.

		Quantitative or qualitative?	Closed or open?
1	What do you enjoy most about studying Informatics?		
2	Which subjects do you think prepare you the most for life after VCE?		
3	Do you prefer Windows, Mac or Linux operating systems?		
4	What do you think you need to most improve on before taking VCE exams?		
5	Why did you enrol in Informatics?		
6	What do you think will be the most difficult part of Unit 3, Outcome 2?		
7	Why do you prefer to use my computer instead of yours? It's right there!		
8	How long do you think it will take you to formulate a hypothesis?		
9	How many computing devices do you use at home?		
10	When will you be home tonight after netball?		

Faults in questions

Questions can be flawed regardless of whether they are open or closed.

A **loaded question** contains controversial or unjustified assumptions that may serve to support the **hidden agenda** of the person asking the question. For example, the question, 'Why do you think immigrants are involved in so much crime in Australia?' is a loaded question because it asks you to assume from the beginning that immigrants are involved in a lot of crime. Similarly, 'How did you feel when the red car slammed into the blue car?' is a loaded question, because it implies not only that the red car was not only responsible for the collision, but also that it was speeding at the time. Both of these questions try not to give you room to suggest something else – that immigrants are no more or less law-abiding than anyone else, or that perhaps the blue car slammed into the law-abiding red car.

A **leading question** directs respondents to answer in certain desired directions. For example, the question, 'Do you think murderers should get life in prison or the death penalty?' directs the respondent away from suggesting any other alternative penalty. Similarly, the question, 'Do you think mothers or fathers make better single parents?' tries to suggest that one gender is *always* better than the other.

Most courts of law have strict rules controlling the use of leading and loaded questions by lawyers when questioning witnesses. The following question would raise an objection for being a leading question:

Did you see the accused come out of the house at 2 a.m. on the fifth of May this year?

But it *would* be appropriate to ask:

Did you see the accused on the fifth of May this year?

Then:

Where did you see him?

And:

When did you see him?

This question would also raise an objection for being a leading question:

Was the accused wearing a red jumper that day?

But you *may* ask:

What was the accused wearing that day?

Conducting interviews

Interviews are real-time personal interactions between the data gatherer (you) and data providers (the respondent). Unlike a survey, an interview can be shaped by the answers given. You can pursue a line of questioning further to elicit more information upon receiving an interesting answer. The presence of the interviewer can subtly change the attitude of the interviewee. Wanting to be more helpful and avoiding disappointing the interviewer, the interviewee is often more cooperative and less likely to skip questions than he or she would be in a survey.

Using the face-to-face technique, an interviewer can pick up on an interviewee's tone of voice – detecting interest or disinterest, anger or happiness – during the interaction. Body language, facial expressions, pauses, hesitancy, energy, lethargy and other non-verbal elements of communication are all there to be observed. Personal interviews are more likely to elicit accurate information on sensitive subjects than impersonal surveys.

To make the most of the benefits, interviews are best conducted face-to-face, but this can be time-consuming and expensive. The next best alternative technique is internet video conferencing (such as Skype or Google Hangouts), which offers some of the benefits of face-to-face questioning, but is cheaper and more convenient. Phone interviews are the last choice of technique for personal interviews because all visual cues and intimacy have been lost.

Conversely, a potential negative of interviews is that interviewees may sometimes try to say the 'right' thing to the interviewer, to give the correct answer or avoid embarrassment. For this reason, observation is also a useful method of gathering data.

Interview technique

You can prepare yourself for your interviews by following a few simple guidelines. These are not hard and fast rules, but they may help you to set the right tone.

- Introduce yourself and clarify the purpose of the interview and the way the information will be used.
- Start with easy, quick questions to put nervous interviewees at ease.
- If the information may be personal, follow the spirit of the Australian Privacy Principles (APPs) even if you are not subject to the Privacy Act.
- Have a list of questions prepared so you do not forget them.
- Avoid getting side-tracked by irrelevant discussion unless it is informative.

- Give interviewees time to think and talk.
- Do not interrupt answers unless it is obvious that they are irrelevant or repetitive.
- Gather the data you want to collect, but leave room for the interviewee to volunteer extra information. This can be done by including open questions, and finishing the interview by asking if there is anything else they did not get an opportunity to say.
- Record interviews for later reference (with the interviewee's prior permission). Writing down answers may distract you from listening actively and closely.

Try to keep in mind that interviews are not foolproof. Sometimes subjects can be unwilling to provide data voluntarily, especially if they think it may be embarrassing, uncomfortable or will have unwanted consequences. For this reason, it may be necessary to use another method to acquire data directly, such as observation.

Observing subjects

Observing people behaving naturally can be very informative, because people do not always report their own behaviour accurately. Sometimes this is because people know there is a correct, desired answer to a question. When your parents ask you if you have cleaned your room, you know the correct, desired answer is yes. You know the same thing when they ask if you have done your homework.

Similarly, when trying to reduce infection rates in hospitals, a survey question such as, 'Do you always wash your hands after going to the toilet?' will generally elicit 100% yes answers. Observing actual behaviour may disclose a truth far from the survey results. Asking office workers to report how long they spend playing games instead of working is less reliable than actually observing them.

When impartial interpretation of behaviour is needed, observation is an important method. If studying a university's performance, shedding light on whether teachers deal with their students in a fair, friendly and open manner is a **subjective** assessment that should not be left to the teachers being evaluated.

You can also observe people to investigate the interaction between them, particularly their non-verbal communication. Like interviews, observation allows you to explore non-verbal communication in a way that other forms of data gathering do not.

Subjects are not always capable of reporting on themselves with clarity, and, because they are reporting on themselves, they are biased. You, as observer, may notice things about your subject that they cannot. Older observers may detect youthful, immature behaviour in younger subjects; female observers may detect sexist behaviour in male subjects. Observation by an outsider can yield valuable data.

Observation may be overt, where subjects are aware they are being watched, or covert, where they are unaware. Covert observation (which may involve using cameras and is sometimes referred to as spying) is more accurate, but is also more difficult to accomplish. It can also be unethical or illegal.

Observation can be direct, where the observer watches behaviours as they occur; for example, watching students as they eat in a cafeteria. It can also be indirect, where the results of the behaviours are observed; for example, measuring the food left uneaten after the students from the cafeteria have finished their meals.

Observation may have electronic aspects, or it may be conducted completely electronically. Techniques for electronic observation include activity logs, key and screen logging software, and equipment that tracks user eye movements when they are using software, watching TV or using other products.

Observation approaches

Prepare

- If observation is overt, gain the trust, consent and confidence of your subjects by explaining the purpose of the observation.
- Prepare objective measures such as recording sheets and checklists. This will help to reduce the effect of observer bias, record data in a consistent format and remind observers what to look for. For instance, a personal judgement of whether a teacher treats male and female students relies on both the quality of an observer's opinion, and interpretation of what 'equal treatment' means. More reliable would be a count of how often the teacher responds to hands raised by males and females, and a measurement of the average amount of time spent helping male students compared with female students.
- If appropriate, reassure subjects that they will not be judged or punished. Offer to share the research results with them.

Record

- Keep ongoing records, notes and recordings. Relying on your memory is unreliable.
- Record exact quotes, not your interpretation of what was said.
- Describe events without interpretation or evaluation. Interpretation happens later.
- Record **field notes**, which are open-ended recordings of things you observed that were not anticipated when research began. They add detail but do not answer pre-planned questions.

Monitor (behaviour)

- Do not let your observing distort the natural behaviour of your subjects. You must not interfere.
- Remain professionally detached. While developing rapport with subjects is desirable, becoming emotionally attached to them can distort an observer's objective judgement.
- Separate your own thoughts, opinions, expectations, experiences and assumptions from what you actually see.

Disadvantages of observation

Although there are many benefits of observation, negatives exist as well. Observation consumes a great deal of time, money and labour. It indicates what subjects *do*, but not *why* they do it. People (subjects) also tend to perform better when they know they are being monitored (overt observation).

You may be susceptible to observer bias when personal judgements are called for; for example, 'Did male workers interact badly with female customers?' Male and female observers could interpret 'badly' very differently.

Given that both observation and interviews can be time-consuming and costly to undertake, you may wish to consider using less demanding methods of data gathering as well, such as surveys and questionnaires.

Surveys and questionnaires

A survey usually gathers primary data, such as opinions, on a variety of broad issues. It is often used when research begins and little is known about the subject. A survey is also a term of reference for a larger study. A questionnaire is specifically a list of written questions, usually devised for statistical purposes. Ultimately, the *set of questions* you write for your survey will be a questionnaire.

In this textbook, when referring to participants in observation and interviews, the term **subject** is used. When referring to participants in surveys and questionnaires, the term **respondent** is used. This difference in terminology has been adopted deliberately, so that you can immediately tell the difference between individuals who participate in research quite directly, by interviews or observation, and those who participate less directly by responding to a survey. The main reason for the distinction is that you are more likely to have opportunities for follow-up with subjects than with respondents, so it is wise to distinguish between them.

Advantages

Questionnaires can be widely distributed in great quantities, particularly online. As a result of this, you can gather a great deal of data quickly and cheaply, and you can compile, process and analyse the responses with spreadsheets, databases or other software. Responses can be counted, averaged and compared easily because everyone is asked identical questions. Without the presence of a questioner, observer bias is removed.

Disadvantages

Questionnaires cannot answer queries from respondents. Likewise, follow-up questioning from researcher to respondent may be impossible. Respondents may not bother completing lengthy written responses. It may be hard or impossible to encourage compliance. If participation is purely voluntary, the only respondents who submit answers may be those with strong opinions on the research topic, thereby introducing a **self-selection bias** into the data. This means you would end up with results suggesting that everyone feels very passionately about your topic, because people who are neutral or indifferent did not feel compelled to respond to the poll.

Similar to the problems with interviews, respondents are more likely to lie or not answer when it comes to sensitive or embarrassing questions. They may also interpret questions differently. For example, consider the question 'Do you visit the supermarket frequently, sometimes, or rarely?' To one person, 'frequently' may mean daily, but to another, it may mean weekly or fortnightly. Responses usually lack explanation or detail about context – statistical results describe average trends, but say little about individual motives or feelings.

Writing questionnaires

When you write your questionnaire, you should know what data you need to gather because you should have a clear focus. Your questions should be specific and concise because vague, wordy questions will yield ambiguous answers and they will confuse your respondents.

Do not allow leading and loaded questions to creep into your questions. You are a researcher seeking the truth, not an artist seeking to create a piece of work. Refer back to page 96 for the discussion of leading and loaded questions.

Protect the privacy of your respondents. Make sure you have not written any questions that are unnecessarily personal in nature. Do not ask respondents about sensitive issues unless it is relevant and crucial.

You may decide that some of your questions should be closed but some should be open. Refer back to pages 94–5 for the discussion on closed and open questions.

- Closed questions are appropriate for basic information and are easy for respondents to answer.
- Open questions probe issues more deeply, but require more time and effort to answer. Use them sparingly. Reserve them for important questions, because pages of blank lines or large, empty textboxes for too many questions can make a respondent lose patience and tempt them to stop participating.

Road test your questions to get feedback before you launch your questionnaire. Ask the respondents who have agreed to help you to road test your questions:

- Are the questions clear and logical?
- Is the language coherent and clear?
- Are the questions respectful of (your) privacy?

Determine whether the questions have been eliciting the responses (data) you need. Find out if any questions have been skipped, and, if so, why they have been skipped.

When you write your final questionnaire, include important information such as the purpose of the research, assurances of confidentiality and acknowledgements to any people or organisations who assisted you in the development of your questionnaire and the larger survey.

Use Figure 2.12 as a guide to writing a questionnaire.

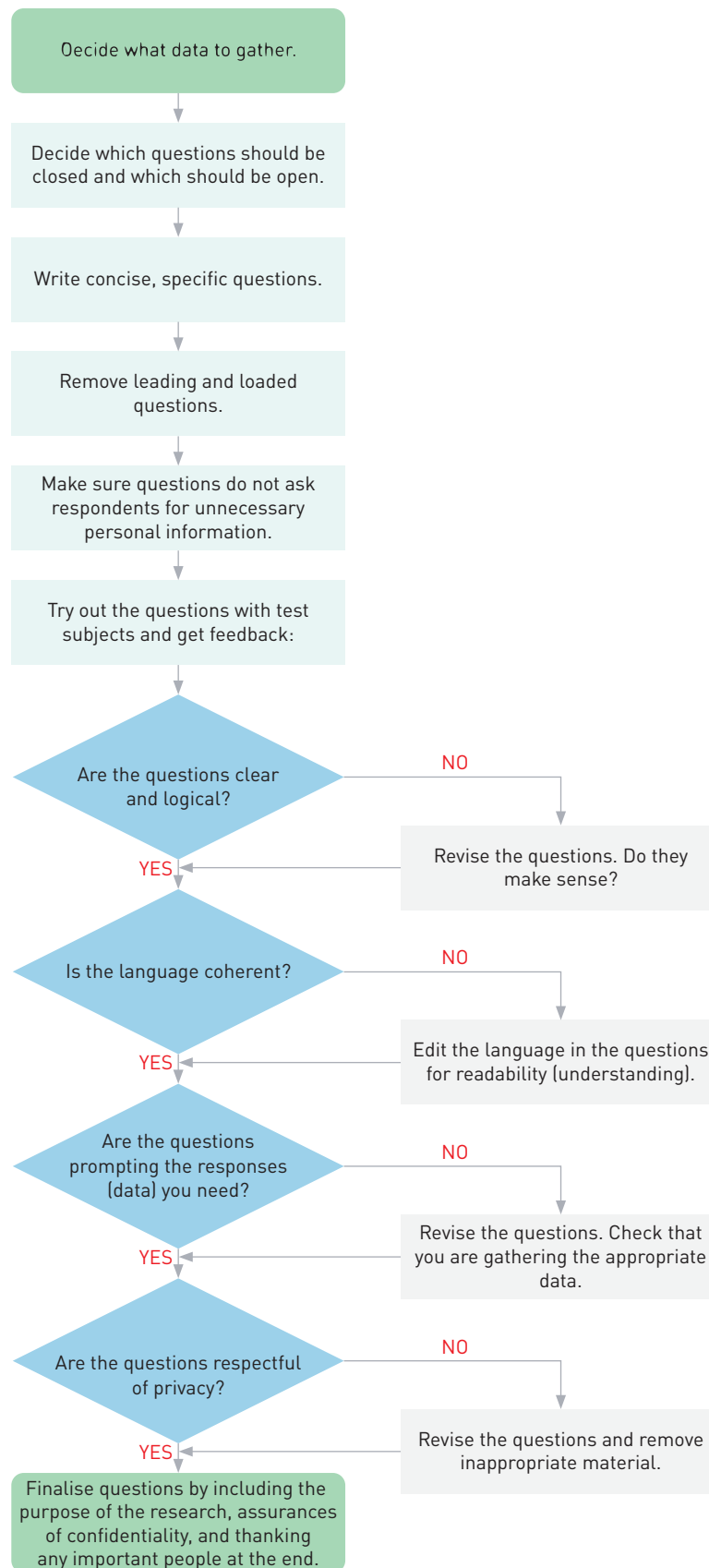


FIGURE 2.13 The process for writing a questionnaire

Paleo diet questionnaire ☆

File Edit View Insert Responses (1) Tools Add-ons Help All changes saved in Drive **Send form**

↶ ↷ **Edit questions** Change theme ↶ View responses View live form ↗

▼ Form Settings

- ☐ Show progress bar at the bottom of form pages
- ☐ Only allow one response per person (requires login) ?
- ☐ Shuffle question order ?

Page 1 of 1

Paleo diet questionnaire

Attitudes to the paleolithic diet

Question Title **Which of the following most closely matches your opinion of the 'paleo diet'?**

Help Text

Question Type **Multiple choice** ☐ Go to page based on answer

- ☐ It has some good ideas and some bad ideas
- ☐ I don't know anything about it, so I can't say
- ☐ I think it's great.
- ☐ It is total bunk.
- ☐ Click to add option or Add "Other"

▶ Advanced settings

Done ☒ Required question

How likely are you to order a meal off a menu when it is labelled 'paleo'?*

0 1 2 3 4 5

Extremely unlikely ☐ ☐ ☐ ☐ ☐ ☐ Very likely

Which of the following does the paleo diet typically exclude?*

- ☐ Dairy
- ☐ Grains or cereals, such as wheat and rice

FIGURE 2.14 The 'Edit questions' view of a questionnaire created using Google Forms. Notice that several types of questions are available for you to use to collect data.

Paleo diet questionnaire

Attitudes to the paleolithic diet

* Required

Which of the following most closely matches your opinion of the 'paleo diet'? *

- ☐ It has some good ideas and some bad ideas, like many diets.
- ☐ I don't know anything about it, so I can't say.
- ☐ I think it's great.
- ☐ It is total bunk.

How likely are you to order a meal off a menu when it is labelled 'paleo'? *

0 1 2 3 4 5

Extremely unlikely ☐ ☐ ☐ ☐ ☐ ☐ Very likely

Which of the following does the paleo diet typically exclude? *

- ☐ Dairy
- ☐ Grains or cereals, such as wheat and rye
- ☐ Meat
- ☐ Legumes, such as peanuts and beans
- ☐ Processed foods
- ☐ Any foods that contain sugar
- ☐ Salt
- ☐ Caffeinated beverages
- ☐ Alcohol
- ☐ Soft drinks
- ☐ Refined sugar
- ☐ Fruit juice
- ☐ Take-away food
- ☐ Oils

How likely are you to eat a paleo snack, such as baked kale chips instead of potato chips? *

0 1 2 3 4 5

Extremely unlikely ☐ ☐ ☐ ☐ ☐ ☐ Very likely

FIGURE 2.15 The 'View live form' view of the same questionnaire from Figure 2.14, created using Google Forms. The question types visible in the screenshot are multiple choice, Likert scale and checkboxes. The required questions are highlighted with a red asterisk at the end of each question.

Selecting representative samples

You will rarely, if ever, be able to survey 100% of the target population, so you must choose a sample. Ensure your samples are representative (typical), not biased or atypical, or you will collect biased data that will ultimately generate inaccurate information.

The American *Literary Digest* in 1936 used a survey to find who people would vote for in the upcoming presidential election, which pitted the incumbent President Franklin D. Roosevelt, a Democrat, against Alfred Landon, the then Republican governor of Kansas. The poll predicted a huge victory for Landon. However, the election result was a massive victory for Roosevelt. The poll's main problem was **selection bias**. Despite being one of the most expensive and largest polls of all time with a huge sample (10 million, with 2.4 million respondents), the sample was not representative because it was based on a mailing list that was drawn from every telephone directory in the United States, club membership lists and lists of magazine subscribers – things that were all geared towards middle class and wealthy families. At the time, wealthy people tended to vote conservative (Republican). The poll failed to sample everyone else.

Timing can also play a part in selection bias. A company conducted a door-to-door survey asking whether government pensions were adequate, and an unexpectedly high number of respondents answered in the negative. Why? Perhaps contributing to the result was the fact that the survey was conducted during business hours. During that time, there was a disproportionate number of unemployed, aged and infirm people – that is, those most directly affected by government pensions – at home at that time to answer the door and respond to the poll.

Another type of selection bias is self-selection bias. A pollster in a supermarket found that an unusually high number of shoppers did charity work. Why? Perhaps because the sort of person who does charity work is also the sort of kind, empathetic and helpful person who would stop to help a pollster in a supermarket. Just choosing to participate in the survey self-selects a certain type of respondent, which can lead to bias in the results.

Distributing questions

You cannot force people to answer your research questions, so finding a representative sample will be a challenge. To find your sample, you could try approaching people in public places, being careful not to select or avoid people of certain sexes, ages, races or appearance.

You could use an online questionnaire with Google Drive, Survey Monkey or similar software, but be aware that any online poll will automatically bias the sample towards younger, wealthier, more technologically capable respondents. Polling your peers is worthwhile, but it will automatically bias your sample towards young people living in a particular social and cultural group. (Your peers may or may not share the same socioeconomic group, religion, gender or ethnic background as you.)

You could try leaving copies of your questionnaire in shops, but be aware that self-selection bias will likely occur.

Querying resources

The data sources you use to generate information to **inform** your hypothesis will include large, complex and remote data sets. You will need to use queries to find the data that is relevant to your hypothesis within them. A resource is an organised collection of data, such as a database, data table, library, data file, file server, metadata, website or **XML** file. The data you use to generate information about your hypothesis may well come from a relational database.

Querying a database

To find and extract relevant data from a database, you need to use the built-in services that allow a database to interact with an external entity – whether human or digital. As long as the user (you) requests data using correctly formatted queries that the database services can understand, the database can comply. The following section discusses two types of correctly formatted queries: **query by example (QBE)** and **structured query language (SQL)**.

Query by example (QBE)

QBE uses a database's graphical user interface (GUI) to let a user type the data they wish to find into the field to be searched. If you enter data into multiple fields, records that have matching data in *all* of the specified fields will be selected.

For example, the query 'Women living in St Kilda', searches two fields:

- 1 the `Sex` field, for a value representing females (for example, 'F') AND
- 2 the `Suburb` field, for records containing 'St Kilda'.

Only those records with successful matches in *both* fields are selected for display. The query would be written as:

```
Sex = "F" AND Suburb = "St Kilda"
```

To select the relevant records, you must use the correct logical operators.

If more than one value could select a record, use OR.

To find people aged 17 or 18, the query would be:

```
Age = 17 OR Age = 18
```

If *either* condition is true in a record, it is selected.

Parentheses are used to logically group logical tests. The following finds males aged 21 and older as well as females 19 and older.

```
(Age >= 21 AND Sex = "M") OR (Age > 18 AND Sex = "F")
```

You will notice that in queries, double quotations marks should be used.

Other examples follow in Table 2.4.

THINK ABOUT INFORMATICS 2.4

Explain how the parentheses change the meanings of these three queries:

- 1 If A=true and B=true and C=true or D=true ...
- 2 (If A=true and B=true) and (C=true or D=true) ...
- 3 (If A=true and B=true and C=true) or D=true ...

TABLE 2.4 Other query examples

Query expression	Finds all records where:
Age > 18	The contents of the Age field are greater than 18.
Sex = "M"	The Sex field contains the letter 'M'.
Age >= 18 and Sex = "M"	The person is a male who is 18 or older.
(Age >= 18 and Sex = "M") or Sex = "F"	The person is either a male who is 18 or older, or is a female.
(Age >= 18 and Sex = "M") or (Sex = "F" and City <> "Melbourne")	The person is either a male who is 18 or older, or is a female not living in Melbourne.
(Sex = "M" and (Age >= 18 and Age < 35)) or (Sex = "F" and City <> "Melbourne")	The person is a male aged between 18 and 34, or is a female not living in Melbourne.
(Sex = "M" and (Age >= 18 and Age < 35))	This query was meant to find males between the ages of 18 and 35 inclusive, but it fails. Find and fix <i>two</i> errors in it.

Pay close attention to the parentheses that group logical tests. Always make sure that you pair your parentheses; that is, that there are equal numbers of left and right parentheses. Also check that tests are grouped correctly to give the right logical result. Be mindful that:

- AND reduces the number of records found because it adds another selection criterion that each record must satisfy. So, each time you add AND, your results are reduced.
- OR can *increase* the number of matching records because it adds a different way of each record being selected. So, each time you add OR, it can increase your results.

Microsoft Access queries can specify selection criteria and also indicate how the found records should be sorted, and which fields should be displayed. In cases where the selection criteria keep changing, you can instruct the query to ask the user for the criteria every time you run it.

Structured query language (SQL)

SQL is a standard method of requesting data from databases. It is a textual equivalent of QBE. SQL is commonly used to save a query as a script so it can be reused easily.

The following shows a query as SQL (the same query is also shown in Figure 2.16 on page 107):

```
SELECT BOOKS.title, BOOKS.price, BOOKS.price, BOOKS.isbn
FROM BOOKS
WHERE (((BOOKS.price) > 100)) OR (((BOOKS.price) < 10))
ORDER BY BOOKS.title;
```

Querying other resources

Like databases, other data resources must have some type of protocol (set of communication rules) to enable them to receive data requests and send back what was requested. For example, data stores such as web and email servers negotiate data transfers with browsers and email clients using the HTTP and IMAP protocols.

Another common method of querying resources is an **application programming interface (API)**, which programmers create as a gateway so other programs can interact with their software. Large data resources such as Flickr, Facebook and Twitter publish APIs to create an 'open architecture' to share their content. The next time you send a photo from Flickr to your Facebook page, login to a blog using your Facebook credentials, or use a phone app to get a weather forecast from the Department of Meteorology, you are taking advantage of their APIs.

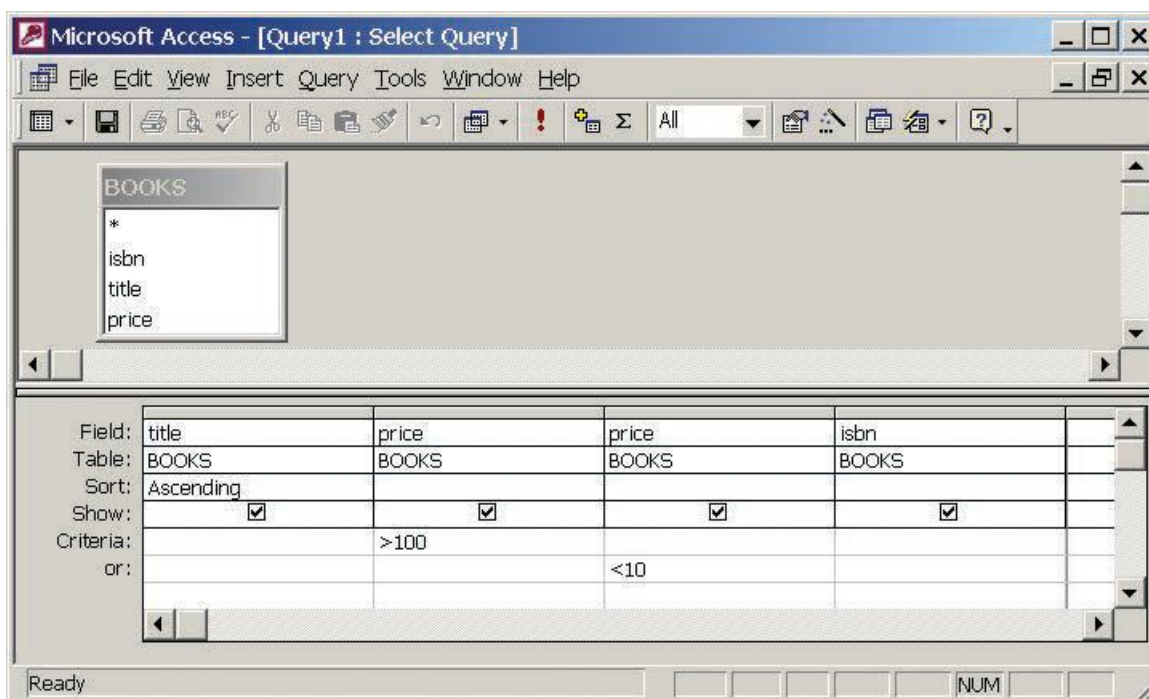


FIGURE 2.16 QBE in Microsoft Access to find, sort and display records containing different types of data

Referencing data sources

You must acknowledge any sources of information you use in your work. This means not just direct quotes: it also includes ideas, summarising and **paraphrasing**. The sources of intellectual property (IP) that influence your work should be acknowledged using one of several accepted referencing styles.

- American Psychological Association (APA)
- Harvard
- Chicago
- Institute of Electrical and Electronics Engineers (IEEE)

The referencing styles described here are not just applicable to this Outcome. You will find they are relevant to the rest of your VCE studies and any studies or writing you go on to do in the future. (However, you must still specifically provide evidence of acknowledgement of intellectual property in Unit 3, Outcome 2.) Each referencing style will enable you, as the author of your work, to acknowledge the author(s) of any intellectual property used within your work, so that:

- it is clear that you are not claiming another author's work as your own
- readers can find the original source to get more information
- copyright 'fair use' legalities are observed
- moral rights are observed.

It is wise to start recording your sources as you find them, rather than at the end of the Outcome. Having to locate them again to cite them may be difficult and time-consuming, especially if you have forgotten.



APA style guide:
Swinburne University
library (pdf)



APA style: refer to
the 'Quick answers –
References' section at
the bottom left of the
homepage.



Harvard style guide:
Swinburne University
library (pdf)



Chicago Manual of
Style: quick guide



IEEE Citation reference
guide (pdf)

If you fail to acknowledge someone else's intellectual property in your work, you may be accused of, and guilty of, plagiarism. Plagiarism, which is when you pass someone else's original work off as your own, is very serious. It can have repercussions beyond embarrassment: you may find yourself in academic and legal strife. Your school or university may take disciplinary action. You may face prosecution or civil legal action under copyright laws.

All referencing styles use a brief identifying **citation** in the body text, and a matching detailed reference in a reference list (bibliography).

Styles

The APA style is widely used and is one of the most common styles students are expected to use in referencing. Citations appear within the text and their corresponding source details in a references list at the end of the work. When using the APA style, the author's surname and the publication date are featured in the text.

Despite the fact that stock and broth seem, at first glance, similar enough to bone broth, and tend to be used interchangeably, the topic of bone broth can be contentious. 'A bit of the bad press surrounding the paleo diet stems from bone broth, and the notion of feeding it to babies and toddlers as formula.' (Green, 2014, pp. 23–4)

Cite all references to a publication within the main body of a document. The format usually follows author-date-page where the author's last name, the year of the publication and the page number of the quote should be referenced. These are usually separated by commas, and are placed within parentheses following the text. The page number is preceded by a lower case 'p' or 'pp' with a period (full stop) after it.

Harvard and APA are very similar styles, both being author-name, date systems. In the Harvard system, there are subtle differences, mostly in terms of punctuation in the *reference list*. However, for acknowledging sources in the *body text* of your work, Harvard and APA styles are the same.

The Chicago style comes in two varieties. The author-date style, favoured by the sciences, is similar to APA and Harvard. Its in-text citation style differs only slightly:

... 'A bit of the bad press surrounding the paleo diet stems from bone broth, and the notion of feeding it to babies and toddlers as formula.'
(Green 2014, 23–4)

(Note that the differences are that there is no comma separating the author name and year, and no 'p.' or 'pp.' preceding the page reference.) Its notes and bibliography style, favoured by the humanities, uses footnotes instead, with a superscript notation in the body of the text:

... 'A bit of the bad press surrounding the paleo diet stems from bone broth, and the notion of feeding it to babies and toddlers as formula.'¹

Bibliography

- 1 Green, Nina. 2014. *The Paleo Way: Method or Madness?* (4th ed., vol. 1, pp. 23–4). South Melbourne, VIC: Coconut Press.

TABLE 2.6 A comparison of all four reference styles using a simple website reference list citation. Note that the first two (Harvard and APA) treat the example website as though the author is clearly distinguishable from the sponsoring website. The last two assume that the author cannot be distinguished, and therefore reference the sponsoring website only. You should check the legal section of any website if the webpage and document you are using has no information about the author on it.

Harvard	Author or organisation who created the webpage, if clearly identifiable	Year the webpage was published, if known, or most recently updated,	<i>Title (of the webpage or document),</i>	Name of government agency, organisation or sponsor hosting the page on their website (this may be identical to author name),	Date of access viewed DD-MM-YYYY,	<URL>.
	Evans, Pete	2015,	<i>The Paleo Way: 10 Week Info</i>	Pete Evans The Paleo Way Pty Ltd,	viewed 31-12-2015,	< https://thepaleoway.com/10wkinfo >.
	Evans, Pete, 2015, <i>The Paleo Way: 10 Week Info</i> , Pete Evans The Paleo Way Pty Ltd, viewed 31-12-2015, < https://thepaleoway.com/10wkinfo >.					
APA	Author or organisation who created the webpage, if clearly identifiable	(Year the webpage was published, or most recently updated).	<i>Title (of the webpage or document),</i>	Name of government agency, organisation or sponsor hosting the page on their website (this may be identical to author name).	Retrieved from URL.	
	Evans, Pete,	(2015).	<i>The Paleo Way: 10 Week Info,</i>	Pete Evans The Paleo Way Pty Ltd.	Retrieved from https://thepaleoway.com/10wkinfo .	
	Evans, Pete, (2015). <i>The Paleo Way: 10 Week Info</i> , Pete Evans The Paleo Way Pty Ltd. Retrieved from https://thepaleoway.com/10wkinfo .					
Chicago	Author or organisation who created the webpage, if clearly identifiable.	Year the webpage was published, if known.	'Title (of the webpage or document)'.	Date website was most recently updated, if known, or accessed, if not known.	URL.	
	Pete Evans The Paleo Way Pty Ltd.	2015.	'The Paleo Way: 10 Week Info.'	Accessed March 11.	https://thepaleoway.com/10wkinfo .	
	Pete Evans The Paleo Way Pty Ltd. 2015. 'The Paleo Way: 10 Week Info.' Accessed March 11. https://thepaleoway.com/10wkinfo .					
IEEE	Author or organisation who created the webpage, if clearly identifiable,		'Title of webpage or document,'	Year the webpage was published, if known. [Online].	Available: URL.	[Accessed: Date.].
	Pete Evans The Paleo Way Pty Ltd,		'The Paleo Way: 10 Week Info,'	2015. [Online].	Available: https://www.thepaleoway.com/10wkinfo/ .	[Accessed 31 December 2015].
	Pete Evans The Paleo Way Pty Ltd, 'The Paleo Way: 10 Week Info,' 2015. [Online]. Available: https://www.thepaleoway.com/10wkinfo/ . [Accessed 31 December 2015].					

Referencing tips

Quotations

- Do not correct errors in quotations. A quotation should be represented as it appears originally, even if it has errors. You can use '(sic)' or '[sic]' to flag to a user that you have noticed an error in the quotation and that it has not come from you.
- Do not change the spelling in a quotation or a name to match Australian English. Preserve correct names and original spelling. For instance, if you quote the World Health Organization, or need to make note of the World Trade Center, you would not change their names to World Health Organisation and World Trade Centre.
- Use square brackets to add explanatory information or give context to a quotation; for example, Darcy's claim that 'during the [American Civil] war, infections killed more men than bullets ever did ...'
- Use ellipses '...' to indicate when part of a quote has been omitted.

Reference list

- You can use Microsoft Word to save a list of your sources in a document and produce a bibliography for you. On the 'References' ribbon, look for 'Bibliography' or 'Citations & Bibliography' (this will depend on your version of Microsoft Word).
- Write 'n.d.' if no date is given for a website or document.
- If no page number is known, write 'n.p.' and provide an estimated page, paragraph number, or nearby heading.
- Use unspaced dashes rather than hyphens in your page reference spans. To insert an en dash in Microsoft Word in Windows, hit CTRL and minus on the number pad, or ALT and 0150.
- In the reference list, use the hanging indent paragraph style.
- If the source has a **digital object identifier (DOI)**, place it at the end of the reference. The DOI is a permanent, unique identifier for an object, such as a serial number for publications, and may look like doi:10.1037/0022-006X.56.6.893. It never changes even if the document's location or name changes.

Integrity of data

For a computer to produce useful information, the data that is input into a database must have integrity. Whether you are storing, transmitting or archiving data, you must be sure that its integrity is maintained. Otherwise, the data may not be accessible when you need it.

In your Outcome, you will rely on quality data to support your hypothesis. It is important for you to develop and apply knowledge of integrity of data to your Outcome. The following sections discuss factors that influence integrity of data: timeliness, authenticity, relevance and accuracy.

Timeliness

Data must be timely for it to produce usable information. That means it needs to be processed while it is current and there should be no significant delays in retrieving it. The methods used to process data should be efficient enough to be complete by the time the data is actually needed.

Make sure that the digital systems (hardware and software) used are powerful enough and appropriate for the task at hand to avoid causing delays. It is crucial that decision-making is never based on outdated data.

Protect users of your data as much as possible from potential delays caused by power outages, hardware failures, downtimes due to system upgrades, and deliberate threats from outsiders, such as denial-of-service attacks or malware infection.

Make sure the age of data is known before using it: it may be meaningless to draw conclusions from very old data; for example, using data about food consumption that is 50 years old to draw conclusions about dietary preferences today.

Authenticity

Your data and information are only authentic as long as the origin attributed to them is correct, and providing that the data has not been improperly or inappropriately changed since it was published.

Digital documents are easy to fake and distort convincingly. It can be difficult to tell if an image, document, database or webpage is genuine, a parody (for example, The Onion) or a deliberate attempt to lead, fool or defraud people, as with spam, links that trick visitors into clicking ads, and phishing sites.

Characteristics of authentic data

Digital data can only be considered authentic (genuine) if it:

- comes from the author and/or source it claims to be from
- has not been deliberately corrupted
- is not faked or disguised as something else
- has not been changed without authorisation
- is what it claims to be and does not misrepresent itself
- does not aim to mislead or deceive by pretending to be anything else
- does not lie.

Challenges to authenticity of data

The challenge of identifying inauthentic data becomes more difficult every day. Spoofing attacks occur frequently online. In a spoofing attack, one person (or multiple people) masquerades as another for the purposes of fraud, advantage or amusement. This usually requires falsifying data, such as caller IDs, email addresses and IP addresses.

Authors plagiarise other people's publications and steal data. Similarly, copyright, source and author information can be edited out of pictures and documents, making their true source, age and legal status unknown. Watermarks can be removed from digital images.

Legitimate media and advertising productions add to the challenges in several ways. Music producers use auto-tune to repair flat notes their singers have hit during recording. Computer-generated imagery (CGI) in movies and TV is becoming undetectable, while magazines edit photos of models into impossibly perfect specimens. Viral ad campaigns and fake YouTube videos grab attention by appearing to be genuine.

Few people use their real names online. When only avatars and nicknames appear, true identities are uncertain. Authentic social networking accounts get hacked and post fake updates or tweets. Fake celebrity accounts are easy to create. Even the 'From' addresses in email are easy to fake. Images that may have already been clearly identified as fake may be re-posted without the warning.

Torrent sites increase the likelihood of downloading data that differs from the original. Although uploaders who post files that have been cracked and corrupted

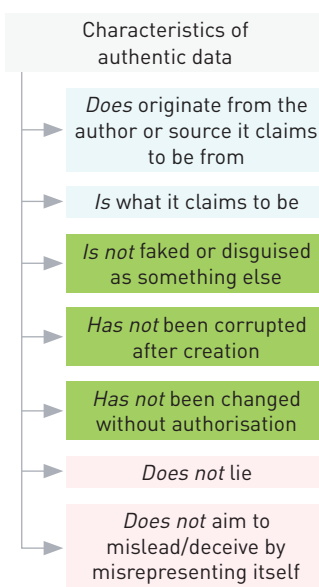


FIGURE 2.17 The characteristics of authentic data

or changed in some way are unlikely to go unnoticed for very long, in the short term, inexperienced users are vulnerable to downloading the corrupted data. While torrent sites often have legal downloadable files, many include cracked software. Cracked software in itself is a challenge to authentic data, but it also poses both legal and ethical implications for the user. Many users who choose to download cracked software knowingly bypass alerts from their firewall or malware detector to install the program, essentially rolling the dice on the possibility that it may have been changed in other ways.

Authenticity techniques

There are a variety of ways of enhancing the authenticity of data. The technique used will depend on the circumstance. Broadly, the techniques can be categorised as digital or non-digital.

Digital

Use digital signatures. A digital signature is a mathematical method used to validate the authenticity and integrity of messages, software and digital documents.

Use SSL or TLS to encrypt web traffic and prevent it being read or modified in transit.

Use security certificates. With each page it sends, a web server verifies its identity by also sending the certificate. Browsers check the certificate's legitimacy to ensure the server is genuine.

Use checksums. MD5 and FFP fingerprints are values (checksums) calculated using the data in a published document and embedded into the document. Publishers can give the checksum of a file they offer, and anyone can re-calculate the fingerprint of the file they receive. If the calculated checksum does not match the published checksum, it proves that the document has been damaged or altered.

Use email validation to verify that an email address exists and belongs to the correspondent.

Non-digital

Compare original documents with any allegedly accurate copies. Contact the original authors of documents to verify their authenticity, if that is possible.

Relevance

People look for information that relates to a topic that interests them. Relevance measures how closely a resource, such as a book, database or webpage, corresponds to people's desire for information.

Relevance is not always easy to measure. For example, it *is* obvious that income is relevant to spending habits. However, it is less clear whether age or gender are relevant to development of schizophrenia. Data about schools in the United States could be relevant to Australian schools, but this is not absolute. Data about men may apply to women in some circumstances, but not in others. Assuming that data is relevant when it may not be can lead you to make invalid conclusions.

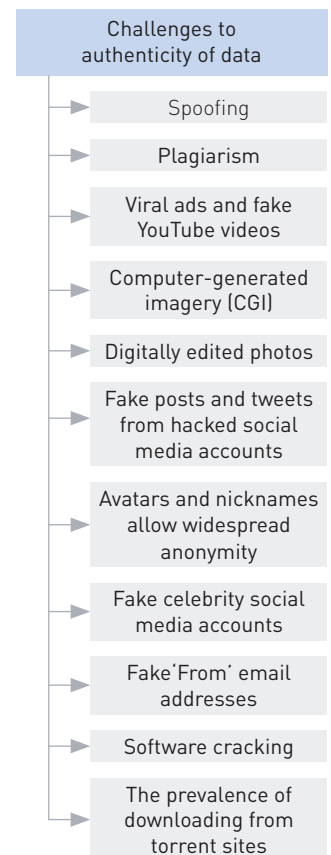


FIGURE 2.18 Challenges to authenticity of data

TABLE 2.7 Circumstances when your data is in danger of becoming less relevant

Reason	Example
It is on a different topic. Using internet research to drift from link to link may lead to finding information that is irrelevant to the original topic.	A discussion of Scottish history mentions the sport of caber tossing, which leads to a discussion of how the sport is scored, which somehow drifts into detailed coverage of kilts.
It is from a place where conditions are not comparable.	Comparing gun ownerships in suburban Australia with that of wild Canadian bear country
It is from a different time.	'Boys in the 1950s were happy to leave school at 15 and get married at 18. What's wrong with kids nowadays?'
There are significant differences in history, conditions or circumstances that prevent two data sets being compared.	Cultural differences, war status, climate, characteristics of interview subjects and unusual recent events

Accuracy

There are two main characteristics of data accuracy: **content (functionality)** and **form (appearance)**. Content is divided into two parts: **correctness** and **completeness**. The following sections discuss each aspect of accuracy, followed by challenges to data accuracy and the measures to improve them.

Correctness

Correctness means that the values stored for a given object must be correct. For example, if you create a database that includes a list of all the subjects participating in interviews you are conducting for your Outcome, you would need to ensure that you enter each subject's details correctly into the database. The content would be incorrect if you keyed the wrong date of birth into the `BirthDate` field or misspelled a surname in the `LastName` field.

Data is also correct if it is a truthful representation of the real-world construct to which it refers. For example, the weather forecasts by the Bureau of Meteorology are considered (usually) correct, in that they truthfully represent our *concept* of the weather in terms of predicted temperatures, rainfall, winds and so on.

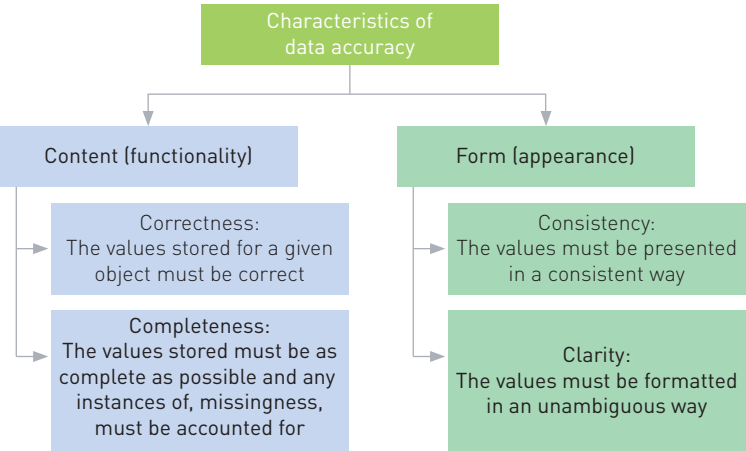


FIGURE 2.19 Characteristics of data accuracy

Completeness

Completeness means that your data set is just that: complete. You have data from all research participants on all variables at all relevant points in time and space. Completeness can be very difficult to achieve.

Clarity

The form of data is important as well as the content because it will remove ambiguity about the content. Accordingly, form is divided into two parts: **clarity** and **consistency**.

Clarity is about formatting data in an unambiguous manner to prevent misinterpretation. For example, you have entered the dates of birth of all of your interviewees into your database and they are the correct values. They are still not completely accurate, because it appears that some of the birth dates have been entered using the US date format (MM/DD/YYYY), and others using the Australian date format (DD/MM/YYYY). Worse still, some of the dates have both days *and* months under 12, making the actual interviewee subjects' dates of birth ambiguous:

04/11/1996	Grey, Elijah
12/02/1991	Kaur, Sandeep
06/09/1994	Imran, Al

The birth dates are inaccurate because you cannot tell what the true, correct values are. You need to be stricter when entering data. In doing this, you can apply consistency.

Consistency

Correct, unambiguous data can still cause a problem in a database if it is not consistent. Inconsistent data is unwelcome because it means the data is unreliable. This is why consistency is part of data accuracy in this Outcome.

If you need to compare demographics as part of your Outcome, and some of your data came from subjects or respondents in inner Melbourne suburbs, you may make the mistake of entering both 'St Kilda' and 'Saint Kilda' into your database. Both of these would be correct and they would be unambiguous. However, they would not be consistent. This is a problem because you would not be able to compare the data you have gathered properly because you would have data for two different suburbs instead of one. If you looked directly at the data once it is entered into your database, you may notice that the numbers are incorrect and fix it. However, it may seem insignificant enough that it does not catch your eye, even though it is there. Anomalies of this nature or similar prevent data being accurately aggregated and compared, and increase the likelihood of generating inaccurate information.

Consistency is also a concern on a larger scale when it occurs between multiple data sources where conflicting versions of data appear in different places. If a 'true' value cannot be easily determined between multiple data sources, an entire data source loses integrity and becomes tainted.

The following section discusses the various challenges to data accuracy, covering each of the characteristics in turn. Covering this material will help you to prepare for your Outcome and carry out your research more effectively.

Challenges to data accuracy

Correctness

Faulty data acquisition equipment, such as inaccurate surveying software, and poor acquisition procedures, such as ambiguous questions, will pose a problem and introduce an accuracy problem from the moment you begin collecting data.

While you acquire your data, you may find that respondents misinterpret questions and answer inappropriately; for example, you ask a question about eating meat, and they assume the question excludes fish, and answer it incorrectly. They may also simply misremember facts.

At times inaccuracy is caused by genuine deceit – in the form of fraud, vandalism or criminal deception. At times it is less clear cut than that. Bias creeps into data if it is from a pressure group, which is an organisation driven by morals, special interests, profit, or even manipulative governments with a particular motive. If data providers have a **vested interest**, they should be acknowledged. For example, sometimes journalists need to declare that their travel expenses have been paid by the company whose products they are discussing.

Most research is based on samples from a much larger population. If the sample size is too small, this can mean the data is unrepresentative of the entire population. Sometimes the data collected is inaccurate because it is not detailed enough to provide a true answer. For example, a pie graph alone does not offer enough detail for a serious statistical study.

Accuracy may be lost through data-entry errors, such as simple typing mistakes. Data may also be distorted during processing through improper statistical methods or poor language translation. Data may be stored, processed or communicated poorly.

Accuracy may also be lost as time passes. Changes in the real world since the data was collected may have a serious impact on the accuracy of the data. An overnight price rise can instantly make yesterday's advertisements or cost estimates inaccurate. Some data is dynamic (constantly changing), so if it is not maintained or synchronised, it becomes outdated. Most data needs to be updated, synchronised and have the necessary corrections applied over time to keep it current.

THINK ABOUT INFORMATICS 2.5

- 1 Identify the data formats shown in Figure 2.21.
- 2 Explain how you would access data found only on one of these obsolete formats.

Completeness

Sometimes the data needed is no longer available because the records may be missing or the files have been lost as a result of fire, flood or war. Data from certain time periods may be unavailable. Digital data can be lost by accident or equipment failure. Disk rot or link rot may occur, or the format used to store the data may become obsolete. Obsolete data itself may be deleted rather than archived. Valid data may be removed accidentally during data cleansing.



Mark Fergus

FIGURE 2.20 Data may be incomplete because of water-damage to paper records that have never been transcribed to digital form



FIGURE 2.21 Data may be incomplete because it is only available on obsolete data formats

In some circumstances, the data exists, but is inaccessible. For example, the data may have been collected but never published, or it has been censored by a court, privacy legislation, confidentiality agreement or oath. Sometimes data is deliberately ignored or buried because it does not support a hypothesis. For instance, counter-evidence may be ignored because researchers wish to show data from only one side of a debate. Cherry picking occurs when some data is selectively chosen while other data is omitted; for example, researchers have a hypothesis about courts being too soft on crime, and try to support the hypothesis using data showing crime rates rising from 10 000 in 2015 to 12 000 in 2016, but in doing this they choose to exclude data showing that the crime rate in 2014 was much higher – 36 000.

Simple error also introduces a great deal of incompleteness. Data-entry errors occur. Researchers may fail to acquire the data for a relevant field; for example, forgetting to ask about alcohol intake when interviewing people about weight loss. They may ask irrelevant questions, such as asking single people how long they have been married. Sensors, both human and digital, can fail or measure inaccurately. If a database has been set up with overly stringent validation rules, it can prevent the input of unusual but valid data; for example, a 12-year-old genius cannot enrol at a university because its database refuses to accept the ‘unreasonable’ age of 12.

In some cases, respondents fail to answer certain questions. Sometimes they may refuse to answer sensitive questions. Participants may also drop out of studies before they are complete, leaving gaps in the research.

Regardless of the cause, incomplete data can be just as misleading as incorrect data. Try to avoid drawing conclusions from incomplete data, particularly if the missing data could be vital to supporting or refuting your hypothesis.

Clarity

Data-entry errors and using multiple data-entry operators can introduce ambiguity into the formatting of the data you have collected. Using multiple tables with different rules for formatting dates or numbers, such as dates of birth or calculated ages, can also pose a problem.

The Official Secrets Act still suppresses information of some events from World War II.

Consistency

Consistency problems can occur on several levels. For example:

- the date of birth listed in one table does not match the age calculated in another table
- a respondent's answers to one survey question conflict with another
- data stored in one location, such as a local database, does not match corresponding data in another location, such as a linked website database.

You may cause consistency problems with your data if your formulas or queries have errors that calculate ages incorrectly or there are mistakes in data entry.

You may also cause consistency problems with your data if you save multiple versions of your files without carefully managing them to avoid versioning issues.

Your respondents may cause consistency problems if they are not completely truthful or misinterpret questions.

This means that consistency does have a degree of overlap with correctness and therefore functionality. When the data is inconsistent, it becomes hard to tell which is correct and therefore accurate information cannot be generated from the data.

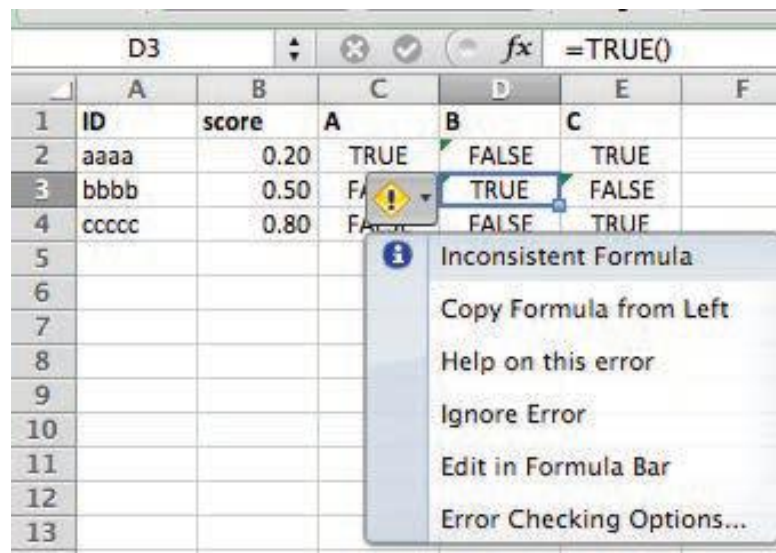


FIGURE 2.22 Microsoft Excel can detect inconsistent formulas and warn the user

Measures to improve accuracy

Correctness

You can perform data quality assurance to cleanse or scrub data. This will identify and remove or repair data that is incomplete, inaccurate, irrelevant or inconsistent. It may also standardise data; for example, by changing all instances of 'Street' to 'St' or adding data to existing records, such as Medicare or tax file numbers.

Recalculate digital signatures or fingerprints to check if digital files have been modified or suffered from disk rot. You should also frequently synchronise copies with source data and conduct a quality check on secondary data obtained from a third party.

Completeness

For data that you suspect has gaps because it has been collected but not published in full, you could contact the original data collectors to ask if they have unpublished data that they could provide you.

This is well beyond the scope of VCE Informatics, but you can also improve correctness by conducting a cyclic redundancy check (CRC) to detect data copy errors.

If the topic is controversial, look at data from proponents of both sides of the argument to see what inconvenient facts may have been ignored. Look for the hidden counter-evidence and find what has been cherry-picked. Researchers are often keen to point out the faults in evidence used by their opponents.

To minimise the potential for data-entry errors, use existence validation to ensure that essential fields cannot be left empty when a database record is being entered. You can still use an incomplete record if it has valid data for a particular question, but ignore it for questions where it lacks relevant data. A single person who left the 'How long have you been married?' box empty could be used for calculating the number of times people get married, but ignored when calculating the average length of a marriage.

You can also interpolate the missing data. **Interpolation** is using a trend to estimate a missing value within a data set. For example, 1990 had 1000 sales, 1991 had 1100 sales, 1992's data is missing and 1993 had 1300 sales. If you interpolate 1992's missing data, it would have had approximately 1200 sales.

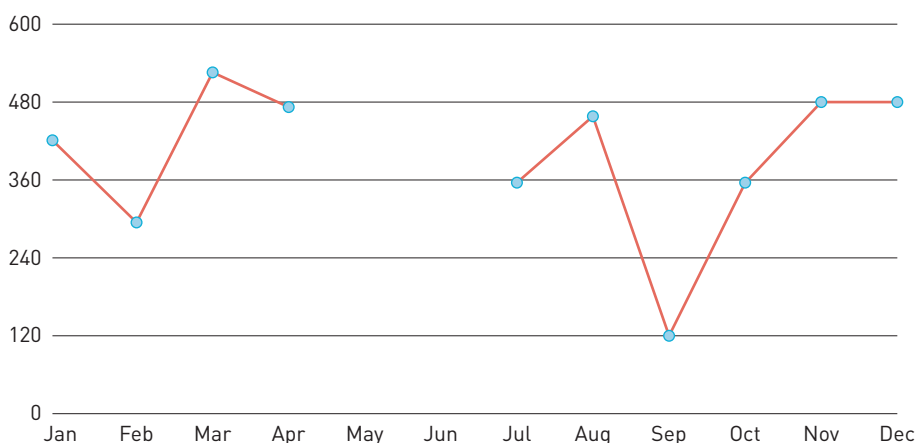


FIGURE 2.23 Missing data usually means trouble for researchers

Avoid dubious methods of dealing with incomplete data.

Do not use **mean substitution** to fill in missing data with the average value calculated from other records. Instead, use **imputation** to fill in a record's missing values using data from records that are otherwise similar to it.

Do not simply ignore incomplete records during processing. This is just as troublesome as ignoring missing fields. The data may be incomplete for highly significant reasons, so removing the records could introduce bias into the information generated from the records and skew the results and conclusion.

Be transparent. If your data is incomplete, do not try to hide it or ignore it. Sometimes incomplete data is unavoidable, so acknowledge it, explain what was missing and how you dealt with the issue.

Make sure you investigate reasons for data being **missing not at random** (see Table 2.8). Data missing not at random may be significant to your research.

If fields are being left empty, you could try using sophisticated analysis techniques, such as hierarchical linear modelling and survival analysis, to deal with incomplete data, but these are also beyond the scope of your Informatics study.

Validation techniques are discussed in Chapter 1, and are also covered in *Computing VCE Units 1 & 2*.

THINK ABOUT INFORMATICS 2.6

A real-life study ignored all records with missing data, and cherry-picked to use only records that were complete.

It later reported results saying that those who drink more alcohol have fewer problems with depression and anxiety, which contradicted common sense and decades of other research.

After appropriate handling of missing data, new results arose that were more consistent with the literature.

Suggest another case where the removal of incomplete records could lead to wrong conclusions.

TABLE 2.8 Discussing incompleteness: Some researchers discuss the data missing from their analyses, and apply categories to it. If you have data missing in your Outcome, you may wish to apply similar categories.

Category	Explanation
Legitimately missing	It is okay for it to be missing; for example, home phone number for a person who does not have a landline phone, years married for an unmarried person or hair colour for a bald person.
Illegitimately missing	This data should be complete, but is not.
Missing at random	There is no pattern to the missing data.
Missing not at random	There is a pattern to the missing data; for example, most men answered this question, but many women did not. This type of missing data should <i>not</i> be ignored.

Clarity

Enforce data formatting and validation rules in your spreadsheets and database that prevent misinterpretation of the data. Try to avoid rekeying data where possible – copy and paste or import instead.

Try out your questions with test subjects, such as your classmates or family, before conducting your research, so you can check that any interview questions or questionnaires make sense and are not ambiguous.

Put dummy data in your database and spreadsheet to check for errors and potential problems with ambiguity before you begin inputting your real data, so that you can troubleshoot. Do this more than just one row or record – input a few so that you can test properly.

Consistency

When formulating questions that deal directly with people, ask the same question again in different ways to check the answer consistency. This is important for questions about sensitive issues, because they may lie, avoid or bend the truth. They may do so for any number of reasons, from embarrassment to exaggeration and simple memory lapse. If you have a hypothesis that requires you to collect data about gambling habits, consider inserting ‘Do you gamble?’ at question 7 and then perhaps ‘Last year, how often did you bet on a horse, buy a lottery ticket or use slot machines?’ at question 38. People may forget they fibbed at question 7 by saying no and tell the truth on question 38.

You can later use electronic checking to detect inconsistent answers. For example, if your questionnaire asked how many children someone had in one question, and the age of their oldest child in another, the following would detect inconsistent answers:

```
IF (NumberOfChildren = 0 and AgeOfOldestChild > 0) then Consistency Error
```

To prevent inconsistency caused by data entry, enforce consistent data formatting and validation rules. Try to avoid rekeying data where possible. Instead, copy and paste or import it after it has been initially keyed in. This advice is repeated because it is important for data accuracy – you should try to only key data in once.

If people can enter a date of birth in two different places in the same database, make sure you use the same data format in both locations. Do not format to accept YYYY-MM-DD (2016-07-27) on one form and DD-MMM-YY (27-Jul-2016) on another. You can use an input mask to force a particular data-entry format, such as XXXX-XXXX-XXXX-XXXX for credit card numbers.

YYYY-MM-DD is the international standard (ISO) format for dates. 2016-04-01 translates to 1 April 2016.

THINK ABOUT INFORMATICS 2.7

Why do you think there is an ISO standard for date formats?

Standardise data so it is always in a consistent, predictable and comparable format.

- Store family names in uppercase so that, when running comparisons or queries, 'Smith' is not overlooked as a match for 'smith' or 'SMITH'.
- Strip spaces from phone numbers and the start and end of all text data.
- Always use the same format when dates are stored as text.
- Use a standardised time setting, such as Greenwich Mean Time (GMT) or Australian Eastern Standard Time (AEST).
- Store all currency values in the same units: AUD, USD or GBP and so on.

However, if some data **must** differ, use **fuzzy logic** to handle trivial differences in data. For example, accept 'Robert' if the user is also known as 'Bob'. Treat 'St Kilda' and 'Saint Kilda' the same way.

Enforce referential integrity in databases to ensure key fields cannot be deleted without also deleting or modifying related fields. For example, in a company database, if the `Sales` department is deleted from the `Departments` table, everyone working in `Sales` need to be either deleted or moved to another department, otherwise their `Department` data becomes meaningless.

Normalise to third normal form (3NF) to ensure proper synchronisation of table data, and prevent duplicate data. Any data stored in more than one place immediately invites the risk of inconsistency. If employee addresses are stored in the `STAFF` database and in the `CONTACTS` database, they could easily become inconsistent if one table is updated and the other is not.

The domain name system (DNS) is the biggest distributed database on the planet and lists every internet domain (for example, `fred.com`) and its IP address (for example, `192.34.78.101`) by pointing to the location of the web server hosting the site's files. Changes to a site's IP address can take 24 hours to propagate (spread) throughout the DNS, so some people may quickly find the new home for a site while others receive 'file not found' (404) error messages.

Normalisation is discussed in Chapter 1.

Data types and data structures

To store your data, you should be familiar with data types and data structures. Data types are categories of data that have individual properties and ways they can be processed. The data types and structures are discussed in Chapter 1. You may also recall them from *Computing VCE Units 1 & 2*.

Data is categorised into types so it can be stored efficiently and processed effectively. These include the following: numeric (integer and floating point), text (string), Boolean, character and date. The basic data types can also be subdivided into more specific variants, as shown in Figure 2.24. FileMaker understands only the number data type, but a programming language may understand a dozen number types, including **unsigned short integer**.

The basic data types are described in detail in Chapter 1, pages 22–3.

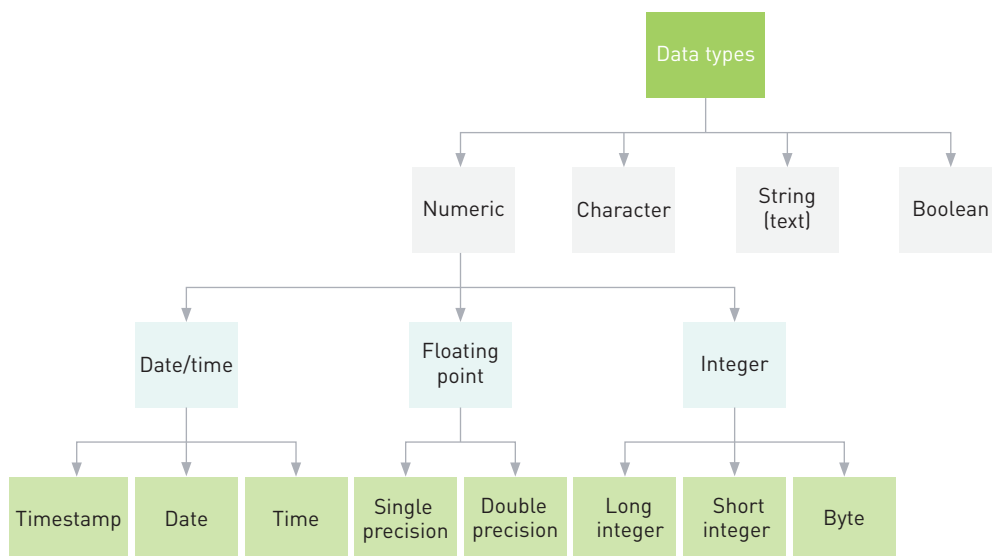


FIGURE 2.24 Common data types used in software tools

Remember: Having only one piece of data in a field is necessary to achieve first normal form (1NF).

Data type tips

Information on data types is also included in Chapter 1. The following is additional information that you may find useful.

- Storing a date as date data type will allow databases, spreadsheets and programs to extract parts of the date (year, month and day).
- A field's data *type* is separate from its data *format*. A field of **date** type can be displayed as 1999/03/28 in a list onscreen, and 28 March 1999 on a printed certificate; one half may be shown as 0.5 or 50%.
- A **timestamp** data type contains both a date and a time of day.
- Do not confuse a **time of day** with a time **duration**. A song's length should be entered as a number. For example, 80 seconds = 1 minute 20 seconds, not a time of day. For example, 1:20 = twenty past one.
- None of the integer type can store fractional data.
- A value such as 1:20 is meaningless to a database. Similarly, 1 minute 20 seconds can only be stored as text, and the software will not be able to interpret the value.
- Small \$1.40, large \$2.20 can be stored and displayed, but the numbers will never be accessible for calculation or individual display.

Spreadsheets

Spreadsheet applications such as Microsoft Excel essentially guess what data type to use, and sometimes guess incorrectly. If you enter 01/02 into a cell in Excel, it will be interpreted as a date. If you enter an apostrophe first, you can force Excel to treat it as text; that is, '01/02. Alternatively, you can change the cell's format to 'Text' to prevent the problem.

Only enter one piece of data (or one datum, the rarely used singular for data) of a single data type into a spreadsheet cell or data field. Do not store units, such as seconds or litres, in cells with data. Put the units in the column's heading, such as Capacity (litres).

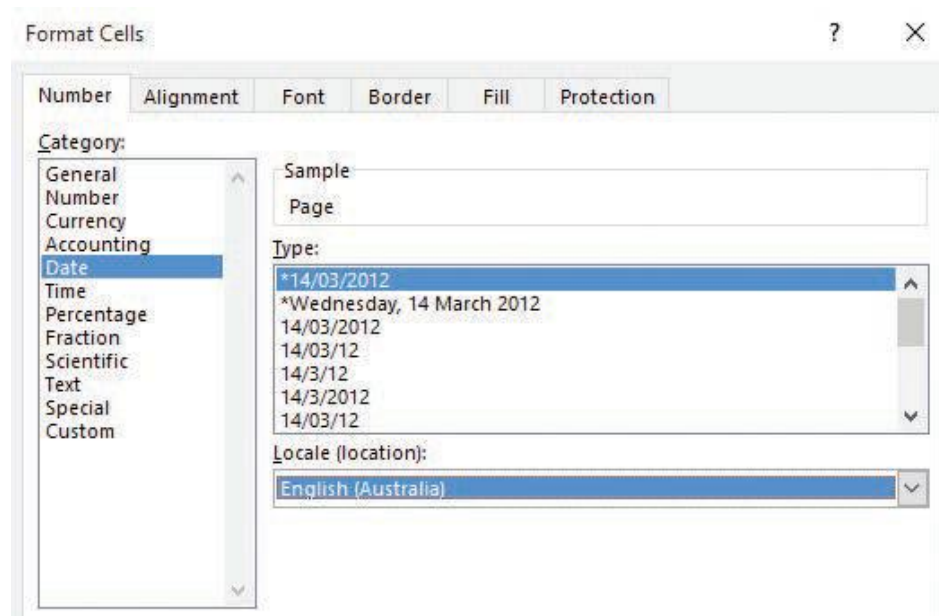


FIGURE 2.25 Formatting the appearance of dates in Microsoft Excel

Data structures

Databases such as Microsoft Access and FileMaker have formal fields, records and table structures, while spreadsheets such as Microsoft Excel do not. In spreadsheets, cells do not need to be defined as containing a certain data type. You can format a cell's appearance, but its data type cannot be specified and imposed. The nature of data in a cell can remain ambiguous until a specific type of arithmetic or function is used to manipulate it, such as date, number or text.

Database structures, including fields, records, and tables are discussed in detail in Chapter 1.

C2			<i>fx</i>	=B2+7
	A	B	C	D
1		Data	Plus 7	
2	Date format	25/12/2016	1/01/2017	
3	Number format	42729.00	42736.00	
4				

FIGURE 2.26 A spreadsheet handling the same data (column B) as dates and numbers

Next steps

As you collect your data, take steps to protect your respondents and subjects, maintain your data's integrity and apply appropriate data types and structures, you should begin to think about relevant legal constraints. Chapter 3 begins with a discussion of legal requirements related to storage and communication of data and information.

ESSENTIAL TERMS

authority a respected reputation for being trustworthy or knowledgeable on a subject; for example, a university professor or an expert in the relevant field

chain of custody the full history of a piece of data, including its origin, acquisition, custody, control, transfer, analysis and processing; a concept important to police, who must be able to account for every piece of criminal evidence from the moment it is found until the time it is used in court

citation a reference to the information's source

closed question a question that offers a limited range of possible answers

field notes open-ended recordings of observations that were not anticipated when the research began, which add details to the observation but do not answer any of the researcher's pre-planned questions or topics

eXtensible Markup Language (XML) a standardised way to add metadata tags to raw data to give it a self-descriptive structure and allow it to be read, queried and processed by any database software

health service assessing, recording, maintaining or improving a person's health; or diagnosing or treating a person's illness or disability; or dispensing a prescription drug by a pharmacist

hidden agenda an undisclosed or disguised true ambition or ulterior motive

inform to give essential guiding information

interpolation using a trend to estimate values within known values in a data set or sequence of values

leading question question that steers answers in certain directions

loaded question question that contains unjustified assumptions that often aim to bias answers

multimodal media consisting of more than one communication mode, including text, images, video, audio, animation, interpretative dance, mime, gesture, icons, sign language, braille, morse code, semaphore, speech, cartoons and music

open question a question that does not limit the length, type or content of answers

paraphrasing using another person's ideas but putting them into your own words, rather than directly quoting them

primary data new facts collected personally by a researcher to answer a specific question

secondary data facts collected by someone other than the researcher, which have often been processed

subjective based on opinion or personal judgement; for example, 'Is that salesperson friendly?'; the opposite of objective

vested interest a situation where an individual, group or organisation has an investment or interest in an issue and stands to gain or lose from it

IMPORTANT FACTS

- 1 **Data** is made up of raw, unprocessed facts and figures.
- 2 **Information** is derived from processing data into a form humans can understand.
- 3 There are many government and non-government **public sources of data sets** in Australia.
- 4 Interviews use **open questions** to gain detailed, personal opinions.
- 5 **Observation** gathers information about people acting naturally.
- 6 Surveys and questionnaires usually use **closed questions** to gather a lot of primary data.
- 7 Choosing samples of a population to interview or survey must be done carefully to avoid **bias**.
- 8 Databases, websites and other data sources can be **queried** using correctly formatted requests (in QBE, SQL or API).
- 9 **Quantitative data** is expressed as numbers, categories or labels and is easy to process.
- 10 **Qualitative data** is rich, unstructured textual data containing opinions and must be encoded to be processed.
- 11 **Likert scales** are often used to encode quantitative data.
- 12 Qualitative data may use **rubrics** to guide how free-form ideas are converted to labels or numbers.
- 13 **Data types** include number, character (text, string) and Boolean (true/false). The data types of database fields must be chosen with care.
- 14 Use one of the following four **referencing systems** for your Outcome: Harvard, American Psychological Association (APA), Chicago, or Institute of Electrical and Electronics Engineers (IEEE).
- 15 Proper referencing is needed to avoid **plagiarism**.
- 16 The **parenthetical referencing style** (Harvard, APA, Chicago science) inserts the author and year of a reference into the body text.
- 17 The **numbered referencing style** (IEEE, Chicago humanities) inserts a reference number into the body text.
- 18 A **reference list** contains full details of every reference given in the body text.
- 19 The federal *Privacy Act* (amended in 2012) regulates the handling of personal information by eight types of organisations – government and private. It is based on 13 Australian Privacy Principles.
- 20 The *Privacy and Data Protection Act 2014* replaces the *Information Privacy Act* and governs the handling of **personal information** by agencies of the Victorian state government.
- 21 The *Victorian Health Records Act 2001* applies to all Victorian government and private organisations that hold **health information**.
- 22 The federal *Copyright Act 1968* protects the **rights of owners of intellectual property**, including electronic forms like software, websites and digital games, books, music, documents and movies.
- 23 **Personal information** can identify a person; for example, name and phone number.
- 24 **Sensitive information** includes data on medical history, politics, sex, religion, race and so on.
- 25 **Health information** includes medical history, diagnoses, drug prescriptions and genetic information.
- 26 The *Victorian Charter of Human Rights and Responsibilities Act 2006* protects the privacy, reputation and freedom of worship, belief and speech of individuals.
- 27 The federal *Spam Act 2003* forbids unsolicited commercial emails, and imposes restrictions on the use of other commercial messaging.



TEST YOUR KNOWLEDGE



Review quiz

PRIMARY AND SECONDARY DATA

- 1 Identify the major difference between data and information.
- 2 Identify the main characteristics and strengths of primary data.
- 3 Identify the main characteristics and strengths of secondary data.
- 4 List five sources of primary data, and five sources of secondary data.

QUANTITATIVE AND QUALITATIVE DATA

- 5 How can qualitative data be prepared for statistical manipulation?
- 6 Give an example of a question using a Likert scale.
- 7 Explain the usefulness of a rubric.
- 8 Contrast qualitative and quantitative data.

ACQUIRING DATA

- 9 State reasons why you would avoid using loaded and leading questions.
- 10 Summarise the advantages of using interviews to collect data.
- 11 When is observation the most appropriate data acquisition method?
- 12 Explain one advantage that surveys and questionnaires have over observation and interviews.
- 13 Explain how QBE and SQL differ.
- 14 Write a query that would find males between the ages of 16 and 18 inclusive, and females older than 25.

REFERENCING DATA SOURCES

- 15 How does proper referencing avoid plagiarism?
- 16 Give examples of using parenthetical (author, year) and numbered reference styles in body text and reference lists.

INTEGRITY OF DATA

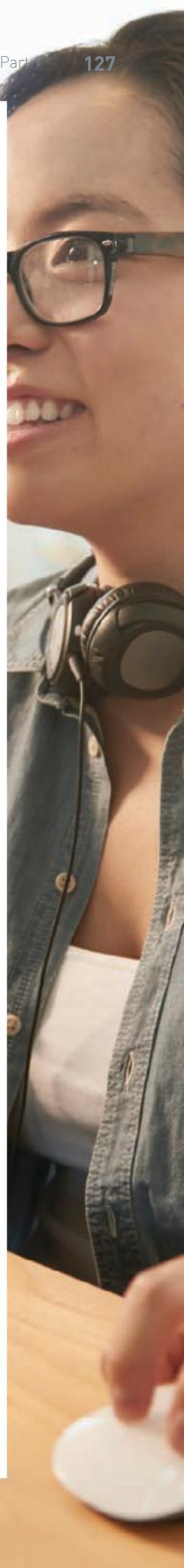
- 17 Summarise the meaning of 'integrity of data'.
- 18 How does lack of timeliness degrade the value of data?
- 19 How does incomplete data lead to faulty information?
- 20 How does database normalisation contribute to data accuracy?
- 21 Describe one way to discover if people are not telling the truth in surveys.
- 22 What does 'data authenticity' mean?
- 23 List three examples of how data may lose relevance.
- 24 What can be done to maintain the accuracy of data over time?
- 25 What are the main differences between open and closed questions, and when should each be used?

DATA TYPES AND DATA STRUCTURES

- 26 Why is it important to choose proper data types?

APPLY YOUR KNOWLEDGE

- 1** Choose a hypothesis for research. Consider data sources to refute or support your hypothesis. Your classmates could be useful primary sources. Examples of potential topics that could be honed into hypothesis could be:
 - teenagers prefer one type of public transport
 - soccer is rising in popularity
 - your school compares well with neighbouring schools
 - internet shopping is better than personal shopping
 - any other operating system is better than Windows.
- 2** The nature of the hypothesis will determine the data types you need. You may consider creating a set of interview questions for a few (for example, three) of your peers to gather qualitative primary data about your topic. Other types of primary data may include photos and observation. For secondary data, you may need to access journals, newspaper articles, television reports, radio programs and statistical evidence.
- 3** Create a survey or questionnaire for a larger number (for example, 10) of your peers on the same topic.
- 4** Conduct the interviews and distribute the surveys or questionnaires. Record any problems you encountered acquiring this information. What questions may have needed changing?
- 5** Encode the interview data to let you summarise it into meaningful categories and trends.
- 6** Write a report on the topic containing:
 - a** a statement of the hypothesis
 - b** results of interviews
 - c** supporting quantitative secondary data from printed or online sources, fully referenced using Harvard, Chicago, APA or IEEE; primary data must also be referenced
 - d** a statement evaluating the quality and integrity of the data used to generate information about the hypothesis
 - e** a statement explaining how the data will be stored and evaluated securely
 - f** a conclusion about the hypothesis using information derived from your data.



DATA ANALYTICS: DRAWING CONCLUSIONS, PART 2

Key knowledge

After completing this chapter, you will be able to demonstrate knowledge of:

Approaches to problem solving

- project management concepts and processes, including milestones and dependencies (concepts), task identification, sequencing, time allocation, resources and documentation using Gantt charts (processes)
- file naming conventions to support efficient use of software tools
- software functions to organise, manipulate and store data
- techniques for identifying patterns and relationships between data

Digital systems

- roles, functions and characteristics of digital system components used to input, store, communicate and output data and information
- physical and software security controls suitable for protecting stored and communicated data

Interactions and impact

- key legal requirements for storage and communication of data and information, including privacy, intellectual property and human rights requirements

For the student

This chapter concludes the discussion of the theory and skills required for Unit 3, Outcome 2.

You will be introduced to the key legal requirements for storage and communication of data and information, techniques and tools for planning and managing the progress of a complex project, identifying patterns and relationships in data, hardware and software, and data security.

By the end of this chapter, you will be ready to frame a hypothesis, gather data to support it, report your findings and cite the data sources you used.

For the teacher

This chapter concludes the theory and skills needed for Unit 3, Outcome 2. Having covered the theory of data in Chapter 2, students are now introduced to the formulation of a reasonable hypothesis, the gathering of quality data to support the hypothesis, project management, computing equipment and data security.

By the end of this chapter, students should be equipped to devise their own hypotheses, search for relevant data from which they derive information to support the hypothesis and report their findings with evidence and a formal citation of sources. Note: You do not provide students with a hypothesis or data.

Continuing Unit 3, Outcome 2

In Chapter 2, you learned how to formulate a reasonable, testable hypothesis. You learned about narrowing the scope of a solution and to identify the relevant constraints. You also learned about data: primary and secondary data, the difference between qualitative and quantitative data, as well as how to protect the integrity of your data and how to reference sources correctly.

These factors are all relevant preparation for Unit 3, Outcome 2. For this Outcome, you will submit the information you have produced from the data you have gathered, which supports or refutes your hypothesis.

This chapter begins with a discussion of the key legal requirements surrounding the storage and communication of data and information. As a collector of data and information, these requirements are relevant to you because you must make sure you follow the law.

To prepare you for Outcome 2 and also for Unit 4, Outcome 1, we will then discuss planning and managing a complex project, including how to use a Gantt chart, because this relates to the project you are undertaking.

We will also talk about identifying patterns and relationships in data, because this will help you to interpret and analyse the data you have gathered for your Outcome.

Finally, we will also discuss components of digital systems (hardware and software), as well as data security.

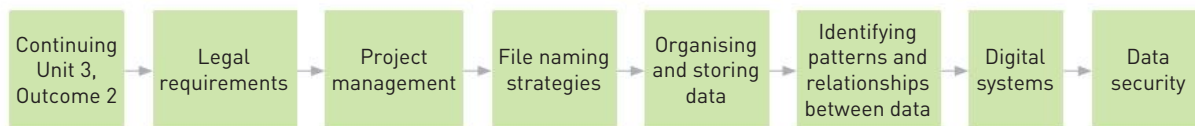


FIGURE 3.1 Chapter map

Legal requirements

Being in possession of other people's personal data and information is a heavy responsibility that you must take seriously. You need to ensure you collect, use and store that data and information in a way that obeys the law, and protects the privacy of your respondents and subjects.

While this chapter deals specifically with Unit 2, Outcome 3, you must *always* protect the privacy of your respondents and subjects. Different laws govern how data and information can be collected, stored, communicated, protected, used and disclosed.

The key legislations, as shown in Table 3.1 and Table 3.2, are:

- *Privacy Act 1988*
- *Privacy and Data Protection Act 2014* (Vic)
- *Health Records Act 2001* (Vic)
- *Copyright Act 1968*
- *Charter of Human Rights and Responsibilities Act 2006* (Vic)
- *Spam Act 2003*.

Both the *Privacy Act 1988* and *Privacy and Data Protection Act 2014* allow exempt organisations to opt-in and become subject to the laws. Although it may seem unusual to choose to become bound by laws, some organisations want to impress the public with the assurance that they take the privacy of individuals so seriously that they volunteer to be subject to laws.

TABLE 3.1 Summary of key federal legislations

Legislation	Effect	Applies to	Does not apply to
<i>Privacy Act 1988</i> including <i>Privacy Amendment (Enhancing Privacy Protection) Act 2012</i>	Regulates handling of personal information about individuals	Federal government departments Private businesses with federal government contracts Health service providers Businesses that buy or sell personal information Credit reporting agencies Banks and credit card companies Energy, water and telecommunications providers Organisations that handle tax file numbers Organisations that choose to opt-in	State/territory public hospitals Individuals not involved in business Public schools and universities Media organisations acting in the course of journalism Political parties and politicians
<i>Copyright Act 1968</i>	Protects the rights of individuals who produce creative and artistic works (intellectual property)	All Australians using other people's intellectual property	
<i>Spam Act 2003</i>	Prevents the sending of spam and harvesting of email addresses; regulates the sending of commercial emails	All Australians who send unsolicited commercial emails	

TABLE 3.2 Summary of key Victorian (state) legislations

Legislation	Effect	Applies to
<i>Charter of Human Rights and Responsibilities Act 2006</i> (Vic)	Protects the rights of individuals to freedom of privacy, thought, belief and expression	All Victorians
<i>Privacy and Data Protection Act 2014</i>	Enforces the Information Privacy Principles on Victorian public service organisations	Victorian state government departments Groups working with Victorian state government departments, including private companies, local governments (councils) and state schools Organisations that choose to opt-in Does NOT apply to police
<i>Health Records Act 2001</i> (Vic)	Protects the privacy of health records	All Victorian organisations and state government departments that hold health records

Privacy legislation

Several federal (Commonwealth) and state laws govern information privacy. All of these are based on the Australian Privacy Principles (APPs). The types of information covered by information privacy can include personal information, sensitive information and health information. Figure 3.2 lists the information that is governed by privacy legislation.

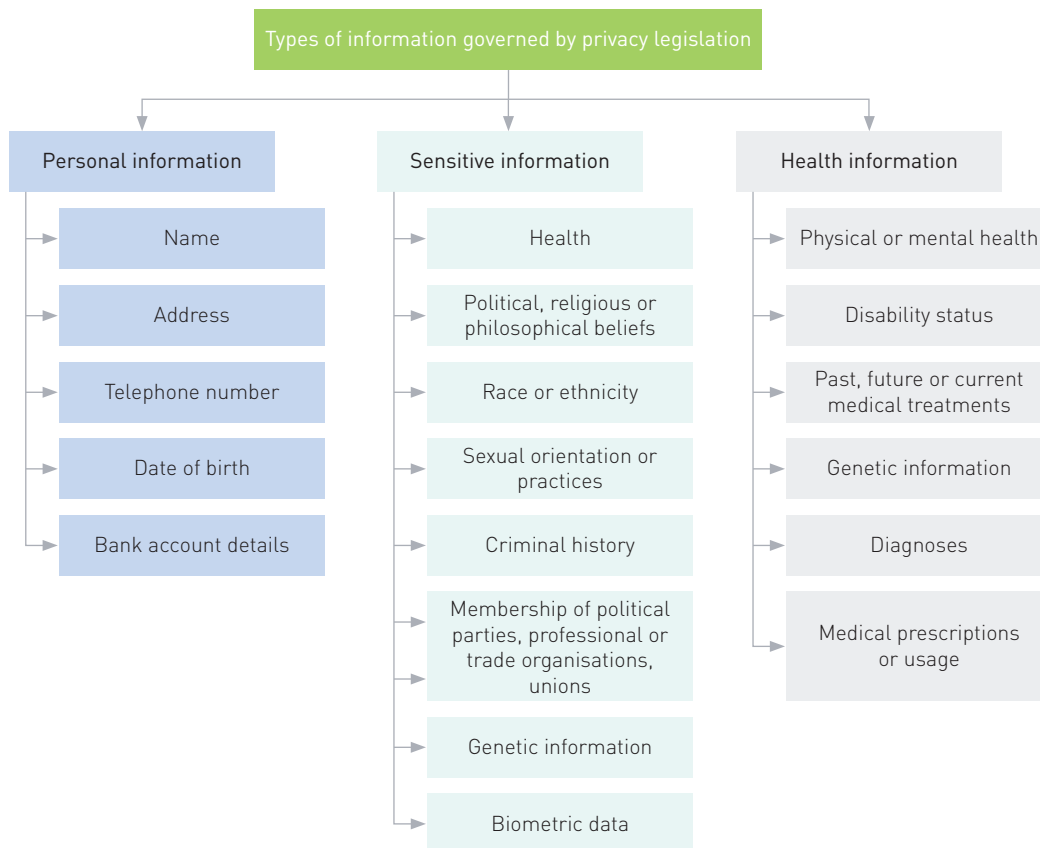


FIGURE 3.2 The types of information governed by privacy legislation

Note that any holder of health information is required to abide by privacy legislation, even if providing a health service is not their primary activity. This includes private hospitals, day surgeries, doctors, dentists, pharmacists, psychiatrists, complementary therapists (for example, naturopaths, chiropractors and massage therapists), gyms and weight loss clinics, child care centres, private schools and universities.

Australian Privacy Principles (APPs)

All privacy legislation in Australia is underpinned by the thirteen APPs. For a list of the APPs, see Chapter 5, page 259–60. The APPs are discussed in brief here, but are discussed in greater depth in Chapter 5.

Most of the APPs have common-sense exceptions. For example:

- if the person consents to the use of the information
- the steps required to comply only have to be ‘reasonable’
- Australian law requires the information to be disclosed
- no APP applies to information held by people relating to their family affairs
- to protect the public interest, health, safety or security
- sharing the information may prevent a serious crime

- it is unreasonable or impracticable to obtain the individual's consent
- to find a missing person
- in court to establish a legal claim, defence, or to resolve a dispute
- during foreign wars, disasters, or emergencies.

Victorian privacy legislation uses 10 Information Privacy Principles (IPPs) that mirror the APPs, except for IPP 10, which differs in how it deals with sensitive information.

TABLE 3.3 Victorian Information Privacy Principles. Note that the IPP 10 is shown in greater detail here because it does not mirror the APPs.

IPP	Effect
1	Collection of personal information
2	Use and disclosure of personal information
3	Data quality
4	Data security
5	Openness
6	Access and correction
7	Unique identifiers
8	Anonymity
9	Transborder data flows
10	Sensitive information: Collect sensitive information only with the person's consent or if required by law. Consent must be voluntary, informed, specific and current.

Specific legislation

The following privacy legislation, also touched on in Table 3.1 and Table 3.2, is relevant for the purposes of your Outcome. You will need to be aware of the Privacy Act in particular when handling personal data and information that belongs to your respondents and subjects.

- *Privacy Act 1988*, including *Privacy Amendment (Enhancing Privacy Protection) Act 2012*. The federal *Privacy Act 1988*, summarised in Table 3.1, is composed of all the APPs and regulates the usage of personal information. If you breach the Act repeatedly or severely, it can result in large fines. The *Privacy Amendment (Enhancing Privacy Protection) Act 2012* introduced the APPs and added several new features, including credit reporting, as described in Table 3.1.
- *Privacy and Data Protection Act 2014* (Vic). This Victorian Act regulates how state government bodies, councils, schools and similar organisations may collect and use personal information. It is very similar to the federal *Privacy Act 1988*.
- *Health Records Act 2001* (Vic). The Victorian Act, which mostly mimics the Privacy Act's rules about the handling of health information, applies to all Victorian organisations, both private and public (government).

Spam Act 2003

Part 1.3 of the Spam Act regulates commercial email and electronic messages, including instant messaging. It does not govern faxes, voice calls or packages sent by postal mail. Penalties for violating the Spam Act include fines of up to \$295 000.

- Individuals may consent to receive spam messages. Consent means clicking a button to say 'I agree to these terms and conditions'.
- Unsolicited (unrequested) commercial electronic messages must not be sent at all.
- Commercial electronic messages must identify the sender, and must contain a working unsubscribe facility.

- Address-harvesting software (or address lists created by such software) must not be supplied, acquired or used.
- Exempt organisations include government bodies, political parties and charities. Schools may send commercial electronic messages to past or current students.
- If you ‘conspicuously publish’ your email address, such as on a blog or a website, you may not be protected by the Act.

Although you will not technically be engaging in ‘commercial’ emailing or messaging for the purposes of your Outcome, it may be a good idea to abide by the general principles of the Spam Act regardless. Doing otherwise may make your Outcome research a nuisance to potential respondents and subjects.

Be wary of clicking the unsubscribe links in spam from unknown foreign sources. This may verify your address and increase your spam or lead you to a dangerous website.

Intellectual property legislation

Intellectual property legislation ties into the reasons for referencing your sources. This legislation protects not just other individuals who have created works – it also protects you.

Copyright Act 1968

The *Copyright Act 1968* is a federal law that recognises that any original creative or artistic work is the property of the person who created it. Any person wishing to use another person’s work must obtain permission and/or pay for a licence. The Copyright Act protects the creator of an original work from unauthorised reproduction, conversion, adaptation, transmission or publication of their intellectual property, which includes:

- original literary, dramatic, musical and artistic works
- websites
- software
- electronically recorded music, films and books.

Copyright protection is automatic as soon as intellectual property is created and recorded in a way that can be seen or heard (for example, written, recorded, filmed or put online). You do not have to register for copyright as you do for patents or trademarks. You do not need to use a copyright symbol, ©, or statement, but they are recommended.

You have the right to protect your own original works using technological devices, such as encryption or copy protection.

Without permission from the copyright holder, it is illegal to:

- digitise a non-digital work, such as ripping a CD to MP3, or converting a DVD to an MKV file
- make or import devices or software to bypass copy protection
- remove or tamper with a copyright notice
- share copyrighted material online, such as torrents and Mediafire
- keep or share programs recorded from TV
- publish unauthorised screenshots from some webpages or software.

There are a few limited exceptions for using IP for education, review, satire or fair use.

THINK ABOUT INFORMATICS 3.1

If you create intellectual property as an employee, even outside of business hours, it is usually owned by your employer. Why might this be the case?

Human rights

Charter of Human Rights and Responsibilities Act 2006 (Vic)

The Charter protects the rights of privacy, reputation and freedom of belief and expression of individuals.

Section 13 protects privacy and reputation. It states that individuals have the right *not* to have their privacy, family, home or correspondence unlawfully interfered with, and not to have their reputation unlawfully attacked.

Section 14 states that individuals have the right to freedom of thought, conscience, religion and belief, including the freedom to demonstrate publicly their religion or belief in worship, observance, practice and teaching. Individuals must not be forced to accept a religion, and may not be restrained from believing as they choose.

Section 15 establishes freedom of expression and the right to hold an opinion without interference. Individuals may seek, receive and give information and ideas of all kinds.

Other laws are used to prevent abuses of these freedoms that would be disruptive to public order or vilifying in some way, such as wandering naked through the streets or using racially abusive language.

Project management

Project management is the process of planning, organising and monitoring a project in order for it to be completed on time and within budget. Building or changing information systems for a project can be expensive and disruptive. If managed badly, it can be damaging to an organisation's operations and profit. Large-scale changes are often approached as projects so they can be planned, organised and conducted appropriately to enable them to finish on time, on budget and fully functional. You will formulate a project plan to manage your progress through Unit 3, Outcome 2 and Unit 4, Outcome 1.

For your project to be successful, you need to identify, schedule and monitor tasks, resources, people and time. You are not required to use project management software to do this, though you may use them if you wish. Rather, you should focus on the **concepts** and **processes** of project management.

One of the items you need to submit as part of your Outcome is a **Gantt chart**.

A Gantt chart is a graphic timeline that:

- lists all tasks in a project
- organises the tasks in order
- shows which tasks must wait for other tasks to finish before they can begin
- allocates people and resources to tasks
- tracks the progress of tasks and the entire project.

Although you can create a Gantt chart with a pen or with a spreadsheet, project management software can be easier and faster. In addition, you are required to use an appropriate software tool to create your Gantt chart for Unit 3, Outcome 2. Suitable software includes the commercial Microsoft Project, and the free, easy-to-use GanttProject and ProjectLibre.

When using software to create your Gantt chart, you will not be assessed on your technical prowess with the software. Rather, you will be assessed on how well your Gantt chart demonstrates your understanding of the concepts and processes of project management.

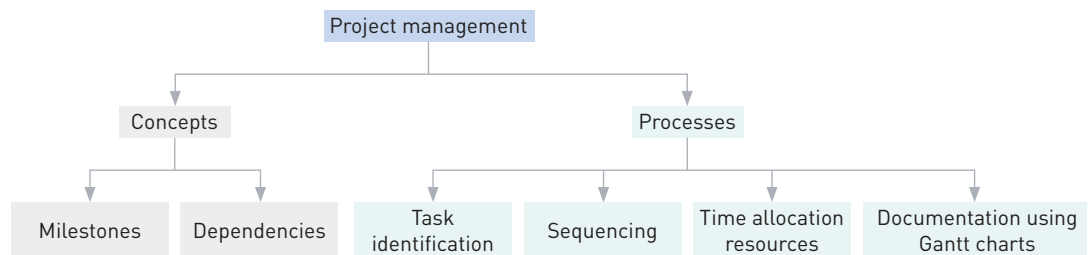


FIGURE 3.3 Key project management concepts and processes

Concepts

Milestones

A milestone represents the achievement of a significant stage in a project and has zero time duration. For example, the completion of the printing of a questionnaire so that it can be distributed to respondents would be a task of zero time and represents a milestone. This follows tasks in which the questionnaire has to be researched, written, proofread and finally printed, all of which take time.

Dependencies

Tasks are interdependent, meaning that they must be completed in a particular order. The commencement of some tasks depends directly on the task that is completed before. For example, you cannot distribute a questionnaire (one task) before writing the questions for it (another task). However, you cannot write the questions for the questionnaire without first determining your hypothesis (a third task). Ultimately, the task of distributing a questionnaire has multiple dependencies. To distribute the questionnaire without first formulating the hypothesis and writing the questions is not possible.

Processes

Task identification

You would break a large project such as Unit 3, Outcome 2 down into discrete tasks, such as the following.

- Research multiple topics
- Decide on a single topic
- Create Gantt chart to identify tasks
- Formulate hypothesis
- Write qualitative questions for interviews
- Conduct interviews
- Write quantitative questions for questionnaires
- Distribute questionnaires
- Locate data sets and secondary data
- Input data into database
- Interpret data and review results
- Conclude whether hypothesis has been supported or refuted.

Note: Not all of the tasks you would undertake for Unit 3, Outcome are included in the list above. You should not use this as your own exhaustive list.

To break down your project into achievable tasks, develop a **work breakdown structure (WBS)** and draw a WBS diagram to accompany it. Do not leave any tasks out of the WBS. For large projects, a WBS will often be hierarchical, breaking major tasks into subtasks and even sub-subtasks. Although this may sound confusing, it will actually keep your tasks organised and in context, allowing them to be collapsed or expanded to view overall task progress or fine details about how minor subtasks are proceeding.

Do not leave any tasks out of the WBS. For example, imagine you distributed both a print and online version of a questionnaire, but did not list the printed version in the WBS, and forgot to go collect the printed forms from respondents. All the gathered data would be overlooked, or would be counted too late.

Figures 3.4a and 3.4b show a sample potential WBS for Unit 3, Outcome 2. Again, the tasks may not be exhaustive and you may find that your own WBS requires additional tasks, particularly if you use different types of data, or need to revise your hypothesis.

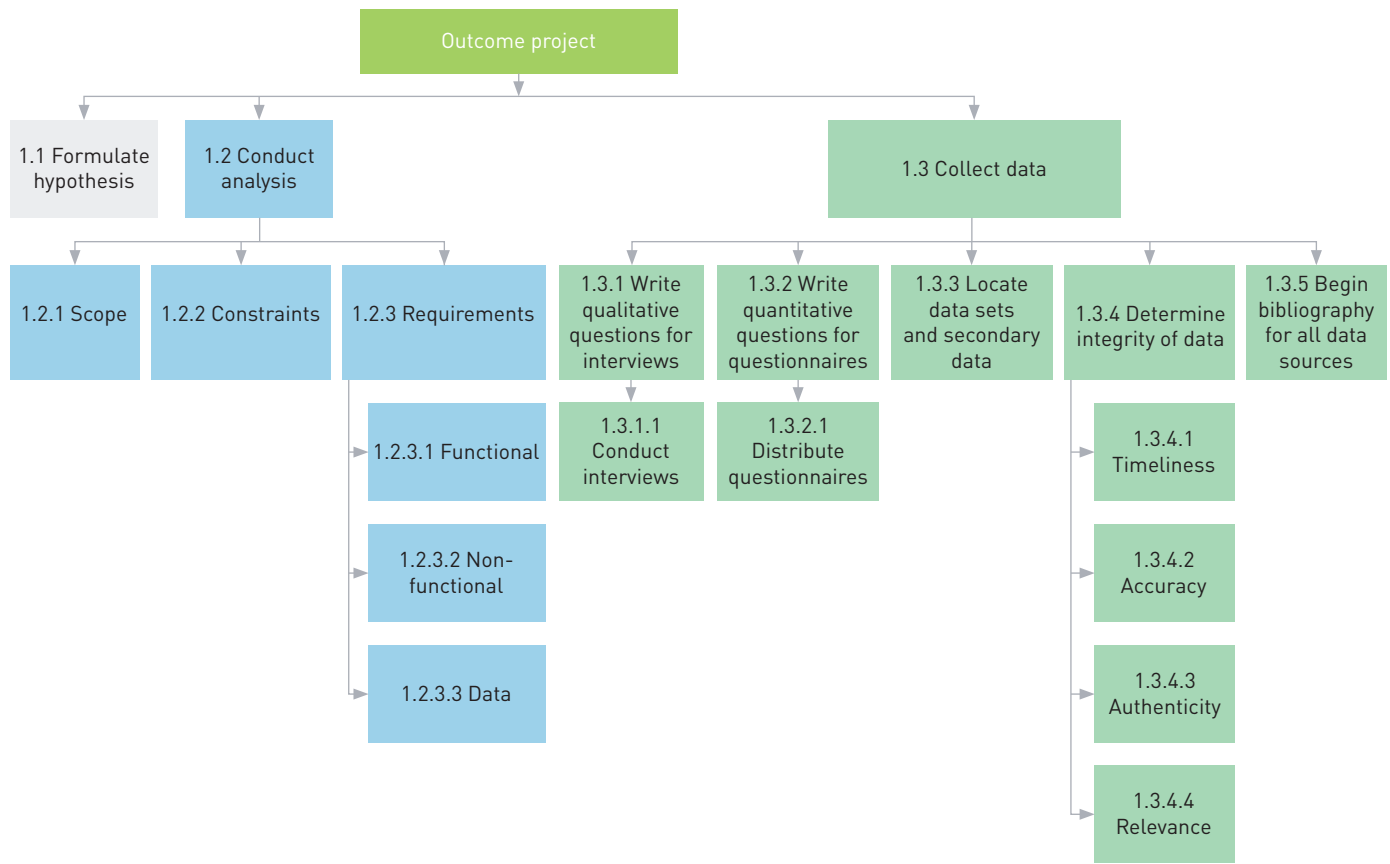


FIGURE 3.4a Sample breakdown structure (WBS) diagram for Unit 2, Outcome 3

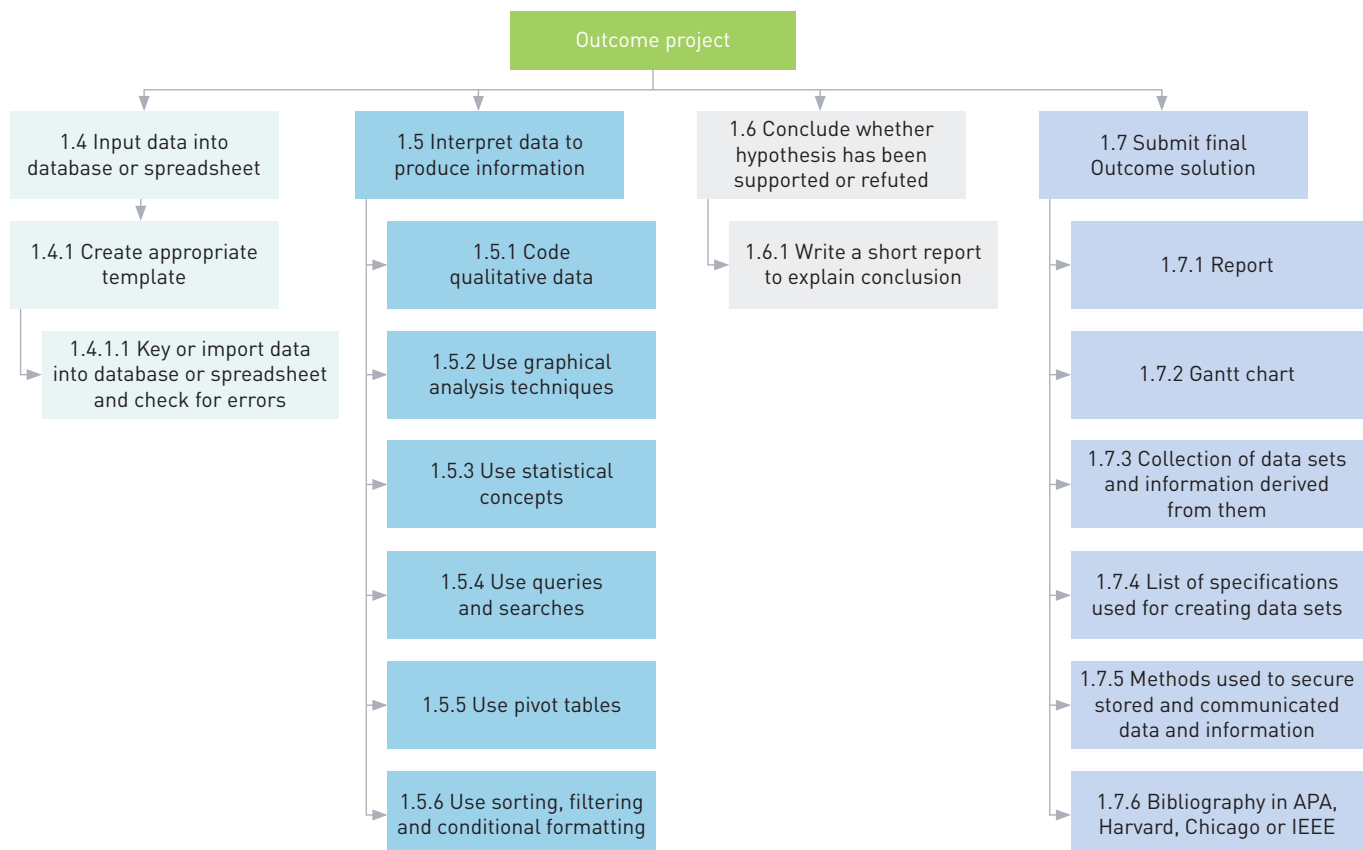


FIGURE 3.4b Sample breakdown structure (WBS) diagram for Unit 2, Outcome 3 (continued)

Sequencing

When you have finished identifying every individual task, you need to decide how long each task will take and then put them in a sequence; that is, arrange them in a particular order. As discussed in dependencies, quite often one task cannot be started before one or more other tasks have already been completed.

Decide which tasks can be worked on **concurrently**, but are dependent on other tasks that have been completed already. For example, you could work concurrently on conducting interviews, researching data sets and writing quantitative questions for a questionnaire, but they are all dependent on the hypothesis having been formulated first.

Similarly, you cannot write the report concluding whether your hypothesis has been supported or refuted before you have interpreted your data.

Tasks that must be completed before another task can begin are called **predecessors**.

The dependent tasks are called **successors**.

If a predecessor runs overtime, all of its successors will be delayed, causing problems for other tasks and deadlines. This is where a Gantt chart becomes very useful – it helps to monitor tasks and reach deadlines on schedule. It also helps you to visualise the problems that will occur down the line if a predecessor is late.

The length of time that a task runs overtime before it affects other tasks is called slack time. When workers have **slack time**, you can reassign them to other tasks.

Time allocation resources

A Gantt chart shows tasks as horizontal bars. Each horizontal bar is of a length proportional to its task's duration. A very short task will have a very short bar, while a very long task will have a very long bar. Figure 3.5 displays a number of features typical to a Gantt chart.

The names of each task are shown in the left pane, along with start and end dates, while the right pane shows task timelines. Tasks that overlap in time are concurrent and can be carried out at the same time using different teams.

Arrows are used to indicate dependency. For example, neither 'Write qualitative questions for interviews' nor 'Research data sets' can begin until 'Formulate hypothesis' has finished, because they depend on this task.

The diamond shape indicates a milestone. Milestones are points of significant progress in a project. They are often the start or end of major stages, and can be used to monitor whether a project is on track. A milestone is an **event** with zero-duration and no allocated resources. It is simply shown as a diamond-shaped 'task'. In this instance, 'Submit final solution', the end of the project, is one such milestone.

An event differs from a task because though something happens (for example, a major task ends), no resources, work or time are allocated for it because there is nothing that people need to do to make it happen.

Also notice that 'Input collected data' begins before both 'Research data sets' and 'Distribute and circulate quantitative questionnaires' have finished, even though the task is dependent on both of these predecessors. This is because 'Conduct interviews' *has* finished, so you can get a head start by inputting the data from the interviews while continuing to look at data sets and waiting for responses to questionnaires.

A project's critical path is the sequence of tasks from the beginning to the end, which:

- contains no slack time, therefore any delay in a task on the critical path will affect the ending date of the project
- is the longest duration
- is the minimum possible time in which all of the project's tasks can be completed.

While you can use Gantt chart software to identify the critical path, in the example from Figure 3.5, the critical path comprises the following.

- Formulate hypothesis
- Write qualitative questions
- Conduct interviews
- Input collected data
- Interpret data
- Write conclusion
- Write/finalise remaining Outcome components

Sometimes, more than one critical path is possible. No task on the critical path can have its duration changed without affecting the end date of the whole project.



NelsonNet additional resource: Figure 3.5 Sample GanttProject chart, p. 138

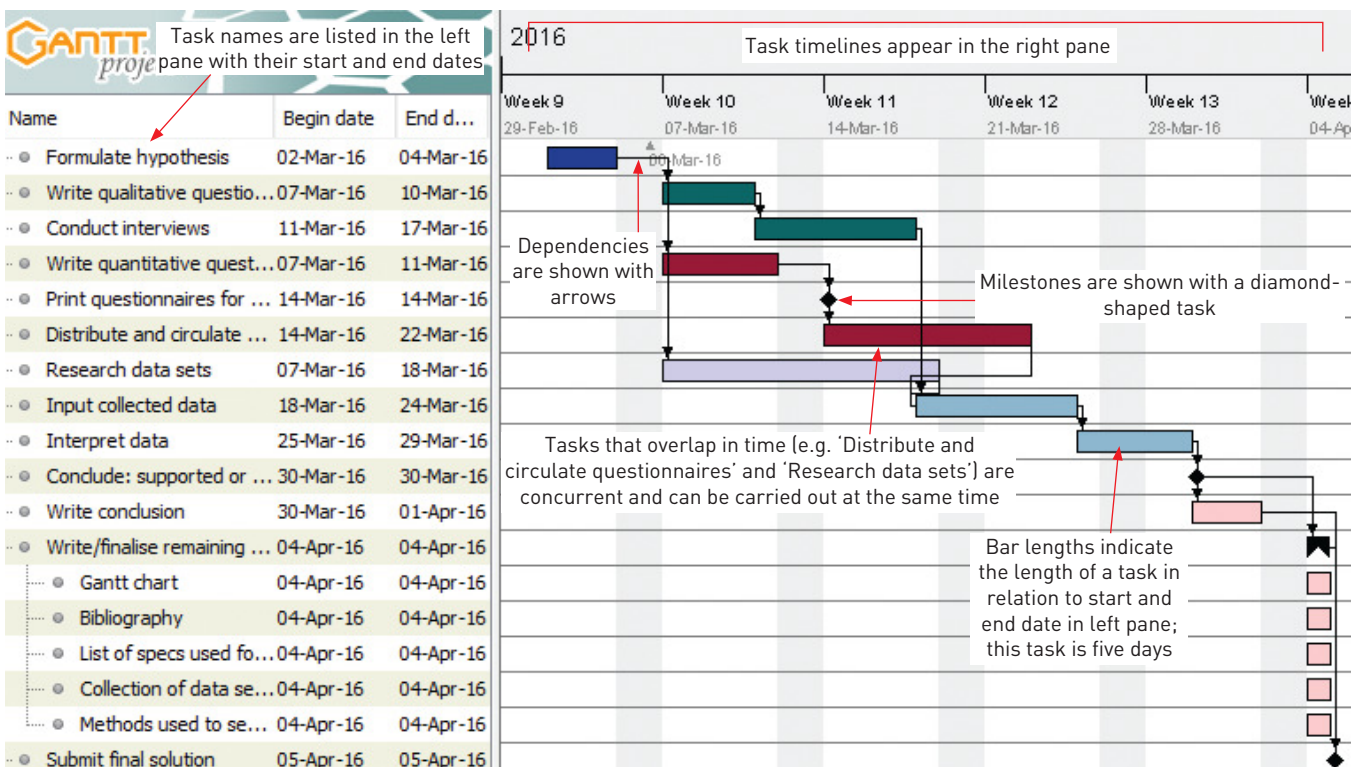


FIGURE 3.5 An annotated Gantt chart. This Gantt chart provides a partial sample model for Unit 3, Outcome 2, with placeholder dates.

THINK ABOUT INFORMATICS 3.2

Project management tools are useful to find the perfect number of people needed on a task so it is finished as quickly as possible without anyone being idle. Use software to develop a Gantt chart to plan the baking of a cake. Assume you can use as many cooks as you want.

Documentation using Gantt charts

You can use the Gantt chart you develop to mark your progress throughout your Outcome. By including information about the progress of the task and the planned versus the actual duration of the task you will help to keep yourself on track.

To manage your solution effectively as a complex project, you should also use your Gantt chart to document the resources you have allocated to it, such as any tools and equipment. While a company or organisation may list consultants and buildings as resources, your resources might be computers, particular data sets and software tools.

You should also frequently modify your Gantt chart so you can reflect contingencies. A contingency is an unforeseen event, incident or emergency. You may find that your interview subjects suddenly become unavailable or unwilling to be interviewed. The data set you wish to use is no longer available. The website hosting your questionnaire crashes and loses your data.

Your Gantt chart should show problems like this and how you react to them, such as finding new interview subjects, different data sets, or switching to paper-based questionnaires.

You should keep your Gantt chart updated throughout both Unit 3, Outcome 2 *and* Unit 4, Outcome 1. You will submit the project plan, indicating times, resources and tasks, in Unit 3, Outcome 2. You will submit the assessment of the plan in Unit 4, Outcome 1.

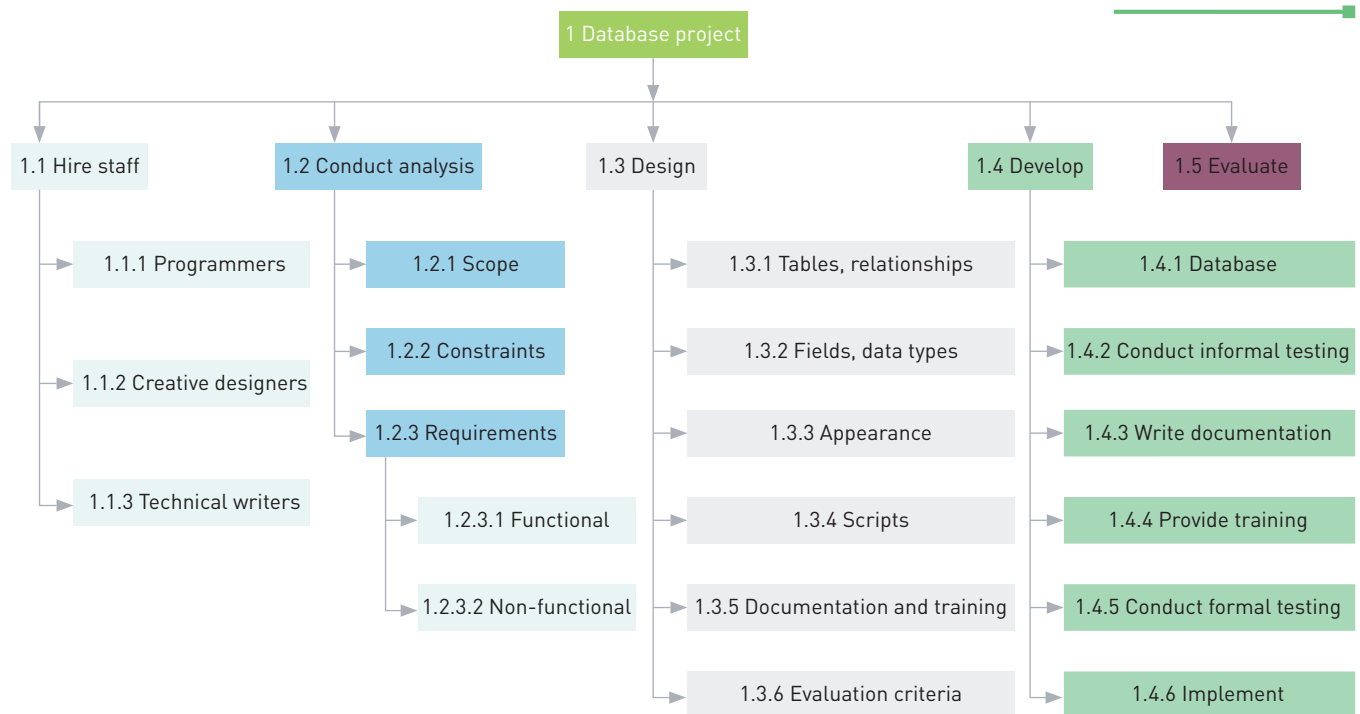
CASE STUDY

GANTT CHART FOR CREATING A DATABASE

A small company of web developers are working on a database project together. As part of their project, they need to build a Gantt chart.

Task identification

They first identify the tasks they need to complete using a WBS diagram:



THINK ABOUT INFORMATICS 3.3

In terms of project management, research the meaning of:

- an 'optimistic' task duration
- a 'pessimistic' task duration.

◀ Case study continued

Next, they enter these tasks into their chosen Gantt software, GanttProject:

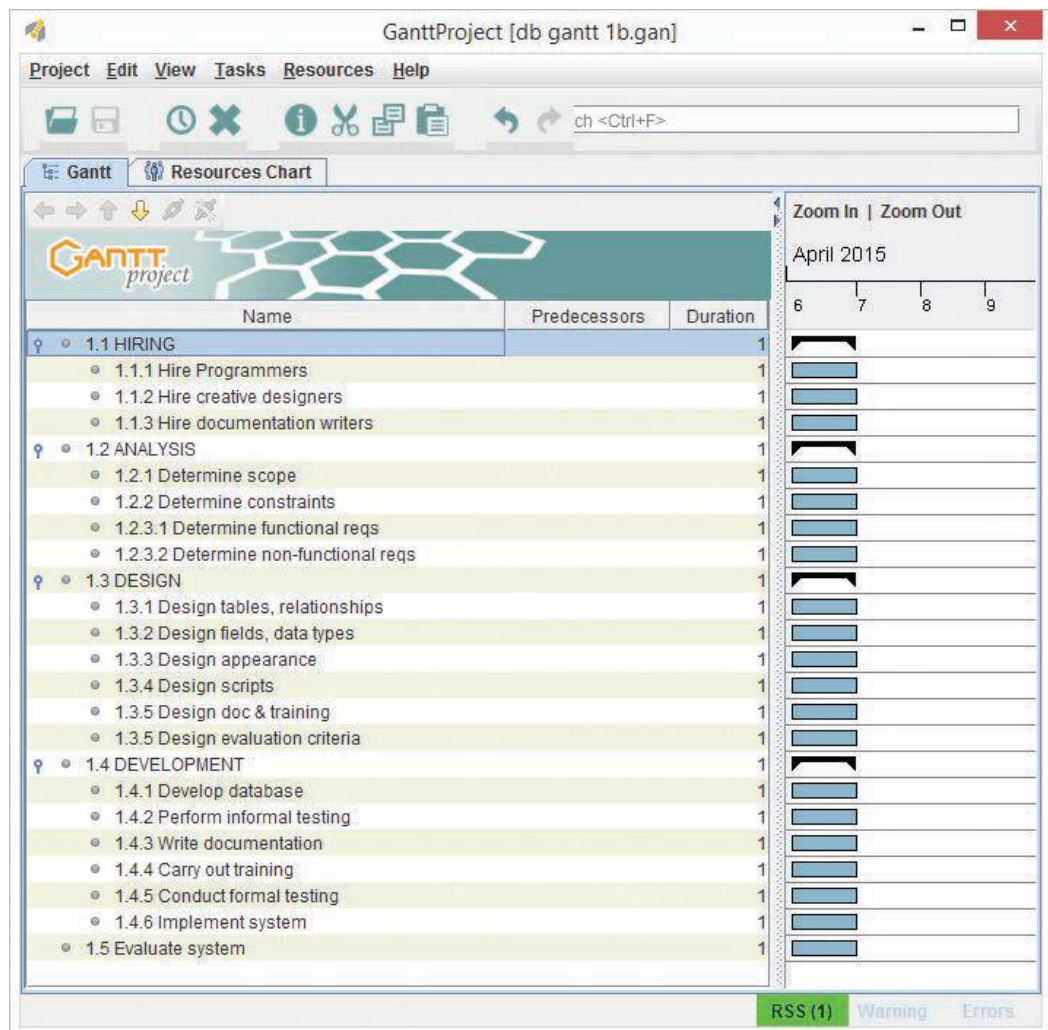


FIGURE 3.6 Entering tasks into their Gantt software

They use a hierarchical structure to group tasks under headings, such as HIRING and ANALYSIS to make task management easier. Groups of tasks can be collapsed or expanded or moved as a group. In GanttProject, multiple levels of sub-tasks can easily be created by just indenting them in the task properties.

Sequencing

The tasks are sufficiently sequenced, but the order can be shifted easily if needed. The developers start creating dependencies, forcing dependent tasks to wait until their predecessors have finished.

- All of the HIRING (1.1) tasks can start immediately.
- Management will complete the tasks in the ANALYSIS group (1.2), which can begin immediately and run concurrently with the hiring tasks.
- The DESIGN (1.3) tasks cannot begin until the ANALYSIS tasks are complete, so DESIGN is made dependent on ANALYSIS.
- The DEVELOPMENT (1.4) tasks cannot begin until DESIGN is finished, so DESIGN is added as a predecessor to DEVELOPMENT.
- Evaluate system must wait for everything else to finish, so it is made dependent on DEVELOPMENT.

Case study continues ▶

◀ Case study continued

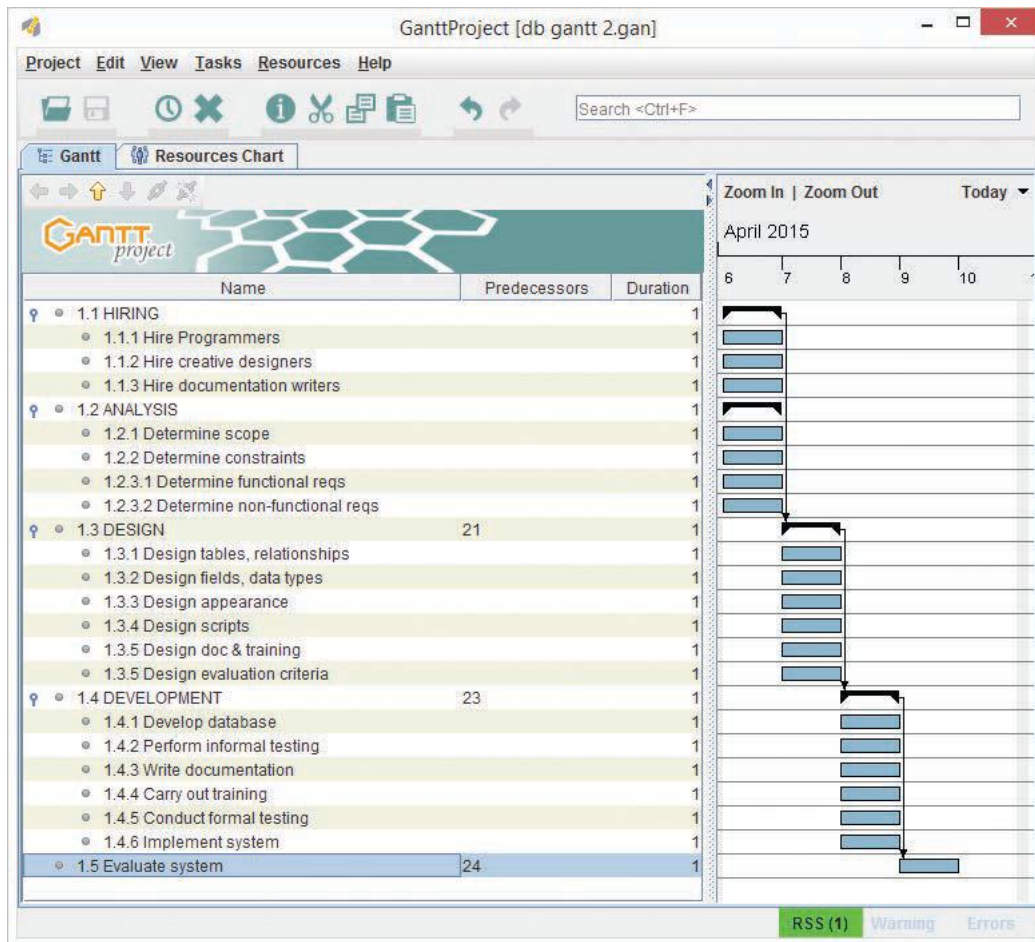


FIGURE 3.7 Major dependencies have been added. Arrows lead from predecessor tasks to dependent tasks.

Subtler dependencies can now be added.

- The team wants DESIGN to be well underway before they start creating documentation and training, because they would be greatly affected by changes to the database. They make 1.3.5 dependent on other design tasks being finished.
- They also make 1.4.4 (Carry out training) dependent on 1.4.3 (Write documentation) being finished.
- Formal testing (1.4.5) must follow all database creation tasks, so they add another dependency.
- Implementation (1.4.6) comes as the last stage of development.

They are now happy with the logical task sequences and dependencies, but they know that if their needs change later, they can still easily adapt the chart to suit their needs.

◀ Case study continued

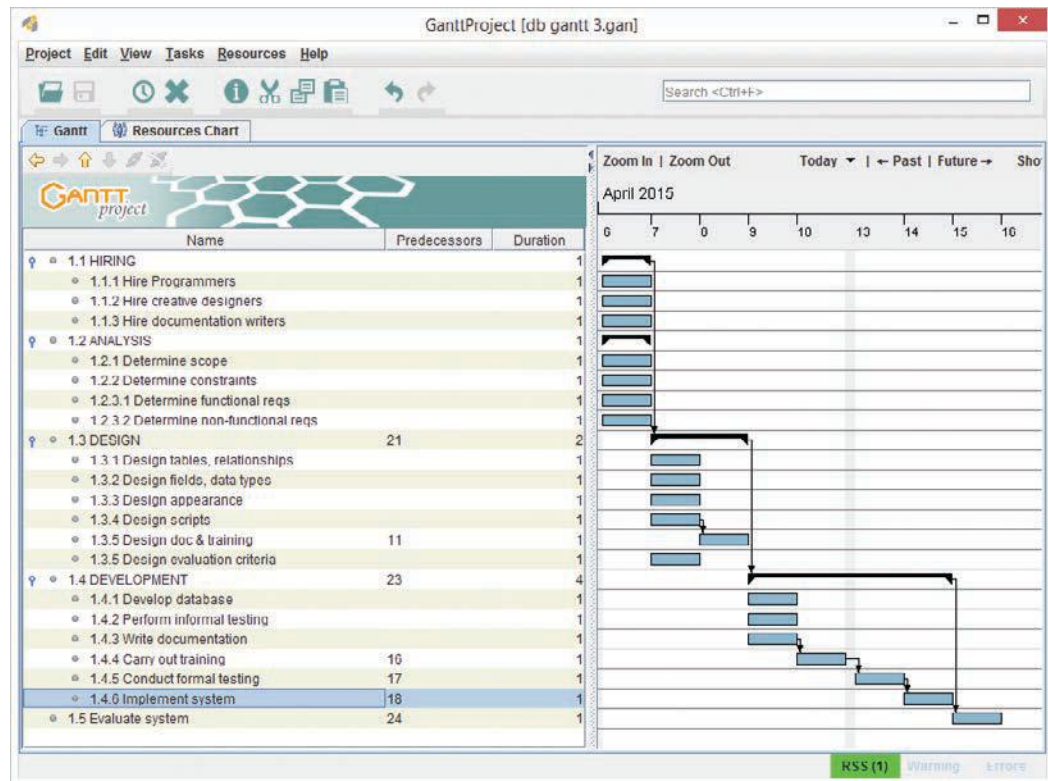


FIGURE 3.8 Dependencies have been created

Time allocation resources

The project developers now tackle the challenge of the time required for each task. They consult with experts and colleagues, and use their extensive experience and knowledge of past projects to guide their estimates.

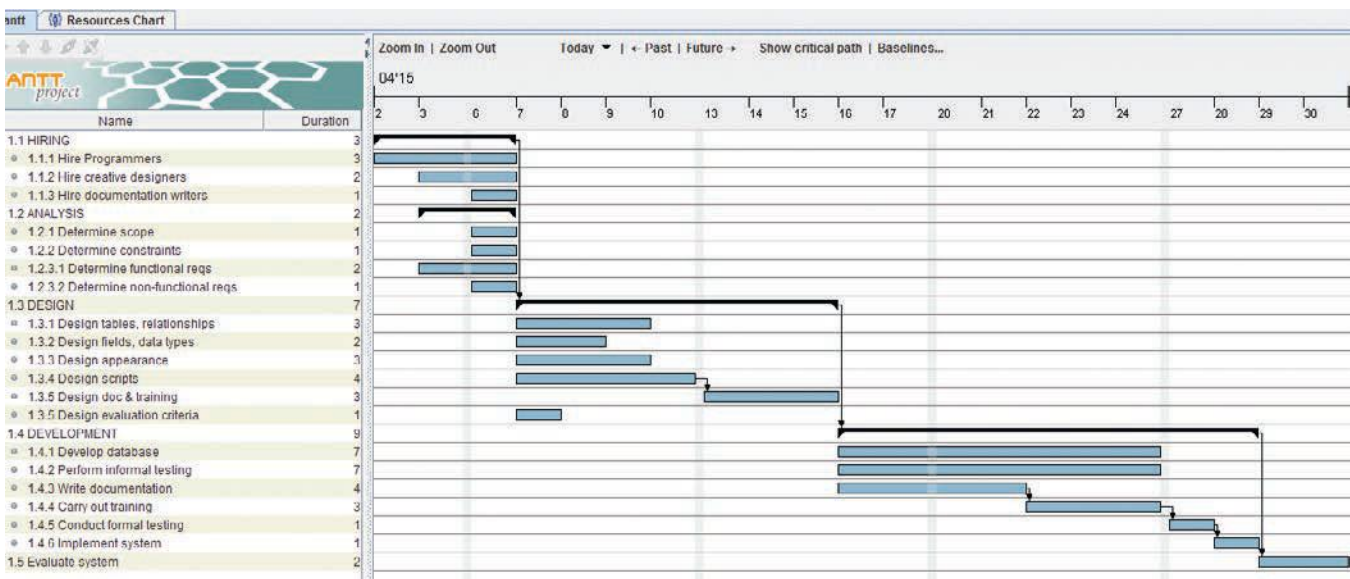


FIGURE 3.9 Task durations have been estimated

◀ Case study continued

The developers can finish with their Gantt chart by:

- selecting a project start date
- showing a critical path (critical tasks are shown cross-hatched)
- adding milestones to mark the end of major stages of the project.



NelsonNet additional resource: Figure 3.10 Gantt chart finished first draft, p. 143

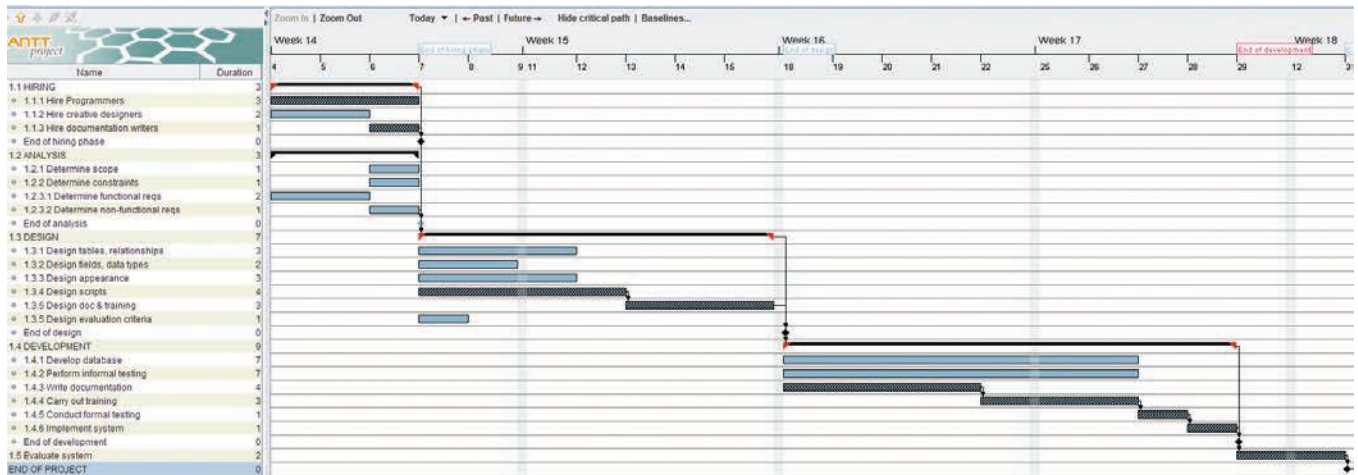


FIGURE 3.10 Finished first draft of the Gantt chart

Questions

- 1 The project team has made a mistake with the starting date of the 'Evaluation system' task. Explain why. How could they fix it?
- 2 How much slack does the 'Hire Creative designers' task have?
- 3 If designing scripts took a day longer than expected, would it affect the project end date?
- 4 The team discovers that the 'Develop the database' task is running overtime. How could they keep the project from running past its planned end date?
- 5 Explain the benefits of using Gantt chart software instead of using a pen and paper or a whiteboard.

Activity

- 6 Obtain a Gantt software tool and create a simple chart of your own.

Documentation using Gantt charts

To ensure project workers are not booked to be in two places at once, or idle, and that equipment is ready at the right time, the project manager allocates resources to tasks using the Gantt chart.

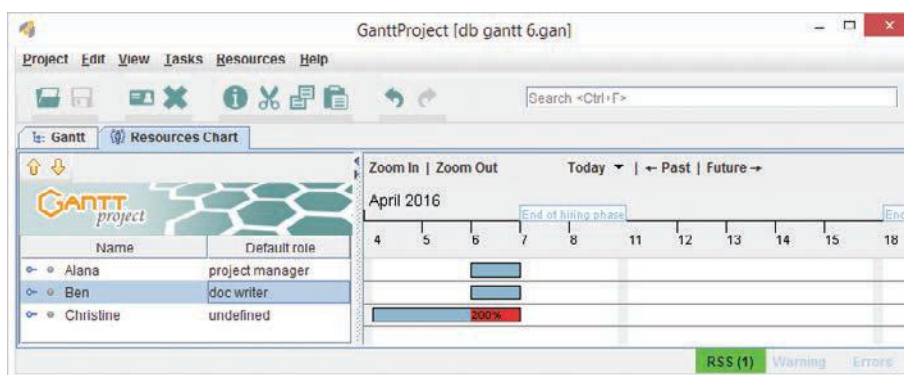


FIGURE 3.11 Adding resources to the project

Once the project is underway, the team will continue to refer to the Gantt chart to monitor their progress, and they will modify the chart when contingencies force plans to change.

While Gantt charts are one crucial aspect of project management, good file management practices are another. Wise file naming strategies are easy to learn and useful in many ways. You will find it easier to keep track of the materials you collect for your solution if you learn to manage files by naming them wisely.

File naming strategies

You will be creating and managing many files as you develop your SAT. Having a grasp of good file naming strategies will be vital to this. Although file naming strategies are immediately useful for your solutions now, you will also find them handy for years to come, particularly for future studies and work, because it is much easier to be organised when your files are organised too.

You should consider how you name your files and folders (or directories), as well as tables, fields, disks, programming storage structures, servers and computers carefully, because those names can endure for many years. The main goal is to use a consistent, sensible system.

You may already know that different operating systems have different file naming rules. Windows, Mac OS, Linux and Android all have different **formats** (which must be obeyed), and **conventions** (which *should* be obeyed).

Formats

File naming formats for Windows forbid certain punctuation marks and special characters, prevent the use of names reserved by the OS and ensure that the file name length cannot be exceeded (this may depend on the length of the file path, but is generally 255 characters). Similarly, Mac OS prevents the use of the colon and question mark in file names, as well as starting file and folder names with a dot. Mac OS X also specifically limits file and folder names to 255 characters. (Older Mac OS have shorter character limits.)

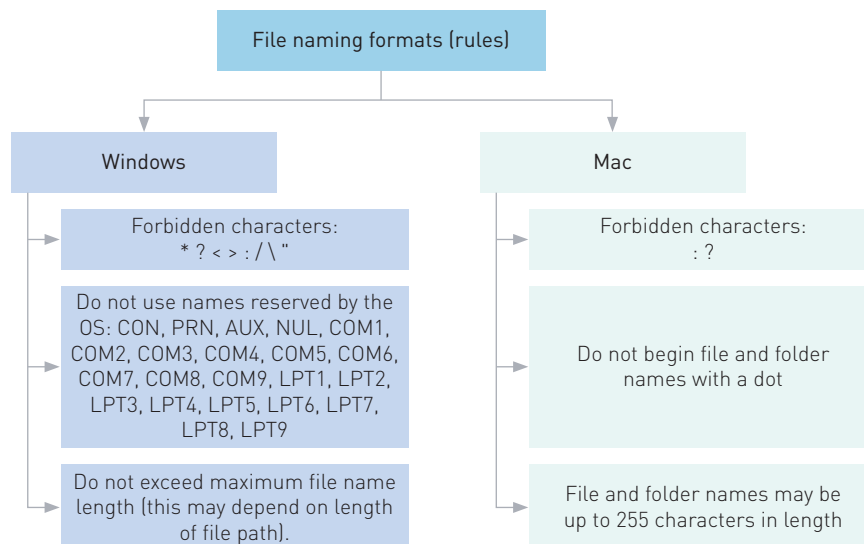


FIGURE 3.12 File naming formats (rules)

Conventions

Overall, remember the golden rule is that your file names should be **short**, **meaningful** and **consistent**. Apply the same careful principles to your folder names. Make sure that anyone who has access to your files and folders knows and applies the same conventions that you do. Other conventions to remember are listed in Table 3.4.

TABLE 3.4 File naming conventions

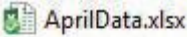
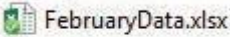
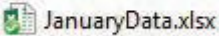
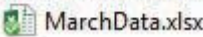
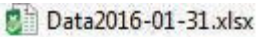
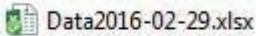
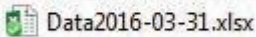
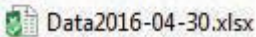
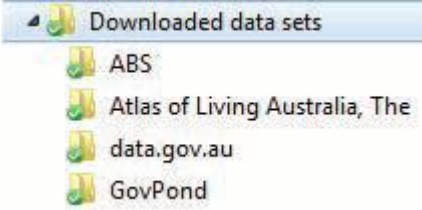
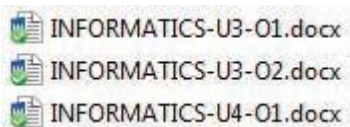
Convention	Do use ...	Why?
Characters	Numerals for numbers, not words	Words will not sort numerically Numerals will make the file names shorter
	Leading zeroes before numerals	Files will sort correctly in all OSs File names without leading zeroes in front of single-digit numbers may otherwise sort into a useless order (Figure 3.13)
	Plain, alpha numeric characters	Most punctuation marks and other special characters are forbidden by Windows, which makes them incompatible for cross-platform compatibility
	CamelCase or dashes instead of spaces or underscores in file names	If you ever intend to put your file on the internet, some programs will interpret a space as the end of the file name, leading to a 404 (file not found) error. Using underscores is not recommended because, linked on a webpage, the underscore character will be hidden by the URL.
	Short names	Although some OSs will allow you to name files up to 255 characters, this feature is not recommended because long names are hard to read and type accurately and will be cut off when viewed in a file manager. Also, some programs, such as DVD burners, use a file's full path (drive + folders + file name) and will fail if the total path is too long.
Language	Australian English	
Case	CamelCase (as above)	
	Capitalisation carefully; that is, case-sensitive links	While Windows and Mac file names are not case-sensitive, Linux file names <i>are</i> case sensitive. Most web servers run on Linux. You may find your links to a file you have uploaded simply do not work unless you make them case sensitive.
Dates	ISO format YYYY-MM-DD, with leading zeroes when necessary	Avoid ambiguity, such as this:     Force logical sort order, such as this:    
Versions	Version numbers	You can easily revert to an earlier saved copy if the current version becomes corrupted



FIGURE 3.13 File names with single-digit numbers may sort into a useless order without leading zeroes

TABLE 3.4 (continued)

Convention	Do use ...	Why?
Word choice and order	'Name, The' and 'Name, A(n)' structure for names that begin with articles (a, an and the)	Sorting files using their first significant word forces a logical sort order 
	Names that omit prepositions and conjunctions if needed	This keeps names short and prevents more important data from being obscured
	File names that start with category names and put more significant data before less significant data	Force logical sort order; for example: 
	'Family name, Given name' structure to refer to people	Force logical sort order Conforms with standards
	A consistent vocabulary: 'folder' and 'disk' or 'directory' and 'disc'	Avoid ambiguity

Data manipulation is discussed in Chapter 5.

Organising and storing data

Data is the lifeblood of an organisation. **Curating** data is a major responsibility, and managing data efficiently is vital to an organisation's effectiveness. Even a modest home computer soon accumulates thousands of documents, data files, videos, audio files, applications, backups and vital pictures of cats.

If you are disorganised, it will take longer to find the files you want to use. Eventually, mishaps will start to occur, such as losing one important file. Mishaps may then spiral into disasters, such as accidentally deleting directories of documents you have failed to back up.

Suffering catastrophic data loss or damage can doom a business. Customers would lose confidence, and the Australian Tax Office would have concerns about GST payments and business activity statements. Business itself would be impeded.

Similarly, if you suffered data loss it would seriously hinder your Outcomes, if not destroy your progress on them.

Organise your data with care, relying on some basic but effective strategies.

Hierarchies

You can organise data using hierarchical pyramid-like structures. Divide your major topics into subsections. (If need be, you can in turn subdivide those into more and more specific subsections until no further division is possible or necessary.)

Hierarchies make it easy to find a logical place to store and retrieve data in very large collections. Hierarchies are used in site maps, data structure charts and file systems (see Chapter 5). The structure of a hierarchy is rarely considered strictly right or wrong. Instead, it reflects the importance of its components to its creator and is either appropriate or

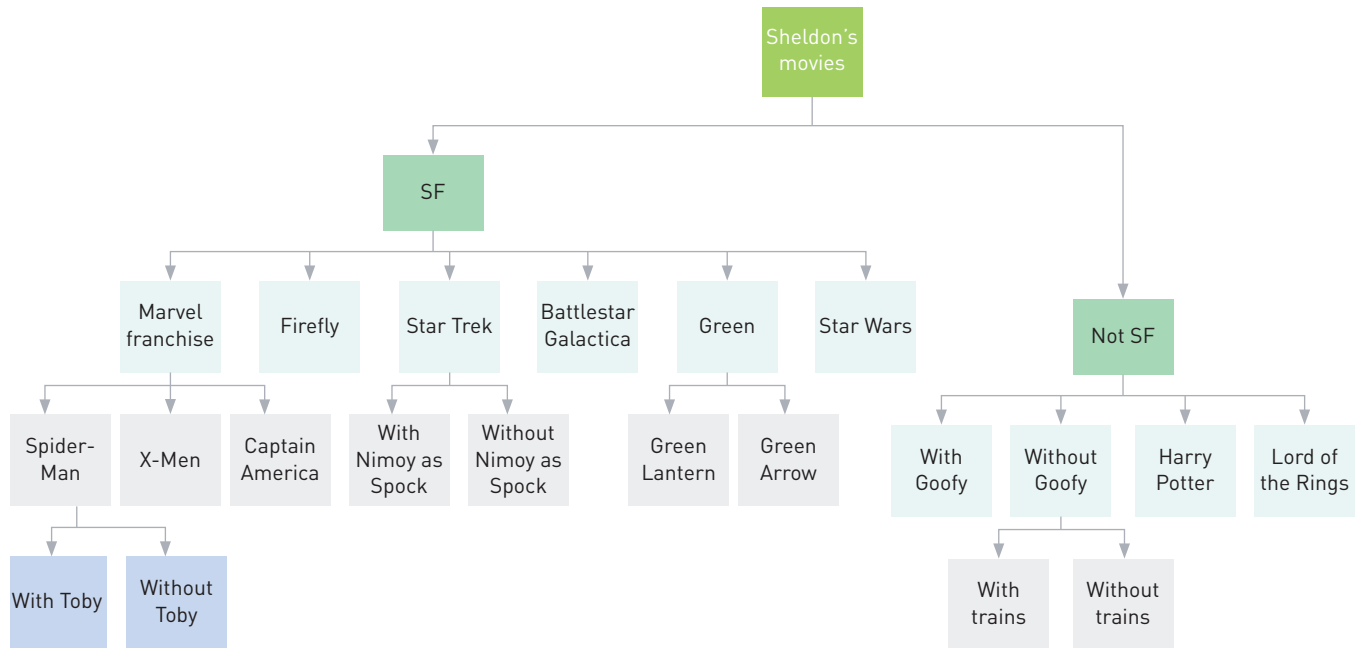


FIGURE 3.14 A hierarchical structure

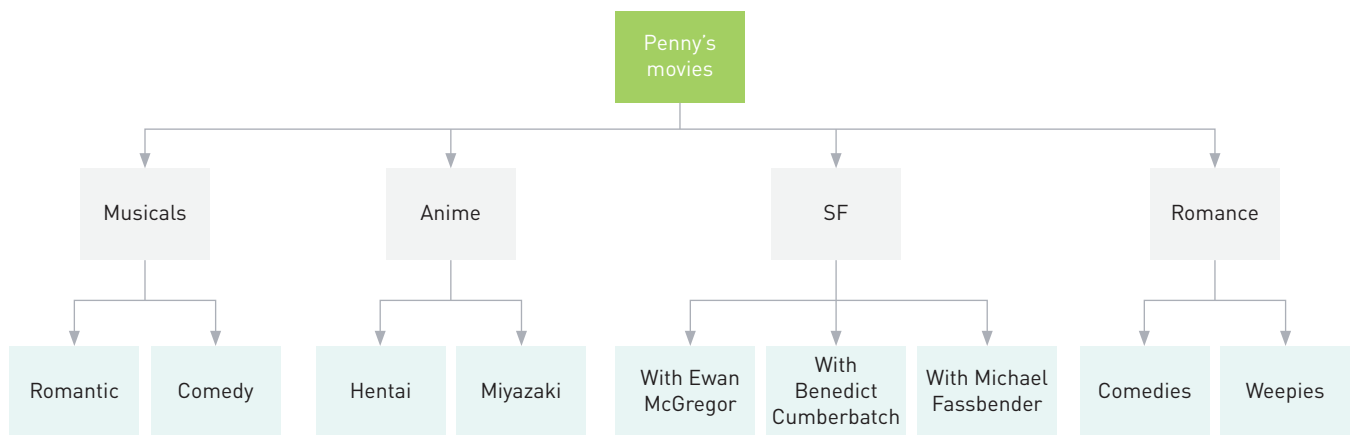


FIGURE 3.15 An alternative hierarchy

inappropriate. One person's hierarchy might be very different to another person's because of their priorities; compare Figure 3.14 with 3.15.

Choosing a particular hierarchical structure without forethought and consideration of priorities can lead to ongoing difficulty with data storage and retrieval. Often hierarchical data structures have a facility to let users collapse or expand the data view according to the level of detail they wish to see. For an overview of data, choose to view only top-level headings. Open lower levels to see more data and, eventually, raw data. For example, Microsoft Word's Outline View lets users choose what levels of headings to view, hide body text and restructure documents by simply dragging headings to different locations.

Databases

Databases are structured as hierarchies. A database is composed of tables, which contain records, which contain fields. A robust and powerful relational database will conform to the rules of **normalisation**, which are discussed in Chapter 1, to ensure that:

- data is not repeated, so data files are smaller and data retrieval is quicker
- changes to data need only be made in one place, saving time and preventing editing errors
- data can be retrieved and manipulated in more powerful ways.

A poorly organised database will be bloated, slow, unreliable and hard to query.

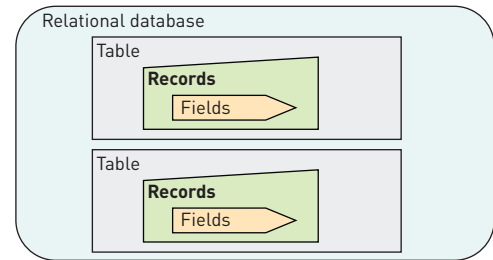


FIGURE 3.16 The hierarchical structure of a database

POORLY ORGANISED DATABASE

A poorly organised database has not achieved 1NF because it stores addresses as a single field, such as:

1 Fred Lane, Melbourne, VIC, 3000.

Explain the process for a database operator to:

- sort the records by suburb
- find all records of people in Bendigo
- print envelopes with street, suburb and postcode on separate lines.

The cloud and file sharing

The crucial question of *where* data should be stored is often overlooked. Contemporary work practices see more and more work being performed by teams that are spread across vast distances. Having to email copies of documents to team members, take in their changes to a single document and then re-distribute the documents is both tedious and likely to introduce errors. A strategy that is becoming increasingly popular is to keep a single master copy, or SPOT, of a document or database on the internet (or cloud), and allow access to multiple users.

This technique:

- removes the need to organise and track multiple versions of documents in different locations and at different stages of development
- permits multiple users to work collaboratively on a document at the same time in different locations and see each other's changes in real time
- permits users to 'roll back' a document to its previous state to undo unwanted changes
- gives users the confidence that the document they are using is the definitive version.

To improve access, store your data in a location where it can be accessed easily from many places: the cloud, a website or **FTP** server, a LAN server or a NAS unit. Storing data on a single computer's hard drive makes it only accessible to someone physically using that computer. Although turning file sharing on grants access to your local files to others on your LAN, it only works when the host computer is turned on.

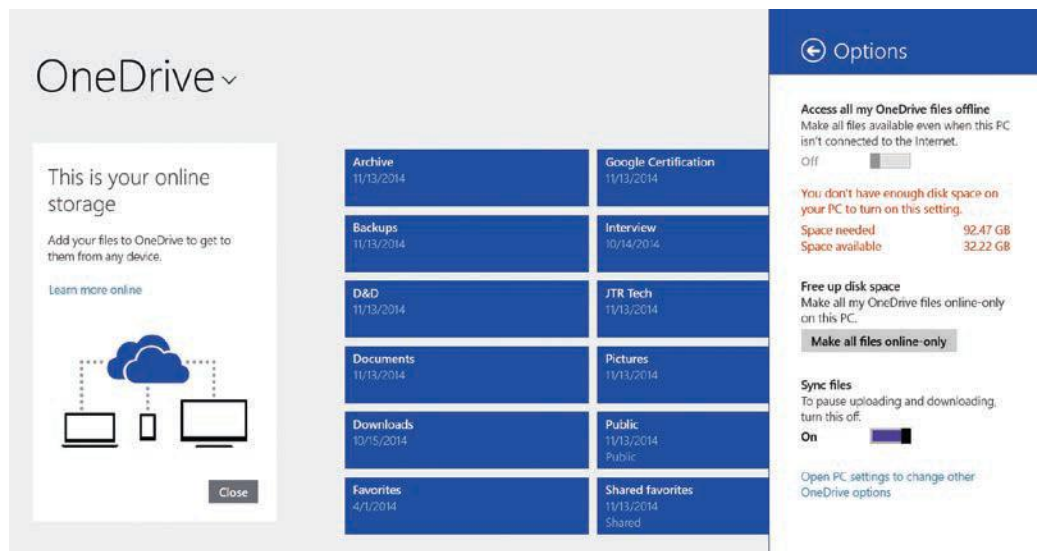


FIGURE 3.17 Many companies offer automated file sync to the cloud, including Microsoft, Amazon and Dropbox

Spreadsheets

Spreadsheets are far easier to use when they are well organised.

Appearance

Format and align data consistently; for example, ensure that all numbers in a specific column have the same number of decimal places and are right-aligned. This makes it far easier to see that 1000.0 is much higher than 1.00000 in a glance.

Use colour and text formatting consistently; for example, shade all formula cells grey; shade all cells that users may edit yellow; use conditional formatting to turn cells red if their value dips below zero; and set headings as bold. Use 'Merge and Centre' to put headings above multiple columns or rows to unify them.

Units

Remove units from numbers and put them in column headings, such as 'Weight (kg)', rather than in the same cell as the number. Use a single consistent unit in a column. Convert mixed units (such as 2 minutes 30 seconds) to a common denominator (such as 150 seconds).

You would not put two data items in a single cell if you expect to process either data item. For example, '2 minutes 30 seconds' or 'small \$1.00 large \$1.50'. Use two cells to store costs of 'large' and 'small' items.

Use one unit of currency only – do not put some prices in US dollars (USD) and others in Australian dollars (AUD).

Ease of use and error tolerance

You will find it easier to work with the spreadsheet if you freeze panes so that column headings do not scroll off-screen. You should label columns and rows clearly and concisely.

Use different worksheets for different purposes; for example, Sheet 1 for summary statistics, Sheet 2 for setup figures, Sheet 3 for lookup tables and Sheet 4 for charts, and rename worksheet tabs to make everything easy to find.



FIGURE 3.18 Renamed worksheet tabs

Also note that records conventionally flow across a row, not down columns.

You can apply names to important cells and ranges to make them easier to understand; for example:

```
=VLOOKUP (ID, StaffTable, 3)
```

is more readable than:

```
=VLOOKUP (C3, E8:J495, 3) .
```

Names also make it far easier to refer to cells on other worksheets; for example:

```
=TAXRATE*TOTALCOST
```

is far more meaningful than:

```
=Sheet2!E7*Sheet3!D9
```

You can also use cell protection to prevent users accidentally changing or deleting formulas.

A spreadsheet's VLOOKUP function is actually just a simple form of a related table in a relational file database.

THINK ABOUT INFORMATICS 3.4

To what design principles do 'Appearance', 'Ease of use' and 'Error tolerance' relate?

TABLE 3.5 Organising spreadsheet rows and columns: Conventional

Family name	Given name	Date of birth
Kirby	Anika	2014-10-03
McLeod	Stuart	2015-10-07
Mensur	Abdul	2015-02-19
Singh	Jaspreet	2014-06-18
Wu	Evelyn	2014-04-18
Wu	Vivian	2014-04-18

TABLE 3.6 Organising spreadsheet rows and columns: Unconventional

Family name	Kirby	McLeod	Mensur	Singh	Wu	Wu
Given name	Anika	Stuart	Abdul	Jaspreet	Evelyn	Vivian
Date of birth	2014-10-03	2015-10-07	2015-02-19	2014-06-18	2014-04-18	2014-04-18

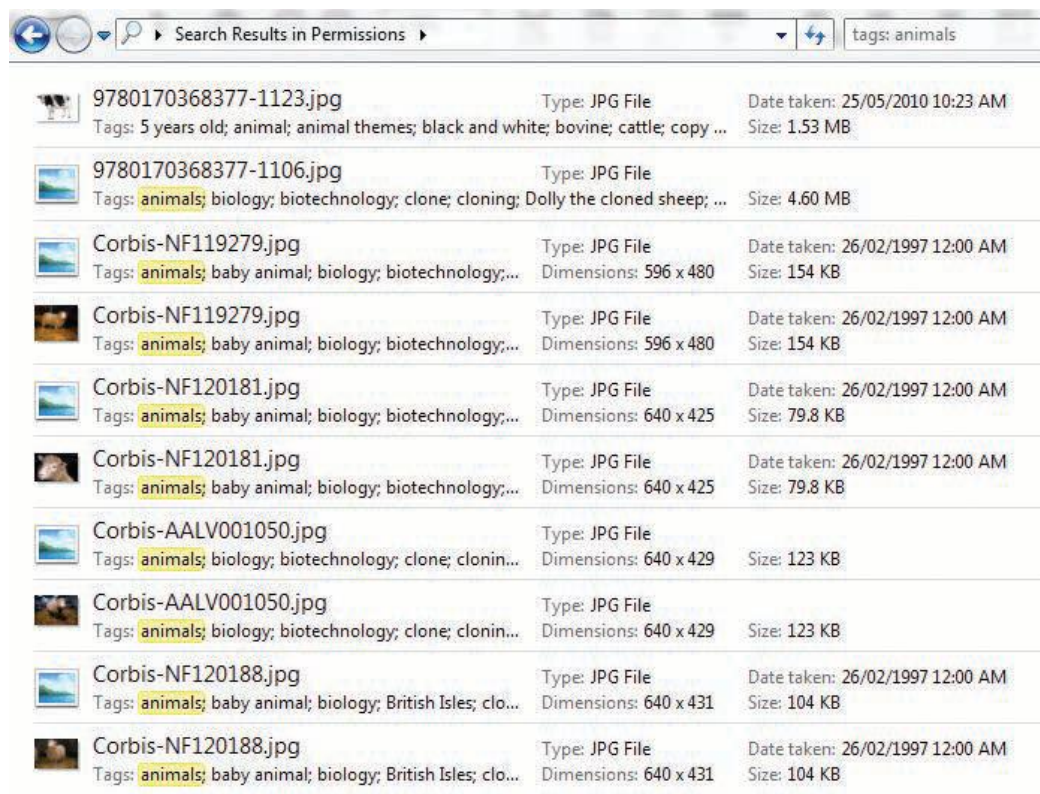
Metadata

To search files by their attributes rather than by their file names, use metadata. Metadata (literally data about data) is descriptive data saved inside a digital data file that can be read by media players and editors. Metadata about an MP3 includes track number, artist, cover art, track length, album, the year of recording and so on. Cameras save metadata such as exposure settings and GPS coordinates (location data) inside digital images.

You can edit metadata from within application software. Adobe Photoshop allows you to edit metadata of photos so you can catalogue and search images more efficiently and avoid having to add the descriptive detail in the photo's file name; for example, `cat-black-chair-asleep-stretched-onback.tif`.



Dublin Core is a global project intended to describe web resources and physical resources using metadata.



The screenshot shows a Windows Explorer window with the address bar set to 'Search Results in Permissions' and a search filter of 'tags: animals'. The results list several JPEG files with their metadata.

File Name	Type	Date taken	Size
9780170368377-1123.jpg	JPG File	25/05/2010 10:23 AM	1.53 MB
9780170368377-1106.jpg	JPG File		4.60 MB
Corbis-NF119279.jpg	JPG File	26/02/1997 12:00 AM	154 KB
Corbis-NF119279.jpg	JPG File	26/02/1997 12:00 AM	154 KB
Corbis-NF120181.jpg	JPG File	26/02/1997 12:00 AM	79.8 KB
Corbis-NF120181.jpg	JPG File	26/02/1997 12:00 AM	79.8 KB
Corbis-AALV001050.jpg	JPG File		123 KB
Corbis-AALV001050.jpg	JPG File		123 KB
Corbis-NF120188.jpg	JPG File	26/02/1997 12:00 AM	104 KB
Corbis-NF120188.jpg	JPG File	26/02/1997 12:00 AM	104 KB

FIGURE 3.19 Using Windows Explorer to search files using metadata rather than file names

Archiving

Old, unused, redundant and temporary files take up storage space and distract you from finding the files you need. Deleting seemingly unnecessary files can be unwise because you may be unsure if you will need them again in the future. Instead of keeping the files or deleting them outright, you can take them offline by moving or copying them to external storage, compressing them to save space and then deleting the original file.

Files stored on offline storage media, such as DVDs, USB flash drives and external HDDs, are inaccessible until attached to a computer and loaded. Files stored on online storage, such as internal HDDs, networks and the cloud, can be loaded immediately.

Manually compress files in utilities such as WinZip or WinRAR, or automatically using backup software or your OS. Zipping files to compress them not only reduces each file's size, but also can group many files into a single, easy-to-manage package. File formats that are already compressed, such as JPG, MP3 and AVI, reduce very little in size when zipped.

Email

Your email inbox can hold valuable communications and stored information. However, if you do not keep your email well organised, over time it can become a mess of redundant old information. Create folders, or use labels, highlighting or tags to organise your inbox. You can use search facilities to find emails by key words, author, subject and other fields. Regularly delete junk and use spam filters to divert unwanted emails and keep your inbox tidy.

Many web browsers allow settings, bookmarks, emails, passwords and apps to be synced across multiple devices so that they all can share the same data and settings.

Refer to Chapter 2 for information on integrity of data. For Unit 3, Outcome 2, integrity of data comprises:

- timeliness
- authenticity
- relevance
- accuracy.

Synchronising files

Keeping identical copies of certain files on multiple devices, such as your phone, tablet and home PC, can be desirable, but it is difficult to keep them all current when you can change a file on any one of those devices without affecting another. Utilities designed to synchronise files compare the contents of folders, determine which files are the same or different, newer or older, and copy files back and forth to bring both folders up to date, preserving the **timeliness** and thus integrity of the data.

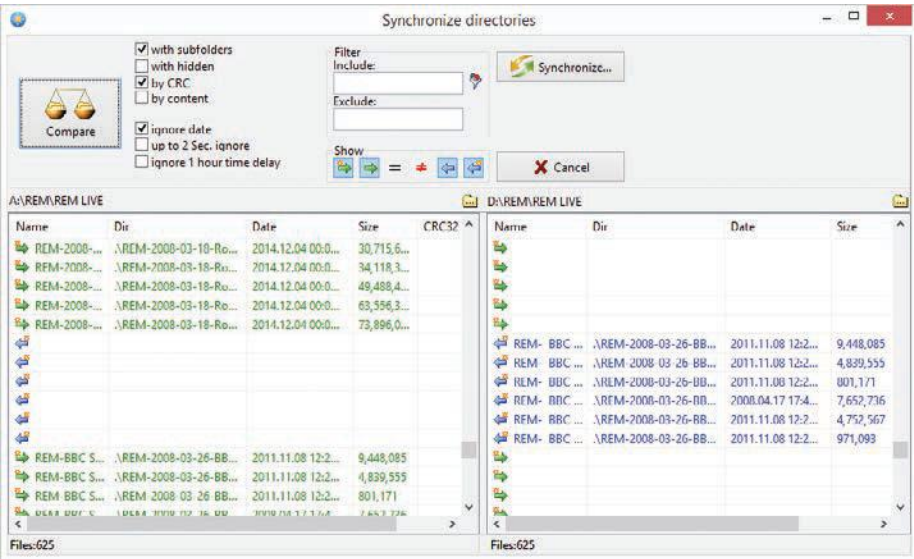


FIGURE 3.20 A file sync utility keeps two data copies updated

Identifying patterns and relationships between data

Use statistics to convert large quantities of raw data into small, informative, meaningful summaries. You are not required to perform very complex statistics when examining your hypothesis for Unit 3, Outcome 2. However, you *should* develop a basic understanding of statistical concepts, because these are essential for manipulating data, and are very valuable for reaching sensible conclusions. The most useful statistical concepts are described in Table 3.7.

TABLE 3.7 Useful statistical concepts

Concept	Purpose
Average	The single value that gives the most representative summary value of a range of numbers; mean, mode and median are different ways of averaging data under different circumstances.
Significance	Indicates whether two data sets are basically similar, or are different in an important way
Standard deviation	Indicates how consistent the values are in a data set. Are the values all close to the average, or do they vary greatly? A low standard deviation means that the mean is truly representative of the data and can be trusted. A high standard deviation suggests you should not trust the mean.
Correlation	Do two data sets show the same trends, as if they might be connected in some way?
Causation	One factor (for example, eating patterns) is responsible for causing changes in another factor (for example, weight).

Basic statistics

Averages

Hiroyuki, a sushi restaurant manager, wants to monitor the performance of his two chefs, Bruno and Eva. He puts each chef on duty for one week, and counts the numbers of diners each day.

Chef	Diners per day						
Bruno	12	16	8	15	22	35	7
Eva	4	20	5	19	7	43	3

Initially, Hiroyuki cannot draw a conclusion from the raw figures. However, he realises that averaging the number of diners would be informative. There are three methods of calculating averages: mean, median and mode.

The mean

Most often, when you hear of the 'average', you tend to think automatically of the arithmetic mean. The mean is the sum of the data divided by the total number of values.

Bruno's mean:

$$12 + 16 + 8 + 15 + 22 + 35 + 7 = 115$$

$$\frac{115}{7} = 16.43 \text{ daily diners (rounded up)}$$

Eva's mean:

$$4 + 20 + 5 + 19 + 7 + 43 + 3 = 101$$

$$\frac{101}{7} = 14.43 \text{ daily diners (rounded up)}$$

The median

When data is sorted in order from smallest number to largest number, the median is the number in the middle of the range. Both Bruno and Eva have a data set that can be sorted in this manner.

Bruno's sorted data:

7, 8, 12, 15, 16, 22, 35

With seven numbers in Bruno's data set, the number in the middle of the range is the fourth – 15. While you can work out medians manually, Excel's `MEDIAN()` function can determine the median in a range of values more efficiently.

Eva's sorted data:

3, 4, 5, 7, 19, 20, 43

Eva also has seven numbers in her data set, which means that the number in the middle of her range is also the fourth – in this case, 7.

You should note that the mean and the median can be very different, so they can lead you to have drastically different interpretations of the average.

The mode

The mode is the value that occurs most frequently. Sometimes, the mode can be the most logical and representative average to use. For example, imagine that you interview four people at random and ask them how many children they have. The answers they give you are 2, 2, 2 and 12. Logically, what would be the most representative average number of children? The mode, which is 2. The **outlier**, 12, would be ignored. (In this instance, the median would be logically representative as well, but the mean, 4.5, would be unrepresentative because it would be inaccurate for all of the surveyed people.)

Excel's `AVERAGE()` function calculates a mean.

An outlier may be caused by experimental error, or it may be a remarkable exception to the general population. Although it may be worthy of investigation, it is still not typical. Excel's `TRIMMEAN()` function will ignore outliers when calculating a mean, but the user needs to indicate when to begin ignoring values.

THINK ABOUT INFORMATICS 3.5

Use a spreadsheet to enter the sample data shown and use the `AVERAGE()`, `MODE()` and `MEDIAN()` functions to confirm the averages.

Published hypothesis conclusions tend to include statements such as ' $p < 0.05$ '. This indicates a significance measure. It means that the probability of the result being purely caused by chance is less than 5%. The lower the number is, the more confident you can be that the hypothesis is supported.

The lowercase Greek letter sigma, σ , is used to represent standard deviation. Lowercase mu, μ , is used to represent mean.

Neither Bruno nor Eva have a mode because they do not have a repeating number.

You will find that each type of average is logical and accurate under different circumstances, but they often will not agree with each other. Choosing the right type of average for the right circumstance is important. It is easy to misrepresent data by choosing the wrong type of average.

Significance

Bruno's mean number of customers is 16.43 and Eva's is 14.43. Should Hiroyuki be concerned about this difference? Is it insignificant, or merely a matter of chance? A difference between data sets is not always significant. A **significance test** can help to indicate when a difference may be important; for example, you can conduct a t-test using a spreadsheet.

Standard deviation

You will often need to know how much variety there is in your data. For example, if diner counts in Week 1 are 8, 7, 9, 8, 7, 9 and 8 (indicating a mean of 8), and data tends to be consistent, differing little from the mean, a sudden value of 33 would be very unusual. Conversely, if the figures for Week 2 are 1, 33, 2, 1, 0, 3 and 16, the mean is still 8, but the data is inconsistent and none of the individual figures are near the value of the mean.

Using Excel, you can calculate the standard deviation of Week 1's figures to 0.816. This means that Week 1's figures have little variation, and its mean accurately reflects the individual values of the entire data set.

On the other hand, Week 2's standard deviation is 12.329. Week 2's data varies greatly. The data is inconsistent and lacks any discernible pattern. The mean cannot be trusted to reflect the general nature of the entire data set accurately.

	A	B	C	D	E	F	G	H	I	J
Week 1		8	7	9	8	7	9	8	0.816	
Week 2		1	33	2	1	0	3	16	12.329	

FIGURE 3.21 Standard deviation in Microsoft Excel using `=STDEV.S`. Note that Excel has multiple standard deviation formulas. `=STDEV.S` calculates based on a sample.

You can also use mean and standard deviation to see whether any value in a given data set could be considered unusual or special. The following rules are applied.

- 1 A value more than 1 standard deviation from the mean (higher or lower) is unusual and noteworthy.
- 2 A value more than 2 standard deviations from the mean is very unusual.
- 3 A value more than 3 standard deviations from the mean is a major anomaly.

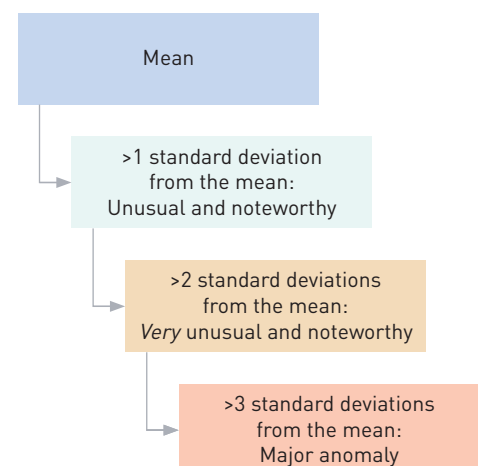


FIGURE 3.22 Standard deviations from the mean

HEIGHTS

The heights of thousands of 18-year-old Australian males are measured. The mean height is 175 cm. The standard deviation is calculated to be 15 cm. In plain terms, this means that the average 18-year-old Australian male is 175 cm – give or take 15 cm.

The classic normal distribution curve looks like Figure 3.23. Most of the data *is* close to the mean, with progressively less varying greatly from the mean. The height of the curve represents the number of men with each height, so the highest point on the curve occurs at the mean value – 175 cm.

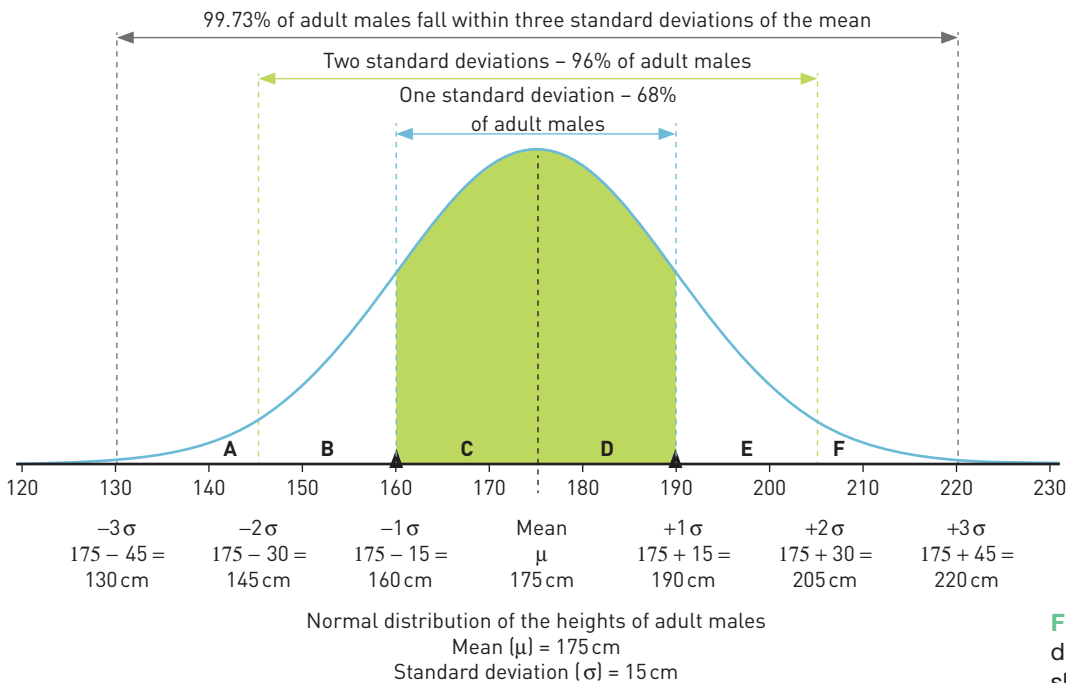


FIGURE 3.23 The normal distribution of heights, shown in a bell curve shape

Heights that are between 160 cm and 190 cm are not considered noteworthy. This is because the mean is 175 cm and one standard deviation is 15 cm. To demonstrate:

The mean (175 cm) + 1 standard deviation (15 cm) =
190 cm, therefore 190 cm is not noteworthy.

Similarly:

The mean (175 cm) – 1 standard deviation (15 cm) =
160 cm, therefore 160 cm is not noteworthy.

Heights that are not considered noteworthy appear in the green shaded area of Figure 3.23 (C to D) – 68% of heights fall into this category.

In Figure 3.23, Area B is 145 cm to 160 cm and Area E is 190 cm to 205 cm. They are both considered noteworthy because they are between 1 to 2 standard deviations from the mean (more than 15 cm and up to 30 cm). A total of 13.6% of heights fall into this category.

Note that 96% of the entire population tends to fall between –2 and +2 standard deviations from the mean.

Area A, 130 cm to 145 cm and Area F, 205 cm to 220 cm, denote the unusual 3.4% of the population who are greater than –2 and +2 standard deviations (2σ) from the mean.

Only 0.2% of the population falls beyond –3 and +3 standard deviations (3σ) from the mean. They are considered exceptionally rare. In this instance, they would be up to 130 cm tall, or greater than 220 cm tall.

A bell curve shape such as that shown in Figure 3.23 could be replicated using software such as GeoGebra, with labels added in Microsoft PowerPoint or Adobe Acrobat.

Correlation and causality

The human brain is hard-wired to look for patterns in information. This tendency can be useful for detecting predators, but it can also mislead us into thinking we see faces everywhere: on Mars, the Moon, or in burnt toast. Similarly, we look for patterns in data. At times, we make mistakes there too.



FIGURE 3.24 Humans see faces everywhere because our brains are wired to seek them out; whether on **a** the surface of Mars, **b** bats or **c** power sockets.

A common mistake in analysing data occurs when you notice a correlation and assume that it is the result of a **causal relationship**. A **correlation** is a relationship between two variables. You will most likely notice it as a similar trend or trends in two data sets. The error here would be if you assumed that one factor **caused** the other. This is often not the case, and you can only tell if you conduct further research.

Sometimes it is very obvious that two trends are not related. For instance, if sales of pepperoni pizzas varied over a 10-year period in Victoria, you would probably not think it was related to varying sales of umbrellas, even if they varied in the same way at the same time. These two factors would not even be in the same data set, so you would logically conclude it was a coincidence.

However, you may consider conducting research to look for a causal relationship between varying sales of umbrellas and rainfall, or perhaps even climate change.

CAUSE AND EFFECT

If you were using one of the paleo diet hypotheses from Chapter 2, you would be trying to establish a cause and effect relationship.

Do increasing numbers of people who live in Richmond eat a paleo diet because they believe it has health benefits?

- 1 Between which two factors would you be trying to establish a cause and effect relationship?
- 2 Are there any potential additional factors to consider?
- 3 Discuss the merits of causality in the following hypothesis:

Being wealthy leads to better health because you can afford expensive medications.

You can calculate correlation between data sets using Excel's CORREL function, but a correlation value will not indicate which factor (if either) is the cause and which is the effect. Here are some examples of potentially dangerous cause and effect reasoning.

Observation: A school that introduced sex education classes later reported increased incidences of students diagnosed with sexually transmitted infections (STIs).

Reasoning: Sex education was assumed to have made the students more sexually active.

Observation: President George W. Bush came into power and the multi-nation 'War on Terror' commenced shortly after.

Reasoning: President George W. Bush caused the war to start.

Data visualisations

If you have difficulty interpreting your numerical data, you can use a data visualisation. Using (analogue) lines, shapes and colours to represent (digital) numbers can make it easier to interpret the data. Appropriate visualisation forms include graphs, charts, spatial relationships, maps, histograms and network diagrams. Each can be used to make trends and patterns more obvious.

You can represent different types of **analogue data** using dynamically changing physical indicators, such as strength, direction, duration or size of the signal. For example, a sound's pitch and volume, the position of a clock's hands, the height of liquid in a thermometer, or the length and angle of a line in a chart. You represent sizes, shapes and direction using sizes, shapes and directions.

Encode **digital data** numerically. Infinitely detailed analogue sizes, shapes and directions are converted to numeric samples.

The most important thing in any data visualisation is that it must be easy to see the main aspect and trends in the data.

Visualisation tools

You can manage large and challenging data sets more easily with various visualisation tools, specialist software and services, such as the following.

- IBM's Watson Analytics (previously ManyEyes) lets you create analogue data visualisations from your digital data sets. You will need to sign up to use the service, and free functionality is limited.
- Tableau (Figure 3.25) offers free trials of their commercial data visualisation and statistics software.



Watson Analytics



Tableau

Worldwide Oil Rigs

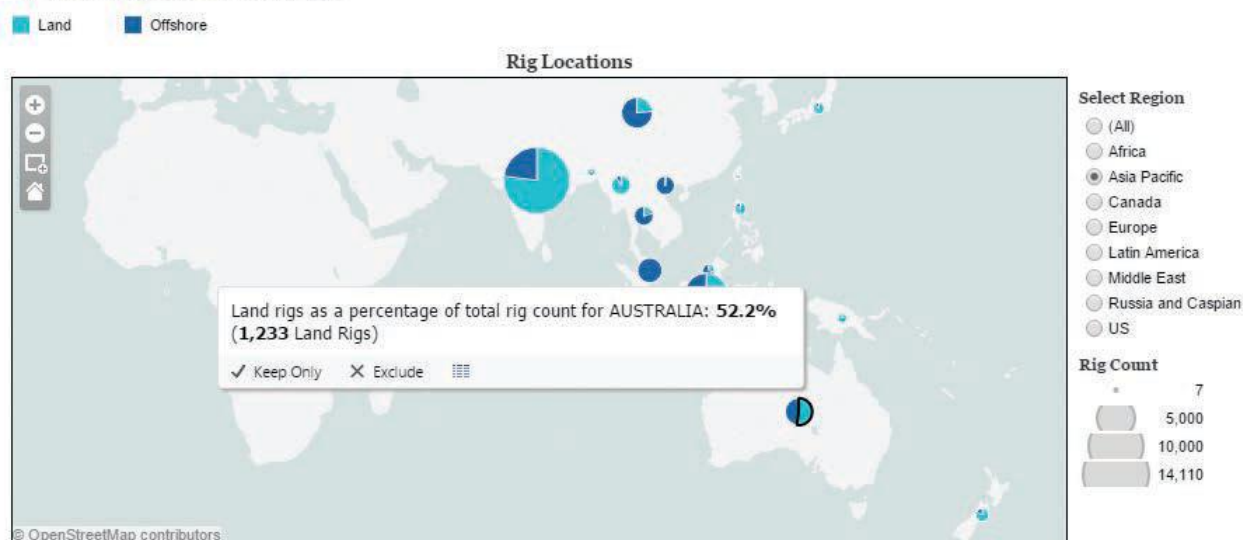


Tableau Software, Worldwide Oil Rig Numbers, <http://www.tableau.com/learn/gallery>

FIGURE 3.25 A Tableau visualisation of worldwide oil rig numbers

Most versions of Microsoft Word have the same range of charts as Microsoft Excel. Click the Insert tab, and then click Chart in the Illustrations group.

Spreadsheet charts

You can use your spreadsheet software such as Microsoft Excel to produce various useful types of charts. Choose an appropriate chart type and follow its conventions.

A **pie chart** is divided into coloured or patterned 'slices' proportional to the percentage of the whole pie. Consider choosing a pie graph if you need to depict approximate proportional relationships (relative amounts) or compare part of a whole at a given point in time. The full circle represents 100 per cent. The angle of each 'slice' is found by multiplying its percentage value by 360° . You should label the slices or use a legend.

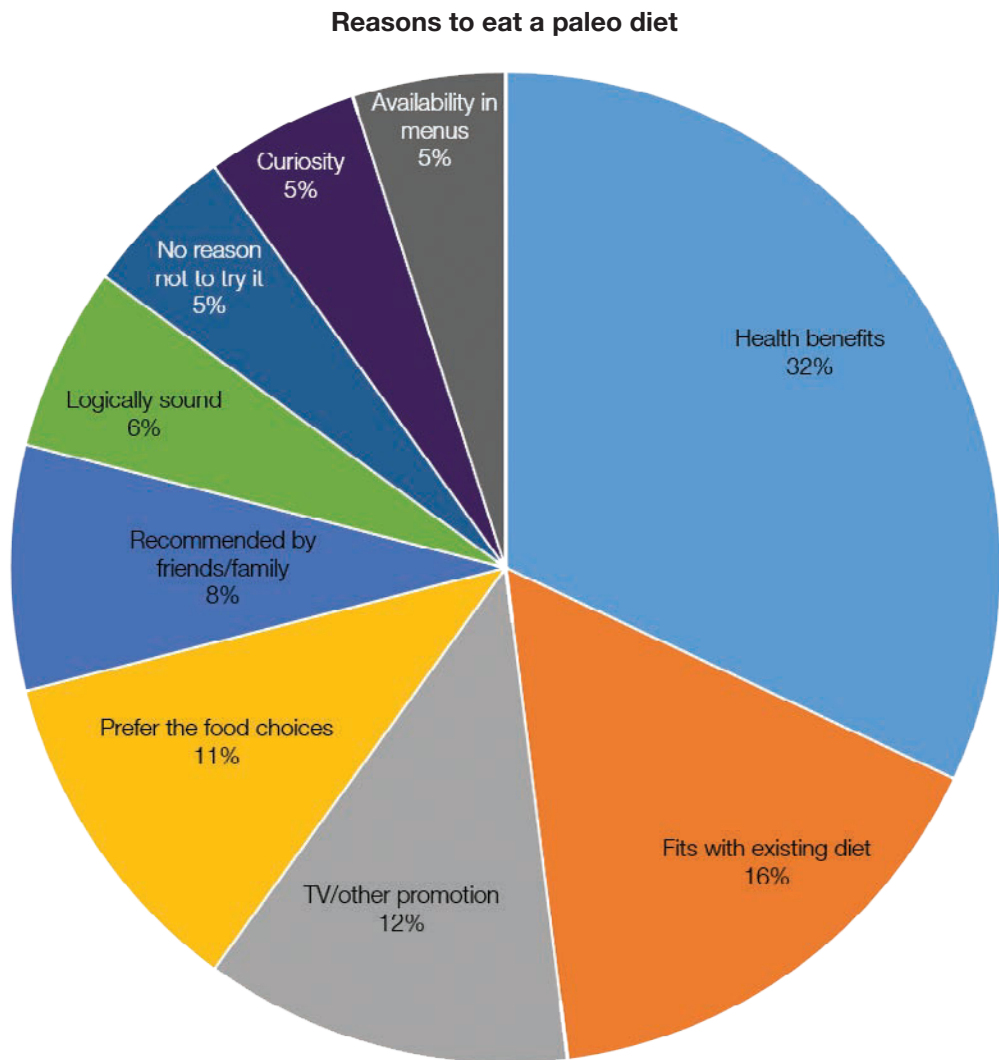


FIGURE 3.26 An Excel-generated pie chart from the paleo hypothesis: 200 people were asked why they first started eating a paleo diet. The raw figures are converted to percentages in the pie chart.

Column graphs arrange data vertically and **bar graphs** arrange data horizontally (in rows). Both are handy for presenting data changes over a period of time or for showing comparisons across different times. They enable visual comparisons very easily so that differences are recognised quickly. You can use colours to distinguish between bars if representing different sets of data, but remember that the colour may be lost if printing.

Line graphs display continuous data over time, set against a common time. They are useful for showing trends in data at equal intervals. If more than one line is shown, use a different colour to distinguish each line.

Scatter graphs show raw data points as dots. This is useful because data rarely lines up precisely, follows the mean exactly, or shows a perfect trend. You can add a **line of best fit** to your scatter graph (Figure 3.27), which will indicate the single trend that most accurately shows the general direction of the trend.

Spreadsheets can calculate lines of best fit for scattered data.

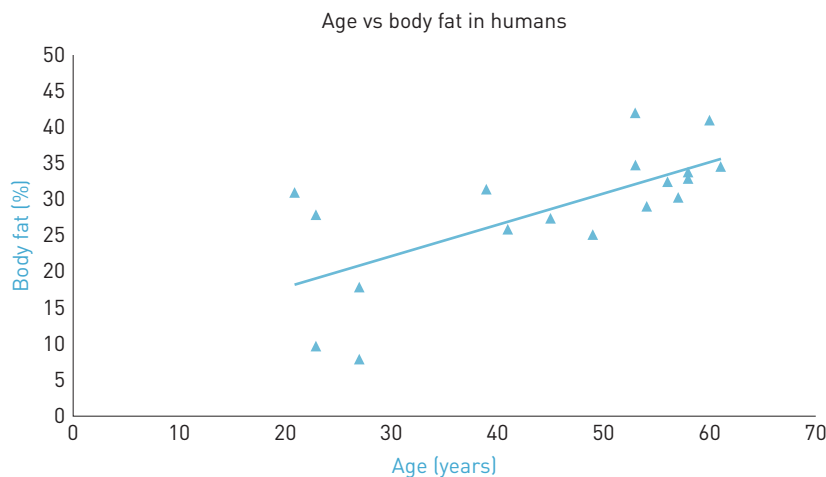


FIGURE 3.27 A scatter graph with a line of best fit: Age vs. body fat in humans. The individual data points are hard to interpret. However, you can see the line of best fit markedly inclining upwards to the right, which indicates that body fat increases with age.

Interpretation of the line of best fit in Figure 3.27 indicates that body fat increases with age. A line of best fit also enables two other important statistical functions: interpolation and extrapolation.

Interpolation involves calculating missing data using trends within the known data. There is no data on body fat for age 35 in Figure 3.27, but using the line of best fit, you can calculate that it would be approximately 25%.

Extrapolation involves calculating missing data that is outside the limits of the known data. To find body fat levels for 10 year olds or 70 year olds, you could extend the line of best fit beyond the known data set and then read the anticipated body fat levels.

Note that interpolation and extrapolation will only be accurate if the line of best fit follows a predictable path. They are both actually predictive hypotheses.

Microsoft has been introducing sophisticated business intelligence tools into recent versions of MS Office, such as Excel, which offer exciting data visualisation possibilities. You might want to particularly investigate Powermap for visualising geographical data.

Queries and searches

It is normal to be overwhelmed by bulk numbers in data sets. Most tools and techniques for handling bulk data involve hiding most of the data and highlighting the interesting parts that remain. You can usually find the interesting data with **queries** and **searches**. Queries and searches will help you to locate the patterns in your data sets so that you can find the larger story that your Outcome data is telling you.

Pivot tables

One of the least-known tools of data manipulation is the pivot table. Pivot tables are a feature of spreadsheets that are highly flexible and useful. You can use a pivot table to reorganise your data set and reveal significant patterns without changing the underlying structure of your data.

If you have been conducting research on the availability of paleo-friendly food in Victoria for your hypothesis and enter it into a spreadsheet, it might look something like this:

Local council regions	Geographic location	Paleo-friendly eatery	Paleo-friendly shops	Fresh paleo bread
Ballarat	North Central Victoria	0	1	1
Banyule	North East	0	1	2
Bayside	South	3	2	2
Boroondara	East	5	7	6
Brimbank	West	0	0	0
Casey	Outer South East	0	2	0
Darebin	Inner North	1	1	1
Frankston	Outer South	1	1	0
Glen Eira	South East	1	3	5
Greater Dandenong	Outer South East	0	0	0
Greater Geelong	Greater Geelong	0	1	0
Hobsons Bay	South West	0	1	1
Hume	Outer North West	1	0	0
Kingston	South East	1	4	2
Knox	Outer East	0	1	1
Latrobe	Gippsland	0	0	1
Manningham	Outer East	2	1	0
Maribyrnong	Inner West	1	1	0
Maroondah	Outer East	0	2	0
Melbourne	Central	7	2	2
Melton	Outer South West	0	0	0
Monash	South East	3	1	1
Moonee Valley	North West	0	1	1
Moreland	Inner North	1	4	0
Mornington	Mornington Peninsula	1	3	1
Nillumbik	Outer North East	0	1	0
Port Phillip	South	8	3	2
Stonnington	Inner South East	9	7	9
Whitehorse	Outer East	0	3	0
Whittlesea	Outer North	0	0	0
Wyndham	Outer South West	0	0	1
Yarra	Inner North East	9	4	2
Yarra Ranges	Outer East	2	0	0

FIGURE 3.28 The locations of paleo-friendly cafes and restaurants and shops, and paleo bakeries, sorted by local council area in Victoria

By going to Insert and then clicking Pivot Table, you can insert a flexible pivot table that allows you to see your data summarised in a variety of ways.

	A	B	C	D
1	Row Labels	Sum of Paleo-friendly eatery	Sum of Paleo-friendly shops	Sum of Fresh paleo bread
2	Central	7	2	2
3	Melbourne	7	2	2
4	East	5	7	6
5	Boroondara	5	7	6
6	Gippsland	0	0	1
7	Latrobe	0	0	1
8	Greater Geelong	0	1	0
9	Greater Geelong	0	1	0
10	Inner North	2	5	1
11	Darebin	1	1	1
12	Moreland	1	4	0
13	Inner North East	9	4	2
14	Yarra	9	4	2
15	Inner South East	9	7	9
16	Stonnington	9	7	9
17	Inner West	1	1	0
18	Maribyrnong	1	1	0
19	Mornington Peninsula	1	3	1
20	Mornington	1	3	1
21	North Central Victoria	0	1	1
22	Ballarat	0	1	1
23	North East	0	1	2
24	Banyule	0	1	2
25	North West	0	1	1
26	Moonee Valley	0	1	1
27	Outer East	4	7	1
28	Knox	0	1	1
29	Manningham	2	1	0
30	Maroondah	0	2	0
31	Whitehorse	0	3	0
32	Yarra Ranges	2	0	0
33	Outer North	0	0	0
34	Whittlesea	0	0	0
35	Outer North East	0	1	0
36	Nillumbik	0	1	0
37	Outer North West	1	0	0
38	Hume	1	0	0
39	Outer South	1	1	0
40	Frankston	1	1	0
41	Outer South East	0	2	0
42	Casey	0	2	0
43	Greater Dandenong	0	0	0
44	Outer South West	0	0	1
45	Melton	0	0	0
46	Wyndham	0	0	1
47	South	11	5	4
48	Bayside	3	2	2
49	Port Phillip	8	3	2
50	South East	5	8	8
51	Glen Eira	1	3	5
52	Kingston	1	4	2
53	Monash	3	1	1
54	South West	0	1	1
55	Hobsons Bay	0	1	1
56	West	0	0	0
57	Brimbank	0	0	0
58	Grand Total	56	58	41



NelsonNet additional resource: Paleo diet pivot table data

FIGURE 3.29 This particular version of the pivot table groups the data by geographic location, allowing you to visualise the hotspots for paleo-friendly food.

1		Column Labels																													
		Inner										North				Outer				Outer				Outer				Grand			
		Central	East	Gippsland	Greater Geelong	Inner North	Inner North	Inner South	Inner West	Mornington Peninsula	Central Victoria	North East	North West	North East	Outer East	Outer North	Outer North	Outer South	Outer South	Outer West	South East	South West	South East	South West	South East	South West	South East	South West	Grand Total		
2	Values																														
3	Sum of Paleo-friendly eatery	7	5	0	0	2	9	9	1	1	0	0	0	4	0	0	1	0	0	11	5	0	0	0	56						
4	Sum of Paleo-friendly shops	2	7	0	1	5	4	7	1	3	1	1	1	7	0	1	0	1	2	0	5	8	1	0	58						
5	Sum of fresh paleo bread	2	6	1	0	1	2	9	0	1	1	2	1	1	0	0	0	0	1	4	8	1	0	41							

FIGURE 3.30 This version of the pivot table may look very different to Figure 3.29, but it is ultimately telling you very similar information, minus the breakdown by local council, and in a horizontal format

It may interest students to know that the figures provided for paleo eateries in Victoria were accurate at the time of writing. If you live by the paleo diet and like to eat out, the best place in Victoria to live is Prahran, followed by Richmond, and then South Melbourne. If you like paleo bread, Malvern is the outlier at the top of the list.

What the pivot tables tell you that you may not have noticed just by looking at Figure 3.28 is that, while the City of Stonnington is ahead of the game on its own, when comparisons go by geographic location, the southern local government regions of Melbourne – Bayside and Port Phillip – combine to overtake it in the paleo-friendly eatery stakes. Most interesting is the fact that Geelong, Gippsland, Mornington Peninsula and Ballarat appear in the lists – while west Melbourne council region Brimbank, and south west region Melton, come away with no results at all.

5	Values	
6	Sum of Paleo-friendly eatery	56
7	Sum of Fresh paleo bread	41
8	Sum of Paleo-friendly shops	58

5	Values
6	Sum of Paleo-friendly eatery 56
7	Sum of Fresh paleo bread 41
8	Sum of Paleo-friendly shops 58

FIGURE 3.31 The sum of paleo-friendly eateries, shops and fresh paleo bread supply in Victoria

You do not need to use the pivot tables to generate large lists. You can use them to create smaller summaries as well, such as Figure 3.31.

Sorting

Imagine trying to find a book in an unsorted library, or a phone number in an unsorted phone book. It would be near impossible to find what you wanted. Sorting is fundamental to both modern life in general, and to data arrangement specifically. Sorting can be used to group data logically, such as by age group, or location, or to find a particular attribute, such as the largest and smallest values.

Sorting on one attribute (for example, local council region, as with Figure 3.28) is often sufficient to organise a small data set. However, if you have **big data** with many repetitions of values, such as phone books with thousands of instances of ‘Smiths’ and ‘Chens’, you would need to use multi-level sorting using several fields.

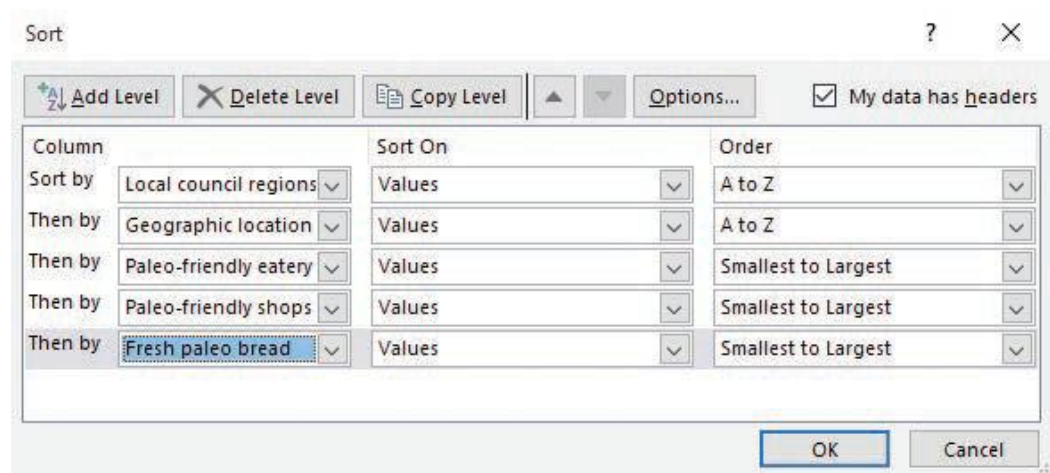


FIGURE 3.32 Data sorting options in Microsoft Excel using the data from Figure 3.28

Filtering

To reduce the bulk of visible data in a spreadsheet or database, use the Filter function. A limitation of simple sorting is that all of the original data is still visible, which makes patterns hard to see.

You can filter by any number of criteria in your data. For instance, you can view ranges that are above or below average (Figure 3.33). You can filter only those values within a certain range (Figure 3.34), above or below average, in the top or bottom 10, or you can create custom filters. You can also use text filters to view by categories (Figure 3.35).

	A	B	C	D	E
1	Local council regions	Geographic location	Paleo-friendly eater	Paleo-friendly shop	Fresh paleo bre
4	Bayside	South	3	2	2
5	Boroondara	East	5	7	6
21	Melbourne	Central	7	2	2
28	Port Phillip	South	8	3	2
29	Stonnington	Inner South East	9	7	9
33	Yarra	Inner North East	9	4	2

Symbol indicates column has a filter applied.

FIGURE 3.33 Filtered data: Regions that are above average in all three types of food resources

	A	B	C	D	E
1	Local council regions	Geographic location	Paleo-friendly eater	Paleo-friendly shop	Fresh paleo bre
4	Bayside	South	3	2	2
5	Boroondara	East	5	7	6
8	Darebin	Inner North	1	1	1
10	Glen Eira	South East	1	3	5
15	Kingston	South East	1	4	2
21	Melbourne	Central	7	2	2
23	Monash	South East	3	1	1
26	Mornington	Mornington Peninsula	1	3	1
28	Port Phillip	South	8	3	2
29	Stonnington	Inner South East	9	7	9
33	Yarra	Inner North East	9	4	2

Faint grey lines between row numbers indicate that filtering has hidden rows.

FIGURE 3.34 Filtered data: Regions with at least one of each type of resource (one eatery, one shop and one fresh bread supplier)

	A	B	C	D	E
1	Local council regions	Geographic location	Paleo-friendly eater	Paleo-friendly shop	Fresh paleo bre
4	Bayside	South	3	2	2
7	Casey	Outer South East	0	2	0
9	Frankston	Outer South	1	1	0
10	Glen Eira	South East	1	3	5
11	Greater Dandenong	Outer South East	0	0	0
13	Hobsons Bay	South West	0	1	1
15	Kingston	South East	1	4	2
22	Melton	Outer South West	0	0	0
23	Monash	South East	3	1	1
28	Port Phillip	South	8	3	2
32	Wyndham	Outer South West	0	0	1

FIGURE 3.35 Filtered data: Only southern regions of Melbourne

THINK ABOUT INFORMATICS 3.6

Use the 'Paleo diet pivot table data' file on NelsonNet to experiment with filtering. Filter the data to show only the following.

- 1 Regions that are above average for numbers of fresh paleo bread suppliers
- 2 Regions that are below average in all three types of resources
- 3 The top 10 regions for paleo-friendly eateries

Conditional formatting

Conditional formatting allows you to change the appearance of data automatically based on its current value in both spreadsheets and databases. This highlights data of special interest and can also point out errors in the data.

	A	B	C	D	E
1	Local council regions	Geographic location	Paleo-friendly eater	Paleo-friendly shop	Fresh paleo bre
2	Ballarat	North Central Victoria	0	1	1
3	Banyule	North East	0	1	2
4	Bayside	South	3	2	2
5	Boroondara	East	5	7	6
6	Brimbank	West	0	0	0
7	Casey	Outer South East	0	2	0
8	Darebin	Inner North	1	1	1
9	Frankston	Outer South	1	1	0
10	Glen Eira	South East	1	3	5
11	Greater Dandenong	Outer South East	0	0	0
12	Greater Geelong	Greater Geelong	0	1	0
13	Hobsons Bay	South West	0	1	1
14	Hume	Outer North West	1	0	0
15	Kingston	South East	1	4	2
16	Knox	Outer East	0	1	1
17	Latrobe	Gippsland	0	0	1
18	Manningham	Outer East	2	1	0
19	Maribyrnong	Inner West	1	1	0
20	Maroondah	Outer East	0	2	0
21	Melbourne	Central	7	2	2
22	Melton	Outer South West	0	0	0
23	Monash	South East	3	1	1
24	Moonee Valley	North West	0	1	1
25	Moreland	Inner North	1	4	0
26	Mornington	Mornington Peninsula	1	3	1
27	Nillumbik	Outer North East	0	1	0
28	Port Phillip	South	8	3	2
29	Stonnington	Inner South East	9	7	9
30	Whitehorse	Outer East	0	3	0
31	Whittlesea	Outer North	0	0	0
32	Wyndham	Outer South West	0	0	1
33	Yarra	Inner North East	9	4	2
34	Yarra Ranges	Outer East	2	0	0

FIGURE 3.36 This conditional formatting rule highlights any cell with a value greater than or equal to 4 in green with green text. Essentially, it is another way of flagging above-average regions.



NelsonNet additional resource: Figure 3.36 Conditional formatting, p. 164

CONDITIONAL FORMATTING

Open **Figure 3.36 Conditional formatting** on NelsonNet to experiment with conditional formatting. There are multiple ways of applying conditional formats: the following procedure is not exhaustive and may differ depending on your version of Excel.

- 1 To apply additional conditional formatting to highlight any cell with the value 'less than or equal to 1' in red font with a pink fill, use the following procedure.
 - a Highlight columns C, D and E in the Excel spreadsheet.
 - b In the Home tab in the Ribbon, click 'Conditional Formatting'. Click 'New Rule' and 'Format only cells that contain' and then, in 'Edit the Rule Description', set the values as follows, from left to right: 'Cell Value', 'less than or equal to', and '=1' (without quotes).





- c Click 'Format'. In the Font tab, choose red from the Color dropdown. From the Fill tab, click 'More Colors' and choose one of the pink shades towards the centre. Click OK, and OK again. When satisfied with the Preview, click OK to close the box.
- d The table should now show a majority of the cells highlighted.
- 2 Modify your new rule by clicking 'Conditional Formatting', 'Manage Rules', highlighting the rule you created and then clicking the 'Edit Rule' button. Change the dropdown from 'less than or equal to' to 'equal to' and change '=1' to '=0' (without quotes). Click OK, Apply, and OK to close the box.
- 3 Modify the existing green highlight rule so that it applies to any cell with a value greater than or equal to 3.
- 4 Experiment with formatting in column B to highlight only regions that contain the specific text 'West'.

Digital systems

While developing your SAT and throughout the course of your Informatics studies, you must use components of **digital systems**. The following sections discuss the composition of digital systems: hardware, software and networks. First, we begin with the hardware of digital systems, which is the physical equipment that is used for input, output, storage and communication.

Hardware

Hardware is made up of **input hardware**, which enables users to input data or provide commands to software, **output hardware**, which gives information to the user, **storage hardware**, which stores data and software, and **communication hardware**.

Input and output

The commonly used types of input hardware are described in Table 3.8.

TABLE 3.8 Types of input hardware

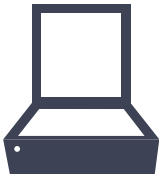
Hardware	Description
Keyboard	Keyboards in Australia usually use the familiar QWERTY layout and are used to type commands
Mouse	Used for controlling a graphical user interface (GUI) that has windows, icons, menus and a mouse pointer
Trackpad	Credit-card sized replacements for a mouse; sensitive to touches and taps and found beside or below the keyboard of laptop computers
Touchscreen 	Common on phones, tablets, laptops, supermarket self-checkouts, ATMs and information kiosks; function as both monitor and mouse replacement
Flatbed scanner 	Digitise printed (analogue) text and images

TABLE 3.8 (continued)

Barcode reader	Used in shops and industry to increase the ease, speed and accuracy of entering product identification codes
Graphic tablet and stylus	Replaces a mouse for inputting intricate hand gestures for graphics, symbols, handwriting and signatures
Custom input devices 	Includes buttons, knobs, switches and dials on game controllers; ticket machines; vehicle dashboards; music players; calculators; phones; remote controls and electric kitchen devices

The most common types of output hardware are monitors and printers. Monitors and printers have both undergone significant technological advancements. Both come in many types. Some of the most popular types of monitors are described in Table 3.9.

TABLE 3.9 Popular types of monitors

Monitor	Description
LCD <i>Liquid crystal display</i>	• Thin • Colourful • Uses comparatively little electricity • To be visible, LCD pixels (picture elements) need a backlight
OLED <i>Organic light-emitting diode</i>	• Pixels glow to produce the display • When turned off, a pixel is black, which produces greater contrast and richer colours
Plasma	• Pixels glow and are truly black when turned off • Uses more electricity than LCD • Cheaper than LCD for very large screens and offers richer colours and better contrast
Electronic paper	• Electrophoretic ink (for example, eInk) • Requires no electricity to hold a text or image display once it has been drawn • Extremely low power consumption • Ideal for battery-powered devices with large displays that are turned on for long periods, such as eBook readers (Figure 3.37)

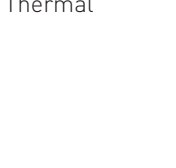
The most commonly used printer types are **laser**, **inkjet** and **thermal**. They are described in Table 3.10.

All printers tend to include lights, dials and indicators to signal activity or status. They also tend to use speakers for alarms and sound effects when there is a problem, or to indicate completion of a task. You have likely often heard the beep your printer makes when it will not print because of a paper jam or because it is out of ink or toner.



FIGURE 3.37 An electronic paper ‘eInk’ display on an eBook reader

TABLE 3.10 Laser, inkjet and thermal printers

Printer	Description
	• Fast, high-resolution, waterproof pages • Monochrome or colour text and graphics • Uses a laser beam to draw an entire printed page onto a rotating drum. The drum passes over ultra-fine plastic powder (toner) and is drawn to the static electricity left where the laser beam hit the drum. The drum transfers the loose toner powder onto a sheet of paper and a heater melts the toner onto the paper
	• Sprays ink onto a page one line at a time • Cheap to buy • Ink is very expensive
	• Used in industry for labels and barcodes, and in retail for receipts • Operates in near silence • Uses very little electricity, which makes them the best choice for battery-powered portable printing, such as for printing parking fines • Printing heads burn heat-sensitive paper to leave text or images • Text and images will fade after a relatively short time, which makes them not a good choice for longevity

Storage

Storage hardware stores data and software for immediate or later use. Storage hardware therefore is made up of **primary (short-term) storage**, which is what you know as random access memory, or RAM, and **secondary (long-term) storage**. Secondary storage includes many kinds of disk drives.

RAM is composed of billions of storage locations in silicon chips. RAM stores program instructions and data when programs are running. RAM chips are volatile and lose their data when power is turned off.

Secondary storage options permanently store data, information and applications when they are not being used. This includes when power is turned off. Table 3.11 discusses three of the most popular secondary storage options: hard disk drives, solid state drives and network-attached storage.

TABLE 3.11 Popular secondary storage devices

Storage	Description
	• Aluminium disks densely crammed with magnetically recorded bits of 1 and 0 (binary code) • Spinning at up to 10 000 RPM, they store and retrieve data at incredible speed, with breath-taking accuracy and reliability • Very cheap per megabyte of capacity; still the largest, most reliable long-term storage media • In 2015, a 6TB (6 terabytes = 6000 gigabytes) HDD cost approximately \$375AUD – around 16GB of storage per dollar

FIGURE 3.38a The inside of a HDD

THINK ABOUT INFORMATICS 3.7

Explain conditions where laser and inkjet printers are either ideal or unsuitable.

TABLE 3.11 (continued)

Solid state drive (SSD)  FIGURE 3.38b The outside (top) and inside (bottom) of an SSD	• Non-volatile memory similar to USB flash drives and SD cards • Run silently, start instantly, generate less heat and use less electricity • Have no motors that will age and eventually fail • Tends to access data faster than a HDD • Stores less data per square centimetre of storage space • Have a limited number of times they can re-write a memory cell; after approximately 1 million writes, a memory cell will become unpredictable, or fail • Quite expensive: in 2015, a 128GB SSD cost approximately \$105AUD – around 1GB of storage per dollar
Network-attached storage (NAS)  FIGURE 3.38c A NAS device	• A networked team of HDDs in a box • Makes file sharing easier • Increases capacity considerably (for example, 12TB) • Offers data protection, such as hot-swap disks • Convenient and reliable

TABLE 3.12 RAM and data storage units

Unit	Symbol	Equivalent to:	
		RAM	Data storage
Byte	B	8 bits (1 or 0), the basic unit of storage	8 bits
Kilobyte	KB	1024 bytes	1000 bytes
Megabyte	MB	1024KB (roughly 1 million bytes – the size of two average novels)	1000KB
Gigabyte	GB	1024MB (PCs have gigabytes of RAM)	1000MB
Terabyte	TB	1024GB (hard disks have terabytes of storage)	1000GB
Petabyte	PB	1024TB	1 000 000GB
Exabyte	EB	1024PB	1 000 000 000GB
Zettabyte	ZB	1024ZB	1 000 000 000 000GB
Yottabyte	YB	1024ZB	1 000 000 000 000 000GB

For more information on input, output, storage and network hardware, see Chapter 3 of *Computing VCE Units 1 & 2*.



Networks and communication

Networks and communication hardware come in a variety of forms.

Ports are the physical sockets that carry data between a computer and peripheral (external) devices. For example, USB (universal serial bus) is an industry standard high-speed protocol that connect many devices, such as Flash drives, printers, modems, keyboards, mice, speakers and smartphones.

Modems convert digital data into analogue data for transmission over non-digital media, such as telephone lines, and convert incoming analogue data into digital data for the computer to use. The two most common types of modem are ADSL and cable. ADSL is short for asymmetric digital subscriber line; asymmetric is a reference to the fact that the modem downloads (receives) data faster than it uploads (sends). Cable modems use pay TV cables, such as Foxtel and Optus.

Switches are boxes that allow multiple network cables to interconnect and exchange data between networked devices.

Routers are used as gateways to interconnect LANs and they guide data packets across networks and the internet. At home, your router is normally built into your modem and it protects your LAN from the outside world.

Cables are used to connect two devices, enabling the transfer of signals from one device to another. The most common networking cable is CAT, a metal-core twisted-pair cable, followed by a number; for example, CAT6. Its maximum length is 100 metres, and its maximum bandwidth is 1000 megabits per second (Mbps). For longer distances and higher speeds, glass fibre optic cable (FOC) is used. FOC can run for kilometres, is immune to electrical interference, runs at near light-speed, and has massive bandwidth because, unlike CAT cable, it can carry multiple signals along a single thread. FOC is used to span oceans and continents.

Wireless access points are devices used on wireless LANs. They act as central transmitters and receivers of wireless radio signals and allow wireless devices such as phones or tablets to connect to the wired network. Although wireless is cheap to install and flexible to configure, it is slower and less secure than a wired connection. If you use wireless data, you should encrypt your data to safeguard your personal and sensitive information.

Software

Software is the programming code that controls hardware. Software comprises:

- application software – for example, a word processor – to do work for the user and create information
- system software – for example, an operating system, device drivers – to control hardware and allocate computer resources so application software can run
- utility software – for example, a text editor – to provide a single, specific service to extend the functionality of a digital system.

Software is written in source code that uses the strict syntax of a particular programming language (for example, C, Basic) that is converted by compiler software into executable code (for example, notepad.exe) that a computer's central processing unit (CPU) can understand and carry out.

We have already discussed spreadsheet software, such as Excel, which is relevant to your Outcome. Databases such as Microsoft Access and FileMaker, and word processors such as Microsoft Word, have also been mentioned.

Gantt charts

GanttProject is downloadable, free, open-source software. All of the Gantt charts used in Chapter 3 were made using GanttProject. Another open-source option is ProjectLibre.

There is also Microsoft Project, which comes with the professional version of the Microsoft Office suite.

If you wish to use web-based software, both GanttPRO and Tom's Planner are free for basic or personal users.

Editing images and creating artwork

To edit images or create new artwork, you could use Adobe Photoshop and Illustrator. For an open-source, free option, you could download GIMP. Another option is to use Microsoft Word's 'Smart Chart' feature to create hierarchies and flowcharts and other types of charts. It is surprisingly customisable and easy to use. To create mockups and annotated diagrams, Balsamiq Mockups is free for 30-day trial usage. For a list of Mashable's ranking of the nine best free image editors, view the weblink opposite.



Switch



Router



Wireless access point



[GanttProject download](#)



[ProjectLibre download](#)



[GanttPRO](#)



[Tom's Planner](#)



[Mashable: 9 best free image editors](#)



[GIMP download](#)



[Balsamiq download](#)



SurveyMonkey

Tableau Public
download

Inspiration download

Survey tools

Google Forms and SurveyMonkey both provide browser-based, free, easy-to-use and easy-to-distribute templates for making questionnaires. Responses to questionnaires are easy to access and you can set up alerts to be notified when new responses are received.

Data visualisations

Most browser-based data visualisation tools are offered on a trial basis: this includes LucidChart and PiktoChart. There are exceptions to this of course, but you may need to look around for free tools that best suit your needs.

Google Charts is another offering within Google Drive's suite of free tools.

To create mind maps of your data, you could download a trial of Inspiration.

You can download Tableau Public to visualise your data for free within minutes – but do be aware this may also *share* your data as a condition of use.

Data security

As part of Unit 3, Outcome 2, you must select and apply methods to secure and store data and information. Losing the data that you have painstakingly gathered for your Outcome would be catastrophic. To a business, the loss of data, depending on magnitude, can be anywhere from minor to fatal. For example, if a company lost records of accounts payable, it would not be able to keep track of what had been paid and what still needed to be paid. If a retail store lost track of stock lists, it would have no way of knowing if stock levels indicated a great deal of shoplifting had been taking place, or if staff had been making mistakes during transactions.

Businesses may lose trade secrets to competitors and lose their reputations as trustworthy organisations if they fail to protect their data and information. They also face prosecution by the Australian Tax Office if their tax records are lost, as well as prosecution under the *Privacy Act 1988* if they violate the Act and personal information is lost, damaged or exposed as a result.

In addition, they may lose income if business cannot carry on because of the loss, and may find they are unable to pay wages.

Clearly, it is very important that businesses protect personal and corporate data from accidental or deliberate loss, damage or theft. This is an ongoing responsibility for them. However, it is vitally important for you as well. Although it may only seem relevant for your Outcome, developing good habits for data security right now is something that is worthwhile for the long term.

Just as businesses can lose their data and privacy, you can lose yours. Unfortunately, malicious behaviours such as identity theft and doxing do happen. There are also a multitude of other data security threats that are not malicious, but are nevertheless real and dangerous. Regardless of how unlikely it may seem, it is best to employ data security measures. Data security takes two main forms: physical and software. The next section discusses the physical and software security measures relevant to you and your data.

Physical security

Physically keeping threats away from your data is a logical and effective first step.

- If you use a laptop or tablet and store your data on it, make sure you keep it in a secure place when you are not using it, such as in a cabinet. Otherwise, store it out of sight.
- Do not let people you do not know very well use your devices.
- If a friend or family member needs to use one of your devices, make sure they cannot access important data.

Doxing involves researching a person, sometimes known only by a handle (nickname or screenname) and then publishing their personal information, such as their full name, address, phone number, workplace and date of birth, online to identify them to as large an audience as possible.

- Keep your doors and windows locked to prevent theft of your hardware.
- If you use a desktop computer, keep it switched off when you are not using it.
- Consider using surge-protector power outlets for all of your devices to protect the data stored on them.

Software security

Software security is extremely important to you and your data. Threats can come from anywhere at any time, and they can be both deliberate and accidental. Threats can come across a network and do not rely on physical access to equipment.

Use strong passwords

Ideally, passwords should be at least 8 characters in length and include a combination of both uppercase and lowercase letters and numbers. Depending on the software, you may also be able to include special characters such as punctuation. Try not to use the same password on every login. If you do, and a hacker gets into one of your accounts, they will then have access to *all* of them. Passwords that contain common words are easily guessed, so are not recommended.

Examples of strong passwords:

tYjkl321pC
n0ts4y1ng!
p4rs3w0rd?!

Examples of weak passwords:

Sunshine1
nothing00
password

It is also a good idea not to use the same ‘Secret question’ information on every account. While it may be convenient to always use your mother’s maiden name or the name of your first dog as your ‘Forgot password?’ reminder, if your hacker finds this out, it will make it very easy for them to take over every single one of your accounts using the ‘Forgot password?’ feature.

For some logins, you can use two-factor authentication. For instance, a login to MyGov.au requires a password, followed by the entry of a secret code that has been sent to the user’s mobile phone. Two-factor identification relies on identifying people because they both know something (a password) and possess something (such as the mobile phone).

Use login passwords

You should use login passwords on your laptop, tablet, desktop computer and any other electronic device that has personal or sensitive information on it. You do not want to risk losing it or having it stolen, and having someone else switch it on and have immediate access to everything you have stored on it.

Use biometric identification

While passwords are the only way to control access to remote computers and resources (currently), biometric identification can be used to control computers and resources when the user is physically present. Biometric data cannot be lost, stolen, guessed or discovered easily. Biometric signatures are unique, and include fingerprints, iris patterns (the coloured part of the eye) and retinal patterns (the blood vessels at the back of the eye). Biometric identification has long been used at Los Angeles Airport. For years, non-US citizens underwent fingerprint scans to enter the country. Now, the airport is considering an upgrade to iris scan technology.

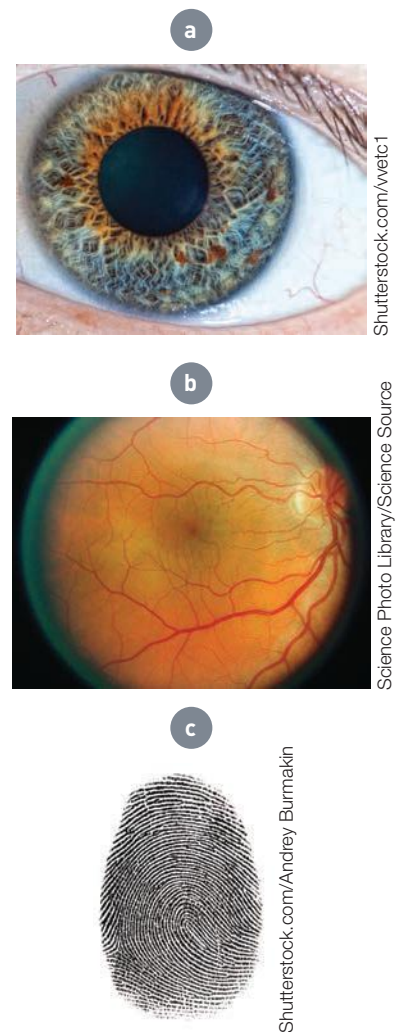


FIGURE 3.39 Biometric data: While most types of biometric data are not yet relevant to you for personal use, such as **a** iris pattern and **b** retinal pattern, **c** fingerprints are relevant.

Some laptop manufacturers, such as Acer and Toshiba, sell laptops with built-in biometric identification systems in the form of fingerprint scanners. If you have a laptop with a fingerprint scanner (Figure 3.40), scan your own fingerprints and use them as your default login for added security.



Gene Anderson-Conklin

FIGURE 3.40 The fingerprint scanner positioned directly below the trackpad (in-built mouse replacement) on a laptop computer. The user only need run their fingerprint over the brown stripe in the gold section to capture their fingerprint scan.



LA airports prepare for biometric ID security



LAX unveils new biometrics

Always log out

When you are not using a computer any more, do not leave it logged in. Log out and turn-off the monitor. It is easy to forget to log out, but leaving the computer logged in to your user account leaves your data vulnerable to anyone who walks past and sees that the computer is still logged on.

Log out of websites when you are not using a personal device. In fact, it is better to log out of websites even when you are using your own personal devices, but people rarely do this, which makes it easier to steal data and identities from people whose devices have been stolen.

Imagine if your phone was stolen. You left yourself logged in on Dropbox on your smartphone, which was storing files from your Outcome. The thief decided to clean out the phone and wanted to see if there was anything of value on Dropbox. Having no interest in Informatics, he deleted all the files.

Next, he wanted to see what you had in your email, so he went into your email account, which was logged in, and changed the password. It was easy for him to do this, because your phone browser gave you the option to view your saved passwords. Knowing it would be easy for you to get control of your email account back, he quickly changed your 'Forgot password?' questions.

Perusing your email inbox, he established that you had not deleted your monthly bank statement email. It included your bank account number. Suspecting you were a creature of habit, he went to the bank's website and logged in using the same password as your email account. Using a few quick steps, a thief can access your bank account. You can prevent this if you avoid storing passwords in your browser, always log out, and do not use the same password repeatedly.

Encrypt your data

Always encrypt your data. This is especially true if you connect using wireless at home. Wireless data is more vulnerable to threats than wired data. Encryption makes data unreadable to unauthorised people, even if they manage to overcome other security measures and steal it.

Modern encryption uses **public key encryption**, which does not need a key to be sent to unlock it. Previous encryption methods required a secret unlocking key to be sent with the encrypted data. If the data were intercepted, the key could also be captured and the data unlocked.

Public key encryption is the basis of SSL and TLS, which is used to encrypt the web traffic between servers and browsers. If intercepted in transit, the traffic is unreadable. PGP (pretty good privacy) is software that also uses public key encryption to protect documents. Wireless signals also use public key encryption to prevent snooping and unauthorised use of wireless networks.

The bigger the numbers used for public key encryption, the harder the encrypted data is to decode. For greatest safety, choose the largest encryption key you can (128-bit or 256-bit is the current recommendation).

Use a firewall



A firewall will prevent unauthorised access to your data and information, and deny network access to outsiders. Essentially, it will separate the internet and other networks from the computer or LAN on which it is installed. A firewall examines the content of incoming data packets and determines whether they should be allowed to pass through.

High-quality free firewall software includes ZoneAlarm, Comodo Firewall, PeerBlock, Privatefirewall and Anti NetCut 3.

Use antivirus software

Most antivirus programs handle a whole lot more than just viruses now. Many kinds of malware are around these days (Table 3.13), so most antivirus programs are equipped to handle at least a few of them.

TABLE 3.13 Common types of malware

Malware	Description
Viruses	- Damaging code that attaches to executable files and travels with them Payload is triggered by human actions, such as running programs - True viruses are now rare, but the name persists as a generic term for malware
Worms	- Copy themselves and travel with no human intervention or need to attach to other files - Most travel via email or over local area networks
Spyware	- Monitors the behaviour and reports browser activity to the spyware's operator - Can hijack browsers to send users to unwanted sites, or show targeted advertising based on a user's browsing history
Trojans	- Any malware that enters a system by pretending to be desirable - Often use 'social engineering' to trick people into downloading or installing them

If you are not really convinced of the dangers of malware on your devices, think again. Payload is a term that describes the destructive potential of malware. Table 3.14 lists some known malware payloads.

THINK ABOUT INFORMATICS 3.8

Breaking a 256-bit key by brute force (simply trying every possible key in turn) is not simply twice as hard as cracking a 128-bit key: it is 2^{128} harder.

To help you visualise that value:

- 2^2 is 4 times harder
- 2^{10} is 1024 times harder
- 2^{20} is more than a million times harder
- 2^{21} is more than 2 million times harder, and so on.

1 Exactly how much harder is 2^{128} ?

2 Calculate how many billions of years it would take to crack a 256-bit encryption key using 50 supercomputers.



ZoneAlarm download

For more information about firewalls, see Chapter 3 of *Computing VCE Units 1 & 2*.

TABLE 3.14 Known malware payloads

Payload	Destructive potential
Keylogger	Can record users' keystrokes including passwords, credit card information and bank account logins, and send the data to the malware operator
Distributed denial of service (DDoS) attack code	Makes a computer vulnerable to remote control by its operator, along with thousands of other infected machines, to participate in a DDoS attack on a remote victim. A computer that might seem sluggish to its user might actually be sending millions of information requests that can render the victim's computer unable to operate. DDoS attacks are often used to blackmail victims into paying protection money, or to attack political or religious enemies.
Adware	Inserts unwanted advertising into visited websites Not deliberately destructive, but often poorly programmed and can dramatically slow computers down or cause crashes
Spam server	Sends thousands of spam emails using the victim's computer. If discovered, innocent computer owners are identified while the spam operator remains undetected.
Ransomware	Encrypts documents on the victim's computer so they are inaccessible to the computer owner. The malware operator then demands payment from the victim to receive the key to unlock their documents.
Root kit	Particularly nasty malware that actively hides from the operating system and works invisibly in the background, often turning a computer into a remote-controlled 'zombie'
Deleting files or damaging operating systems	Such petty vandalism, once common, is now rare. Cyber-attacks are now dominated by large, organised crime syndicates and governments.

While your home computer is unlikely to be deliberately targeted by international hackers who are determined to steal your valuable drafts of Informatics solutions, you do need to protect your computer system. Having a sound understanding of electronic self-defence is necessary for both your digital system and your end-of-year Informatics examination.

Even the most careful computer user cannot guard against worms or 'drive-by downloads' that can cause infection simply by visiting an infected website. However, keep the following in mind.

- 1 You *must* use a reputable, reliable anti-malware scanner.
- 2 You *must* always have it running and scanning opened and downloaded files for known threats.
- 3 You *must* keep your virus definitions up-to-date.

Be aware of false positives. Sometimes a scanner can report a virus that does not exist. Free online scanners sometimes report false positives to scare users into buying their products.

Similarly, be aware of false negatives. Some scanners may be unable to detect existing viruses. This may occur if your virus definitions are out of date, and with newly released 'zero day' threats.

If possible, have a discussion with your parents or guardian about paying for a reputable malware scanner to avoid false positives. (Some free online scanners, ironically, also have in-built adware.) If this is not an option, be careful and research carefully to choose an antivirus program that is right for you.



AVG Antivirus Free



Panda Free Antivirus

Back up your files

Despite all of your best efforts, disasters may still happen. Data backups are the final defence against total data loss.

Back up your files, especially any files you create for your Outcome (and your VCE studies in general), **at least once a day**. It is essential to back up your files for your Outcome. You cannot lodge a request for consideration if you lose your digital files. While a minimum once a day is recommended, it is better to err on the side of caution and back up **every time** you make a significant change.

Store them away from your main device. While companies can store their backups offsite, the best you can do is probably store your backups on external drives as well as on your internal hard drives. Alternatively, consider backing up your files on the cloud using a service such as Dropbox or Google Drive.

Test your backups to make sure they work. Also regularly create system restore points on your computer in case anything goes very wrong and you need to restore to an earlier version.

Next steps

In Chapter 2, we discussed formulating a hypothesis, the requirements of the solution for Unit 3, Outcome 2, and we also discussed data in many forms: primary and secondary, qualitative and quantitative, how to gather it directly, how to gather it using resources and how to reference it correctly in your report.

In this chapter, we have talked about legal requirements of the data and information you have collected for this Outcome. We have also discussed the elements of project planning and management, with a particular emphasis on Gantt charts, as well as strategies for file naming. We have talked about how to organise and store the data you have collected, as well as how you can identify the patterns and relationships within it using a variety of tools and techniques.

Next, we have discussed the digital systems that you have been using throughout your studies: the hardware, software and networks enabling your data gathering and research.

Finally, we have discussed data security, with regard to how it relates to you and your studies.

Your next step, upon completion of the chapter summary, is to work toward completion and submission of the solution for Unit 3, Outcome 2, according to your teacher's instructions.

ESSENTIAL TERMS

averages a method of identifying patterns in data; the type of average selected for a given task – mean, median and mode – should be chosen with care

big data any huge quantity of data that has the potential to be mined for information

CamelCase the use of capital letters to mark the start of words in a file or object name; for example, **AnnualReport2017.xls**

controlled variable a principle of data collection; when collecting data, it is crucial to account for as many variables as possible so only the chosen independent variable can be the cause of the change in the chosen dependent variable

correlation a measurement of how closely changes in one data set relate to changes in a different data set; correlation does not imply causality or indicate which factor is the cause and which is the effect

curating organising, storing, protecting, updating and cataloguing a data collection

digital systems hardware for input, output, storage and communication

FTP file transfer protocol, used to send large files across the internet; webmasters use FTP to send webpages to their site's web server

Gantt charts project management tools that show tasks as bars on a timeline, manage resources and are modified over time to monitor project progress

milestone marker of major points of progress; often shown as diamonds in Gantt charts

outlier a figure that differs so dramatically from the rest of the data that it is not at all typical or representative of the majority

payload the destructive potential of malware

sequencing tasks putting tasks in order of time and specifying which tasks are dependent on other tasks finishing first

significance tests tests that check whether the difference between data sets are more than just random chance

slack the amount of time by which one task is shorter than a concurrent task

standard deviation a measurement of how much a data set varies from the mean

work breakdown structure (WBS) breaking down a complex project's tasks into a complete list of separate, self-contained, achievable tasks

IMPORTANT FACTS

- File names** should be meaningful, concise, consistent and planned, and not use spaces.
- File systems** are organised in a hierarchical structure.
- The **cloud** is an increasingly popular way of storing and manipulating data.
- Databases and spreadsheets must be **organised** to be effective.
- Organise data with **sorting** and **metatags**.
- Regular expressions** are a powerful way to search and replace text.
- Data visualisations make **data patterns and relationships** clearer than lists of numbers.
- Choose **types of charts** with care so they visualise data clearly and accurately.
- Data can be **queried** to find relevant records.
- Use **AND** and **OR** to create powerful selection criteria in queries.
- Pivot tables** are a powerful way to reorganise spreadsheet data in many ways.
- Sorting, filtering** and **conditional formatting** help to highlight interesting and relevant data.
- Data can be secured **physically** (for example, locked doors), with **software** (for example, passwords, encryption, access hierarchy) and with appropriate procedures.
- Backups** must be regular, stored offsite and tested. They are a key component of a **data disaster recovery plan** that all organisations should devise and practise.
- Malware**, such as worms or viruses, can cause significant security problems and must be guarded against.

TEST YOUR KNOWLEDGE

LEGAL REQUIREMENTS

- 1 List entities who are subject to the *Privacy Act 1988* (amended 2012).
- 2 What principles underlie Australian privacy legislation?
- 3 What legislation should you with your Outcome to avoid letting your research become a nuisance?
- 4 What do you need to do to have copyright protection over the work you create for your Outcome?



Review quiz

PROJECT MANAGEMENT

- 5 Explain the purpose of a work breakdown structure.
- 6 In a Gantt chart, how can you identify concurrent tasks?
- 7 Give a real-world example of how a task may be dependent on its predecessor. Explain how this dependence would be represented in a Gantt chart.
- 8 Task 1 is scheduled to take 5 days. Concurrent task 2 takes 4 days. Which task has slack, and how much slack does it have?
- 9 What is a milestone, and how do you recognise one in a Gantt chart?
- 10 Explain the importance of including resources in a Gantt chart.
- 11 Why are Gantt charts modified as a project proceeds?

FILE NAMING STRATEGIES

- 12 Write a brief list of file naming rules suitable for a novice computer user.
- 13 To this list add four file naming conventions.
- 14 Explain how CamelCase and Hungarian notation improve file name effectiveness.

ORGANISING AND STORING DATA

- 15 How do hierarchical systems help data storage and retrieval?
- 16 How can using the cloud (online data storage or processing) help organisations organise data?
- 17 Why is normalisation of a relational database so important?
- 18 Give three tips for organising a spreadsheet effectively.
- 19 Explain two important factors when naming objects that will affect the usefulness of sorting.
- 20 How can metatags be used to organise and manage media data files more effectively?
- 21 How does synchronising improve the integrity of stored data?

IDENTIFYING PATTERNS AND RELATIONSHIPS IN DATA

- 22** Outline the differences between mean, median and mode averages.
- 23** When is a significance test valuable?
- 24** Data set A has a standard deviation of 2. Data set B has a standard deviation of 8. Which data set has more consistent data?
- 25** 'Correlation does not necessarily mean causation.' Explain what this means and why it is important to remember.
- 26** How can data visualisation make your multimodal online solution more effective?
- 27** Write a query that would find all males over 35 years old who live in Melbourne or Sydney.
- 28** How can a spreadsheet's pivot table reveal interesting patterns in data?
- 29** Explain how conditional formatting, sorting and filtering can reveal useful data features.

DIGITAL SYSTEMS

- 30** Create a table listing digital system components for input/output, storage and communication. For each component, outline the main types that exist and their strengths and weaknesses.

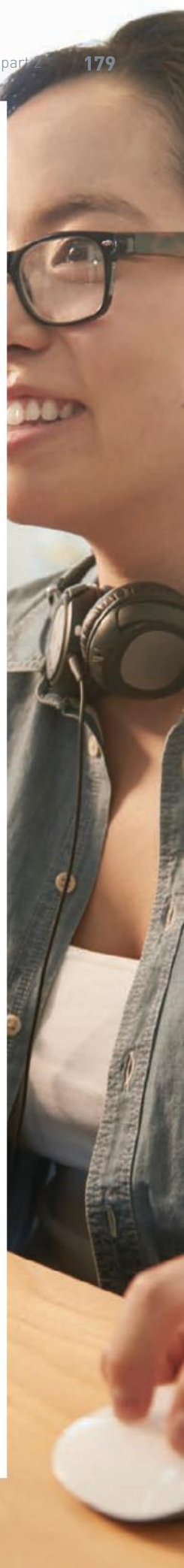
DATA SECURITY

- 31** Your family is concerned that the precious digital photos on your home computer may be eventually lost or damaged. List deliberate and accidental threats to these photos. Recommend physical and electronic strategies and techniques for protecting and recovering these irreplaceable files.

APPLY YOUR KNOWLEDGE

This will continue your work on the hypothesis you investigated in Chapter 2's Apply your knowledge section.

- 1 Look back on the hypothesis you investigated in Chapter 2. Why did you choose this hypothesis?
- 2 Justify the reasonableness of your hypothesis.
- 3 Identify your hypothesis' independent and dependent variables.
- 4 Describe any constraints that affected your data acquisition, including availability of data, time and resources.
- 5 Explain the processes you used to locate, select and acquire qualitative and quantitative data.
- 6 Comment on the integrity of the data you acquired, including possible faults such as inaccuracies, biases and incompleteness
- 7 Explain the software functions and techniques you used to process the data to generate information to support or refute your hypothesis.
- 8 Summarise the functional and non-functional requirements of the information you generated from the data you acquired.
- 9 Complete the Gantt chart covering all of the tasks required for the 'Apply your knowledge' activities in Chapter 2 and Chapter 3. It should include milestones and dependencies, tasks, task sequencing, time allocation and resources.
- 10 Briefly justify the file naming conventions you used for all documents you created for the 'Apply your knowledge' activities in Chapters 2 and 3.
- 11 Explain what led you to discover patterns and relationships in the data that generated information that informed the conclusion you drew about your hypothesis.
- 12 Summarise the hardware and software you used to input, store, communicate and output data and information to research your hypothesis.
- 13 Explain what physical and software controls you took to protect the integrity of stored and communicated data and information.



PREPARING FOR

UNIT 3 OUTCOME 2

Use a range of appropriate techniques and processes to acquire, prepare, manipulate and interpret complex data to confirm or refute a hypothesis, and formulate a project plan to manage progress

To achieve this Outcome, you will draw on knowledge and skills initially outlined in Area of Study 2. This Outcome begins the School-assessed Task (SAT) that will be concluded in Unit 4, Outcome 1.

OUTCOME MILESTONES

- 1 Read the instructions provided by your teacher.
- 2 Get the dates for the start and end of the SAT from your teacher.
- 3 Begin your Gantt chart.
- 4 Formulate your hypothesis.
- 5 Begin your reference list or bibliography.
- 6 Gather data and determine its integrity.
- 7 Establish an appropriate file management system.
- 8 Prepare the data for manipulation, including coding, if required.
- 9 Manipulate the data.
- 10 Interpret the information and conclude whether the hypothesis is supported or refuted.
- 11 Write a short report containing your hypothesis, your conclusion and an explanation of how the information you generated supports that conclusion.
- 12 Prepare for submission both your collection of data sets, and the information derived from them.
- 13 Explain the specifications used for creating the information.
- 14 Finalise your acknowledgment of intellectual property using one of these styles: APA, Chicago, IEEE or Harvard.
- 15 Show evidence of your validation processes and techniques.
- 16 Show evidence of your data manipulation processes and techniques.
- 17 Explain how you secured data and information during storage and communication. Also acknowledge the legislation appropriate to the security and communication of the data.
- 18 Finalise your Gantt chart.

STEPS TO FOLLOW

- 1 Read the instructions provided by your teacher. Note: No case study, hypothesis or data will be provided.
- 2 The start and end dates of the SAT will be provided by your teacher. Note: Your teacher may determine when certain parts of the Outcome need to

be completed, such as submission of the hypothesis. Schools will be provided with a form associated with authenticity. Work needs to be submitted and signed by both you and your teacher. These dates can be supplied by your teacher and included in your Gantt chart.

- 3** Begin your Gantt chart to plan and monitor your progress during the Outcome. As the project progresses, update the chart to match reality. Be sure to include:
 - a** project milestones within the start and end dates provided by your teacher, including development tasks during Unit 4, Outcome 1
 - b** identification of tasks
 - c** sequencing of tasks, including the identification of concurrent tasks
 - d** time allocations for tasks
 - e** allocation of resources.
- 4** Generate a reasonable hypothesis, keeping in mind that relevant data must be available to support or refute it.
- 5** Begin your bibliography to acknowledge intellectual property. You are not required to obtain copyright permissions, but you must acknowledge all primary and secondary data sources using one of these methods: Harvard, APA (latest edition), Chicago, or IEEE.
- 6** Gather relevant, valid and reliable complex data from more than one source. Keep your bibliography up-to-date as you search for data.
- 7** Organise and prepare the data for manipulation with an appropriate software tool, such as a spreadsheet or database.
- 8** Manipulate the data to produce accurate, relevant and meaningful information that will allow you to reach a logical conclusion about your hypothesis.
- 9** Interpret the information and reach a conclusion about whether the hypothesis has been supported or shown to be false.
- 10** Write a short report that sets out:
 - a** a statement of your hypothesis
 - b** the conclusion you have drawn from the information you generated from the data
 - c** an outline of the findings supporting your conclusion.
- 11** Prepare for submission your collection of data sets, and the information derived from them. Printing large data sets may prove to be unwieldy, slow, expensive and environmentally unfriendly. Consult your teacher about options for data file formats, packaging and exchange.
- 12** Explain the specifications used for creating the information. Include:
 - a** functional and non-functional requirements
 - b** constraints
 - c** scope.
- 13** Finalise your bibliography so that all data sources are acknowledged.
- 14** Show evidence of your data validation processes and techniques.
- 15** Show evidence of your data manipulation processes and techniques.
- 16** Explain the methods you used to secure the data and information you stored and communicated.
- 17** Show evidence of a project plan (Gantt chart) indicating times, resources and tasks up to the milestone that concludes Unit 3, Outcome 2.

DOCUMENTS REQUIRED FOR ASSESSMENT

- 1** A short report that sets out:
 - a** a statement of your hypothesis
 - b** the conclusion that has been drawn
 - c** an outline of the findings supporting the conclusion
- 2** A collection of data sets, and information derived from them, that allows a conclusion to be drawn about the hypothesis
- 3** Evidence of:
 - a** the specifications for creating the information
 - b** your bibliography acknowledging intellectual property
 - c** validation and manipulation processes and techniques used
 - d** methods used to secure stored and communicated data and information
- 4** A project plan (Gantt chart) indicating times, resources and tasks

ASSESSMENT

A more detailed set of assessment criteria will be available on the VCAA websites on the Informatics study page. This will include mandated criteria and rubric, and there will be specific requirements about the complexity of the data. The SAT comprising Unit 3, Outcome 2 and Unit 4, Outcome 1 will contribute 30 per cent to your Informatics study score.

INTRODUCTION

In Unit 3, Outcome 2 you formed a hypothesis, collected data on it and reached a conclusion about whether the hypothesis was supported or refuted.

In Unit 4, Outcome 1, you will create a multimodal online solution that uses multiple media types to explain your hypothesis and present your findings to educate a global audience.

You will especially focus on good design (including several alternative design ideas) and communicating your message clearly to the global audience.

You also evaluate how effectively your multimodal solution achieves its goals.

Throughout the creation of your solution, you will use and evaluate the project management plan you began in Unit 3, Outcome 2.

AREA OF STUDY 1: DATA ANALYTICS: PRESENTING THE FINDINGS

Outcome 1

In this Outcome, you will design, develop and evaluate an online solution that is multimodal (uses multiple forms of media). Your solution must aim to educate a global audience about the hypothesis that you formed in Unit 3, Outcome 2 and your findings about whether that hypothesis is supported or refuted. You will also evaluate the effectiveness of your solution and your project plan.

AREA OF STUDY 2: INFORMATION MANAGEMENT

Outcome 2

In this Outcome, you will compare and contrast the strategies of two different organisations in managing the storage and disposal of data and information. As part of this, you must recommend improvements that each organisation can make to their current information management practices.

CHAPTER

4

DATA ANALYTICS: PRESENTING THE FINDINGS

Key knowledge

After completing this chapter, you will be able to demonstrate knowledge of:

Approaches to problem solving

- characteristics of information for educating world-wide audiences, including gender and culture inclusiveness, commonality of language, age appropriateness
- techniques for generating design ideas
- criteria for evaluating alternative design ideas and the effectiveness of solutions
- characteristics of effective multimodal online solutions
- formats and conventions appropriate to multimodal online solutions
- design principles that influence the functionality and appearance of multimodal online solutions
- design tools for representing a solution's appearance and functionality, including relationships, where appropriate
- functions, techniques and procedures for efficiently and effectively manipulating data using software tools
- manual and electronic validation techniques
- functions, techniques and procedures for managing files
- techniques for testing that solutions do what is intended
- techniques for documenting the progress of projects, including annotations, logs and adjustments to tasks and timeframes
- strategies for evaluating the effectiveness of solutions and assessing project plans.

Key skills

- generate alternative design ideas
- select preferred design ideas, based on student-developed criteria
- select and apply design tools to represent the functionality and appearance of solutions
- select and apply software functions, methods, formats, conventions, techniques and design principles to develop multimodal online solutions that operate as intended
- monitor and adjust project plans where appropriate
- apply criteria to evaluate the effectiveness of multimodal online solutions in communicating conclusions to hypotheses
- assess the effectiveness of project plans in managing work practices.

For the student

This chapter discusses the key knowledge required to complete Unit 4, Outcome 1. It covers project management, evaluation techniques, how to write well for global audiences and how to design creatively.

For the teacher

This chapter is based on Unit 4, Area of Study 1, and provides the key knowledge required to complete Unit 4, Outcome 1.

What is a multimodal online solution?

A multimodal online solution (MMOS) is a web-capable solution that presents multiple types of data, such as a website that uses text, sound and images (still and moving) to communicate ideas and information.

In Unit 3, Outcome 2, which was covered throughout Chapter 2 and Chapter 3, you devised a hypothesis, gathered data relevant to it and concluded whether your hypothesis was supported or refuted. You provided data sets and information to justify your conclusion and supplied the specifications for creating your information. You referenced your data sources, gave evidence of validation and manipulation of data and showed evidence of methods used to secure your information. You also provided your project plan in the form of one or more Gantt charts.

In Unit 4, Outcome 1, you will create an effective MMOS to present the findings from the hypothesis you researched in Unit 3, Outcome 2. Unit 4, Outcome 1 makes up the second half of the SAT. In your MMOS, you must communicate and substantiate the conclusion you reached about whether your hypothesis was supported or refuted.

To emphasise: although your MMOS does not need to be published online, it must be *capable* of being put online and it must be created using multimodal authoring software that fulfils the criteria set out by VCAA in the list below. You are not required to write programming code. However, you must use a multimodal authoring software that provides the ability to:

- import/enter, edit and format content
- structure screen layout
- create links
- provide navigation
- create buttons
- create tags (alt tags)
- incorporate images (still and moving), text and sound.

You must also apply each of the functions listed above to your solution, in ways that are appropriate to the solution.

Your MMOS should use an effective design and communicate the findings from your hypothesis and your complex data sets clearly. It should be suitable for a global audience and you should also use appropriate software tools. Other than a website, types of MMOSs that you might use could potentially include slideshows or picture books with pictures and text; webpages with video, text, animation and music; and videos that include, music, spoken word and text, providing that the authoring software you use meets the requirements set out above and you apply each of its functions.

During the development of your MMOS, you will validate input data, test it and evaluate its effectiveness at presenting both your conclusion and your findings. You will apply your project plan to monitoring and recording your ongoing progress, and also assess the effectiveness of your project management strategy.

This chapter will begin with a discussion about the characteristics of an effective MMOS to help you understand how to deliver a solution that meets the requirements of the Outcome.

We will follow on by covering the formats and conventions, and then design principles relevant to your solution. Next, we will discuss ways of generating design ideas and design tools you can use in the development of your MMOS.

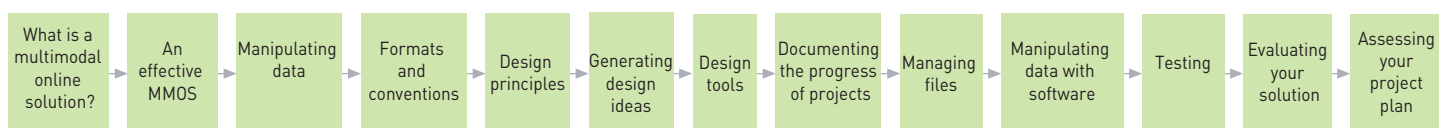


FIGURE 4.1 Chapter map

The concluding parts of the chapter discuss project management, file management, and testing your solution, followed by the evaluation of your solution and assessment of your project plan itself.

To begin, you need to gain an understanding of what constitutes an effective MMOS.

An effective MMOS



FIGURE 4.2 What constitutes an effective MMOS?

Educating a global audience

In some respects, the way that you **communicate your message** can be as important as the message itself. As you develop your MMOS, you need to provide information that is suitable for a worldwide audience. This means that your content should:

- be gender inclusive
- be culturally inclusive
- apply common language
- be age appropriate.

The main thing is that you do not want to make anyone feel excluded from your audience or offended by your content, particularly on the basis of their race, ethnicity, nationality, language, culture, religion, attitudes, beliefs, customs, gender, sexual preference, gender identity, appearance, dietary preferences, disability, illness, marital status, age, income, or education level.

That list may seem long and you may be concerned that it is going to be quite difficult not to exclude or offend anyone. Do not feel discouraged. Just be considerate and sensitive. To a degree you will already have an understanding of how to be sensitive when you communicate with others. All you need to do is broaden that understanding and apply it here. Think about the times when you have had to be tactful in your daily life, both at home and at school.

Also consider the times when you have wished that others could be more tactful or sensitive with you or with someone else.

Gender

Your solution should be written with the assumption that your audience will be composed of females and males, as well as potentially intersex and transgender individuals. Excluding anyone from your solution on the basis of their sex (female, male, intersex), sexual orientation or gender identity (woman, man, transgender) would be unwise – and insulting.

Sexist language expresses bias in favour of one sex and treats another in a discriminatory manner. In most cases, sexist language benefits males at the expense of females. Using sexist language is considered insulting and patronising because it reinforces demeaning stereotypes and traditional gender roles.

It is not difficult for you to avoid using sexist language in your solution. Job titles that include a suffix such as ‘man’ or ‘ess’ to denote that each is a traditionally masculine or feminine role, respectively, are considered examples of sexist language, because they reinforce such stereotypes. Instead of ‘police man’, use ‘police officer’. Instead of ‘stewardess’, use ‘flight attendant’. Reinforce gender neutrality instead of adhering to old stereotypes in your solution.

TABLE 4.1 Reasonable gender-neutral substitutions

Original	Gender-neutral substitution
Mankind	Humanity, humankind
Manpower	Workers, personnel
The man in the street	The average citizen, ordinary people
Sportsmanlike	Fair, sporting
Policeman, fireman, mailman, fisherman	Police officer, firefighter, postie/postal worker, angler
Actress, aviatrix	Actor, aviator
Man-hours	Worker-hours
Spokesman	Representative

GENDER COMPLICATIONS

From 1 August 2013, it became unlawful to discriminate against a person on the basis of sexual orientation, gender identity or intersex status under Federal law. Your solution, when being gender inclusive, should also keep this in mind. The Australian Human Rights Commission explains:

Sex: a person’s biological characteristics. A person’s sex is usually described as being male or female. Some people may not be exclusively male or female (‘intersex’). Some people identify as neither male nor female.

Gender: the way in which a person identifies or expresses their masculine or feminine characteristics. Gender is generally understood as a social and cultural construction. A person’s gender identity or gender expression is not always exclusively male or female and may or may not correspond to their sex.

Gender identity: a person’s deeply held internal and individual sense of gender.





Intersex: people who have genetic, hormonal or physical characteristics that are not exclusively 'male' or 'female'. A person who is intersex may identify as male, female, intersex or as being of indeterminate sex.

- Refer to transgender people using only the pronoun of their identified gender, regardless of their biological sex.
- Never use the pronoun 'it' for people whose identified gender and/or biological sex is unknown. If you must use a pronoun, use 'they'.
- The most commonly accepted terms are now 'gay' and 'lesbian' for homosexual people.

If you find yourself needing to use personal pronouns a great deal throughout your writing, and you need to maintain an even tone, you have two options. One is to use the third person plural pronoun, 'they'. This will allow you to protect the privacy of any respondents or subjects on whom you are reporting (including their gender), without dehumanising anyone with the universally unpopular 'it'.

Your other option is to use the correct pronoun on a case-by-case basis. If you need to discuss a particular interview subject's data somewhere quite specific in your solution, then you could refer to him or her as him or her, as appropriate.

Culture

Think of your MMOS as an environment or a space that your audience enters. In that environment, your audience needs to feel respected, welcomed and safe. In substantiating the results of your hypothesis, and justifying your conclusion, you should not harass or unfairly criticise anyone.

For example, using the paleo diet hypothesis from Chapter 2 and Chapter 3:

Do increasing numbers of people who live in Richmond eat a paleo diet because they believe it will give them health benefits?

It would be quite easy to turn the result of this hypothesis into something exclusive and unfriendly, regardless of whether it was supported or refuted, if you had strong feelings about dietary preferences or if you disagreed with what your respondents and subjects thought. As a researcher, you are not being asked to share your personal feelings. It is important for you to remain objective. One of the best ways to help your audience feel safe and included is to maintain your objectivity so that you share only the truth of your findings.

Culture is made up of the shared ideas, customs and behaviours of a particular group of people in society, which distinguish it from another group in society. Cultures often create a sense of belonging and fellowship by excluding outsiders. The shared ideas, customs and behaviours of a culture can be quite expansive: from religious beliefs and shared sporting teams to dietary preferences and country of birth. Ultimately, your MMOS should not exclude anyone on the basis of their membership in a culture: whether they are Muslim, a Geelong Cats fan, vegan or born in India.

Images

Your MMOS will most likely contain images and video, and you need to be conscious of the possible effect they may have on your audience. If you are concerned that you need to use an image that may identify a child or a subject, or that could contain potentially offensive content, you could use pixelation (Figure 4.3).

You must be sensitive when it comes to religion. Your global audience is diverse on a religious level. Aboriginal or Torres Strait Islander people may find it distressing if a website includes images or names of deceased people. The portrayal of God in human form is forbidden by Jewish law. Representations of the prophet Mohammed can deeply disturb Muslims. Mocking or blasphemous images or cartoons of different cultures or religious leaders must be avoided.

You should avoid appropriating another culture's symbols, such as Native American headdresses, or reproducing another nation's flag, which could be seen as being particularly disrespectful. A nation's flag is an important symbol of that nation. Reproducing symbols that represent oppressive regimes, such as the Nazi swastika, could cause anger and distress.

Other factors that you need to be careful about include sexuality and violence. Some people are affronted by images of non-heterosexuality, sexual innuendo or full or partial nudity. Depictions of bloodshed, torture, injuries or violence could also cause distress.

Language

As the author of a MMOS for a global audience, keep in mind that members of a different culture might not understand your vocabulary or references if they are embedded in your culture, but not in your audience's.

What constitutes swearing can be culturally specific. What seems harmless to you (such as 'OMG') can be blasphemous to others. Humour is another ingredient best left out of a global presentation. Humour varies greatly from person to person, even within the same small cultural group, and what you find amusing may be considered inappropriate to others.

If you need to discuss another culture in your MMOS, be aware of potentially sensitive issues and always be cautious. If in doubt, report only the facts and omit your opinion. For example, if reporting information from the paleo hypothesis:

- ✗ 21% of respondents admitted trying a paleo diet because a TV chef conned them into it.
- ✓ 21% of respondents reported trying a paleo diet in response to TV or other promotions.

In the first example above, the language shows clearly through words such as 'admitted', 'TV chef' and 'conned' that the writer has an attitude towards the respondents – that they are gullible. More than that, the word 'admitted' has the connotation of being a confession, which suggests that the writer feels the respondents have done something *wrong*. Any members of the paleo culture – people who eat a paleo diet themselves – would feel excluded by that language. However, the second example uses more neutral language, 'reported', 'in response' and 'TV or other', which omits the writer's personal feelings.

If you need to discuss other cultures in your MMOS, be aware of potentially sensitive issues and always be considerate. Report the facts – not your opinion.

Rather than using euphemisms in your writing to avoid shocking or offending people, use correct formal language. As a researcher you need to present the facts, so it is not appropriate to use euphemisms, such as 'passed away' to mean 'died', 'expecting' to mean 'pregnant', or 'facilities' to mean 'toilet'.



FIGURE 4.3 Pixelation may be used to protect the identity of a child or a subject, or may be used to hide parts of an image that people may find offensive or distressing

Commonality of language

English is often considered a global language, but some of your global audience may not be as fluent in English as you are. Some speakers of English in your audience will not be native speakers, and some speakers may speak American English, British English or other dialects of English. Your choice of vocabulary should take this into consideration and be easy to understand.

- Replace an obscure word with a simpler word or phrase that means the same thing, but try not to sound **condescending**.
- Keep sentences short, with one main idea per sentence.
- Punctuate clearly. Obey conventions.
- Have someone proofread your work to find spelling errors, culturally bound ideas or vocabulary that may be confusing or incomprehensible, offensive expression, lack of clarity, repetition or ambiguity.
- Be careful using colloquial language – the informal language used in conversation with peers within a culture. Although it seems natural to you, you may not recognise that it may seem unintelligible or misleading to other cultures.

Misunderstandings between speakers of different languages may end in confusion or worse.

Even native English speakers understand some words to have different meanings.



FIGURE 4.4 A 'dairy' in New Zealand is what we call a 'milk bar' in Australia

For example, to an American, a 'bum-bag' would be a sleeping bag for a homeless person, but to an Australian, it is a piece of apparel. To a New Zealander, a 'dairy' is a milk bar (Figure 4.4), but to an Australian, it is a place where cows are milked or where milk is processed. A car that runs on gas in Australia uses LPG, but in the USA, 'gas' is petrol. Scallops and flake in New South Wales mean potato cakes and a chocolate bar. In Victoria, scallops mean shellfish and flake means shark fillets dipped in batter and deep-fried. Australians know thongs as casual summer footwear; Americans only know thongs as skimpy underwear. In Australia, suspenders hold up stockings while in America, suspenders hold up trousers. If you have been to the USA, you will know that you enter a building at the first floor while in Australia we enter a building at the ground floor. To top it off, being caught outside in your pants is quite funny in British comedy shows

because to the British 'pants' are underwear. Australians do not see what is so funny because we consider pants to be trousers or jeans.

In your MMOS, you might not be aware that you are using terms that global readers may misunderstand. Getting feedback from someone else may highlight words, expressions or references that you might need to change or explain.

Age appropriateness

Try to make your writing as accessible as possible to readers of different ages. Your global audience may include children as well as adults. You are not expected to be able to cater to everyone: it would be impossible because of the wide and unpredictable range of your audience's vocabularies, life experiences, comprehension skills and areas of interest. However, you can make an effort to write for a general audience and be aware of techniques you can use to cater to some age-related needs.

You can calculate an estimate of the age or education level a reader would need to understand your writing. Microsoft Word can give you readability statistics of your document. Click the round Office button (or File, in older versions) > Word Options > Proofing > Show Readability Statistics (Figure 4.5).

When you perform a spell check, Microsoft Word shows a rating of how readable the document is. Two of the readability measures, 'Flesch Reading Ease' and 'Flesch-Kincaid Grade Level', are calculated according to the number of syllables and words in a sentence; essentially, the shorter your sentences are, and the fewer the syllables, the easier they are to read, especially for younger audiences. Generally, aim for a higher Flesch Reading Ease number and a lower Flesch-Kincaid number.

The passive sentences percentage is included as a measure of readability because passive voice is less effective and clear in writing than active voice. The higher the passive sentence percentage, the less readable your document is to most of your audience – young and old, native English speaker or not. The active voice is more effective in writing because it is simple and direct.



FIGURE 4.5 Readability statistics for the sentence 'Paleo dieters eat meat, fish, vegetables and fruit.'

- ✓ ACTIVE: Paleo dieters eat meat, fish, vegetables and fruit.
- ✗ PASSIVE: Meat, fish, vegetables and fruit are eaten by paleo dieters.

TABLE 4.2 Break down of active and passive voice in simple sentences

	Subject	Verb	Object
Active	Paleo dieters	eat	meat, fish, vegetables and fruit.
	<i>Agent (the doer)</i>	<i>The action (doing)</i>	<i>Affected (done to)</i>
Passive	Meat, fish, vegetables and fruit	are eaten	by paleo dieters.
	<i>Affected (done to)</i>	<i>The action (doing)</i>	<i>Agent (the doer)</i>
Active	The paleo diet	excludes	dairy, grains, refined sugar and processed foods.
	<i>Agent (the doer)</i>	<i>The action (doing)</i>	<i>Affected (done to)</i>
Passive	Dairy, grains, refined sugar and processed foods	are excluded	by the paleo diet.
	<i>Affected (done to)</i>	<i>The action (doing)</i>	<i>Agent (the doer)</i>
Active	The majority of respondents (56%)	preferred	the non-dairy aspect of the paleo diet.
	<i>Agent (the doer)</i>	<i>The action (doing)</i>	<i>Affected (done to)</i>
Passive	The non-dairy aspect of the paleo diet	was preferred	by the majority of respondents (56%).
	<i>Affected (done to)</i>	<i>The action (doing)</i>	<i>Agent (the doer)</i>

Writing for younger readers

- Create an engaging discussion.
- Use a simple sentence structure.
- Make clear points. Younger children are literal-minded and do not yet reason abstractly or understand irony, sarcasm, analogies, ethical dilemmas, implied meaning or metaphor. They may not detect subtleties.
- Choose from a suitable vocabulary, but do not fear introducing new, useful words.
- Explain any technical terms in simple language.

- Be careful of discussing adult themes, such as sex, violence, drugs, alcohol or death.
- Consider including illustrations for younger readers. Slightly older readers may appreciate simple line illustrations – or none at all.
- Do not be patronising. Younger readers are inexperienced, not stupid.

Writing for senior readers

Writing for elderly/older/senior readers also has its challenges, but it is unfair to generalise about senior people's technological expertise. For every older person who asks for larger text on-screen or simpler instructions, there is another older person who can move swiftly using regular equipment and eagerly embraces new technologies.

In general, on a technical level, be sure that published text is easy to read for those whose eyesight may not be as sharp as younger readers, and is relatively easy to physically navigate. Authors writing for older readers should remember the following.

- Try not to use too many similar colours, nor to obscure text by background colour or images. As eyesight declines, shades of blue can appear to be faded or de-saturated.
- While audio is not usually critical to interacting with a solution, be sure to remove background noise if you are recording your own evidence.
- Always test your solution using screen readers, which convert text into speech for people who are vision-impaired.
- Provide subtitles when video is essential to your information.
- Do not put controls (such as buttons) too close to each other. Make controls a generous size. Include captions that describe their function.
- Explain technical terms. Use your judgement about whether you think you should explain simpler terms or functions that may be new to senior users; for example, 'Click the mouse' or 'Skype'.
- During long tasks, give clear feedback on progress and reminders of goals.

THINK ABOUT INFORMATICS 4.1

Research 'W3C accessibility' and summarise what should be done to make a website accessible.

Accessibility

When developing your online solution, consider how you can communicate your message more effectively to people with special needs. You should avoid making communication unnecessarily difficult for people with special needs, including the very young, the elderly and those with disabilities, such as low vision, colour blindness, limited mobility, limited language skills or computer skills.

Some of the measures you can take have already been listed in the 'Writing for younger readers' and 'Writing for senior readers' sections above. Other suggestions are described in Figure 4.6.

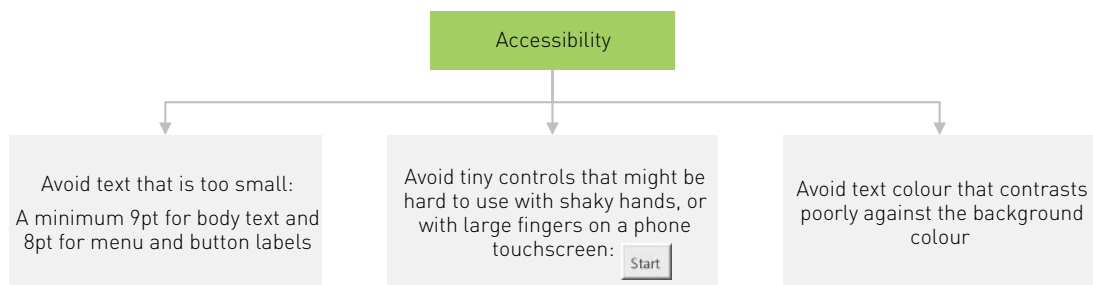


FIGURE 4.6 Suggestions for accessibility; also refer to design principles

Avoid using text that is too small. As a guideline, use a minimum of 9pt for body text and 8pt for menu and button labels – larger, if possible. Avoid using very small controls that might be hard to click easily with shaky hands, or with large fingers on the small touchscreen of a smartphone. Avoid text that contrasts poorly against the background colour.

Refer also to the design principles for accessibility, error tolerance, ease of use and navigation later in the chapter.

Clarity

Whether spoken or written, your solution's language should be clear enough for most audience members to understand. The primary goal of any solution is to convey information to the audience. If you do not convey why you concluded that your hypothesis was supported or refuted and justify this using multiple types of data, your solution will fail. Try to state your goals at the outset of each section.

The advice previously given regarding explaining technical terms and using direct language especially applies here. This is especially true if you cannot avoid using **jargon**. Although it is a good idea to use short sentences and paragraphs, you cannot write completely in short paragraphs, short sentences and dot points. A solution written entirely in dot points and one or two sentence paragraphs will not be very engaging because it will not draw the audience *in*.

You will need to use proper paragraph structure here and there to develop an argument, sustain a narrative and **engage** your audience's interest. Use straightforward phrasing. If you are concerned about needing to use longer paragraphs and are worried about long blocks of text, break them up with subheadings, tables or relevant illustrations as needed. If you need to use a long sentence that worries you, look at the punctuation: can it be broken into multiple sentences? If not, can it be framed with shorter sentences on each side so that it is easier to read? If you are still uneasy, revisit the sentence and ask yourself: is it the right sentence?

**THINK ABOUT
INFORMATICS 4.2**
How could you test your
solution for clarity?

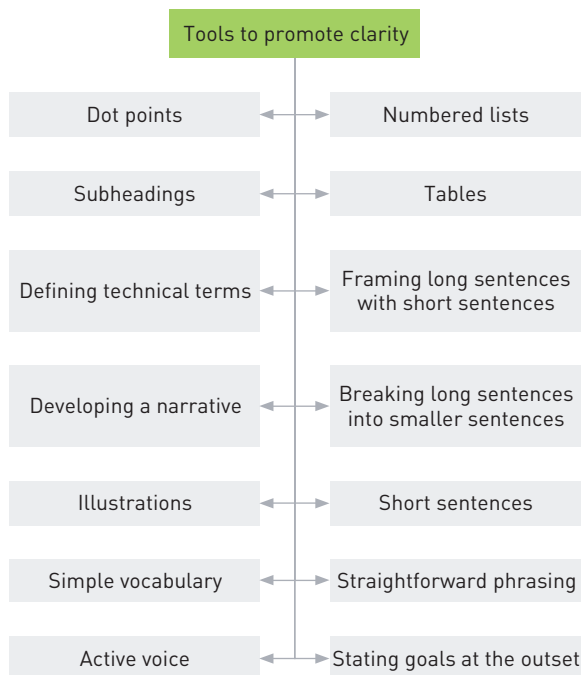


FIGURE 4.7 Tools to promote clarity

Readability

Text, audio and visual media all must be of a size, colour, volume, speed, illumination and fluency that can be interpreted with ease. This overlaps with previous discussion regarding accessibility and educating a global audience. The following section discusses two particular elements of readability: typefaces and white space.

Typefaces

Use a plain, legible typeface for body text. Use either **serif** (such as Times New Roman) or **sans-serif** (such as Arial). Decorative, script and handwriting typefaces are not recommended. If you must use a decorative, script or handwriting typeface, reserve it for headings only.

The slab font 'Courier New' is popular in programming and used throughout Informatics Units 3 & 4.

→ Paleo diet interview results

Serifs make each letter more distinctively different when being read at speed, while sans-serif letters like capital letter I, lowercase l and digit 1 are hard to tell apart. This paragraph is in a sans-serif typeface.

Do not mix more than three typefaces on the same screen, or your solution will begin to look untidy, as if you have thrown it together haphazardly rather than designed it carefully.

White space

White space is a section of a graphic representation that is empty of any colour or object, which is used to create a clean, uncluttered look and is not considered wasted space by designers. In this instance, it refers to the empty parts of the screen that can be used to aid readability.

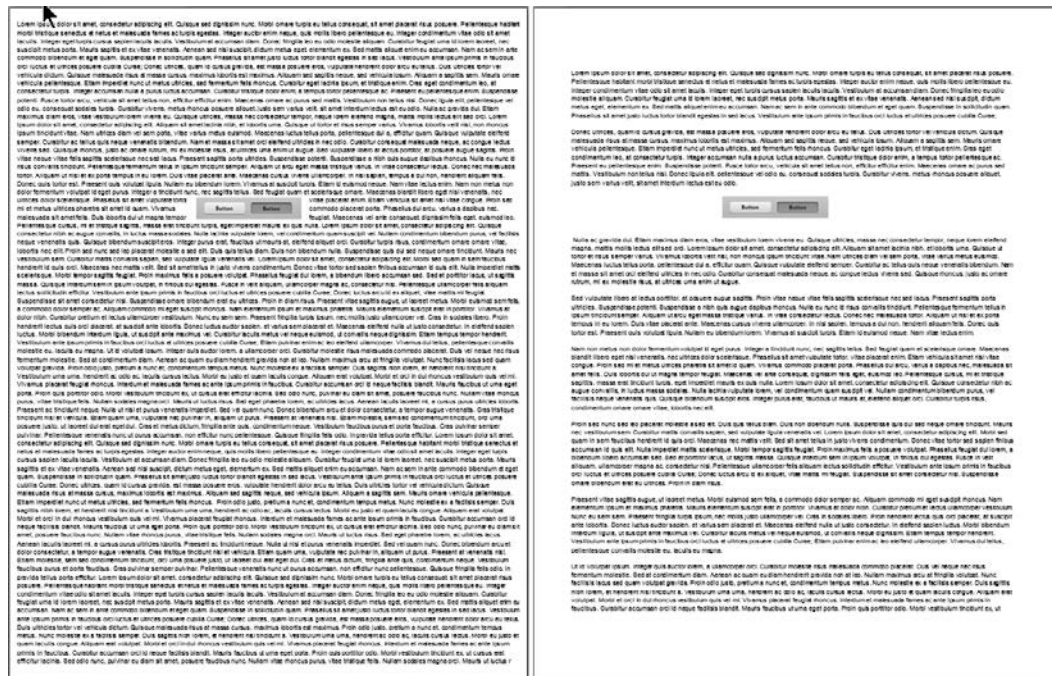


FIGURE 4.8 How white space affects readability: Poor use of white space on the left and generous white space on the right.

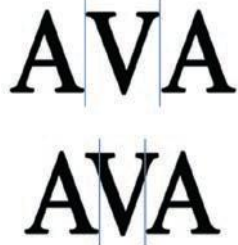


FIGURE 4.9 The top line has no kerning and the 'A' does not share any of the space occupied by the 'V'. With kerning, the A and V share space. The space between the lines is leading.

The important areas of white space are:

- margins at the edges of the page or screen
- visual gaps separating paragraphs and objects (Figure 4.8)
- kerning, which refers to the spacing between pairs of letters. Certain letterforms need space between them adjusted when they are used together to avoid unsightly gaps, such as T and L.
- leading (pronounced 'ledging') or line spacing, which is the distance between two lines of type. Leading affects text legibility. It is set so the eye can flow easily from one line of text to the next. If leading is too tight, text is too close together. If leading is too loose, text is too far apart.

Relevance

Regardless of your solution's topic, make sure you do not veer off onto another topic. Your audience is there for the stated message – your Outcome. They will have little tolerance for off-topic information. Only include material that relates directly to your hypothesis.

Accuracy

Your multimodal solution aims to educate your audience. Make sure that the information you provide is correct based on the conclusion you have drawn from investigating your hypothesis. Any data sources you used during Unit 3, Outcome 2 should have been reputable sources. Remember: Wikipedia is not an acceptable source of data. Any additional facts you cite in your MMOS must be checked in at least two places. Providing incorrect, outdated or misleading information is a violation of the trust an audience puts in an author.

Useability

As a writer of your MMOS, you face many difficulties in reaching global audiences. There is such variety in the hardware and software used by audiences that it may be difficult to create a single solution that is equally usable by everyone. There are several potential factors to consider, including plugins, browsers and hardware.

Plug-ins

Plug-ins are software modules added to applications to enhance their functionality, such as browser plugins to block pop-ups or display PDF documents.

If you use a Flash animation to convey vital information, but some of your audience does not have a Flash player installed, your message cannot be communicated. To prevent this problem, there are other options.

- Produce one version of the site using the plug-in, and another version without it. This doubles your development and maintenance effort.
- Produce a very basic site that every browser can render properly.
- Ignore audience members who lack the plug-in.
- Produce a site that uses a different plug-in that is less likely to be unsupported by a wide range of platforms, such as JavaScript. This still may ignore some audience members who lack the plug-in.

The decision is not easy to make, and such problems are ongoing because of changing technology. The most you can do is make smart choices based on the current technology of your likely audience.

Browsers

The World Wide Web Consortium (W3C) publishes standards that websites and browsers are supposed to obey to be compatible with each other. However, no browser has perfectly achieved this aim and some sites do not work well in some browsers.

As the developer of a MMOS, you cannot know how your audience has configured browsers and devices. Are pop-up windows disabled? Are ads blocked? Is sound muted? Is JavaScript blocked? Is image loading turned off? Are cookies disabled? Any of these can cause your MMOS to perform less adequately than expected.

Try to ensure your MMOS is usable for as many audience members as possible.

- Test your MMOS in all the dominant browsers, including the current versions and the most recent old versions.
- Using online services that perform compatibility checks of your MMOS.



Another nail in Adobe Flash's coffin – Chrome to block Flash ads



Amazon puts another nail in Flash's coffin



Final nail in the coffin for Flash?

THINK ABOUT INFORMATICS 4.3

Adobe Flash used to be the dominant media technology for website animation. Why has it been losing support from major technology companies, such as Apple and Google?



World Wide Web Consortium (W3C)

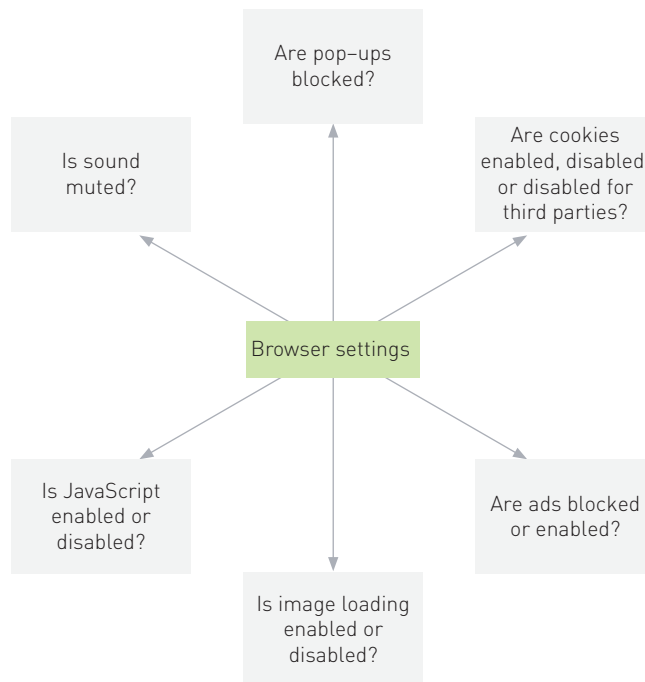


FIGURE 4.10 Browser settings that will differ throughout your audience. Before you finish your MMOS, you will not know what browser settings your audience will have chosen, but these are factors to keep in mind.

THINK ABOUT INFORMATICS 4.4

- 1 Investigate how an average webpage can be made viewable on 22-inch HD monitors, tablets and phone screens.
- 2 Write a summary of the information you have obtained, citing at least two sources.

- Try to anticipate common problems, such as disabled Flash or cookies, and provide backup methods to convey information. If audience members set images not to auto-load, use alt text to describe missing images. Provide email contact links in case feedback forms do not work.

Hardware

A web developer building a page for a 1920 × 1080-pixel monitor may be disappointed to see how unattractive and unusable the page renders on the 4-inch screen of a smartphone (Figure 4.11). Catering equally for all browsing devices is difficult. Not everyone has the time, skill and budget needed to create and maintain both mobile and desktop versions of websites, or write the JavaScript that can cope with browser differences. However, you should try to cater for the most popular current devices – smartphone, tablet and desktop – so you can communicate your message to as large an audience as possible.

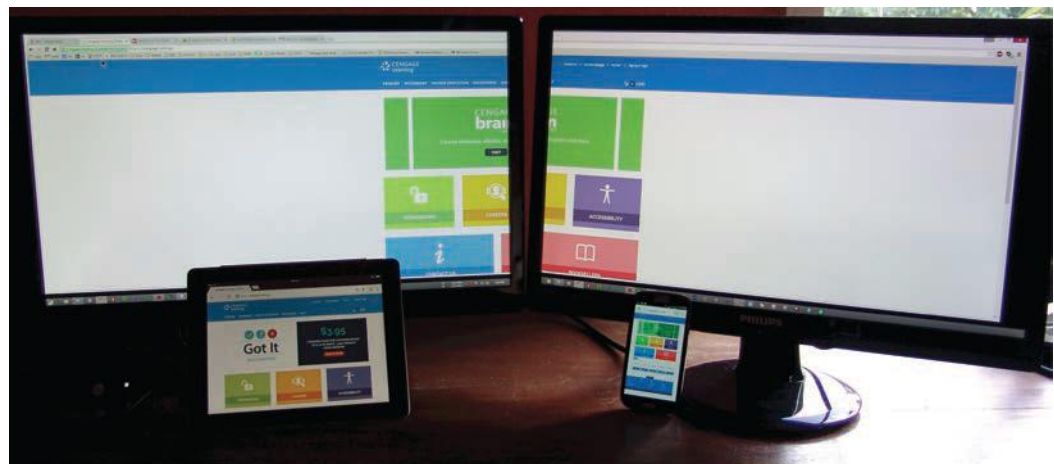


FIGURE 4.11 The same site viewed on three very different displays: dual HD monitors, a 10-inch tablet and a 5-inch smartphone

Media

Avoid media that are only playable on certain platforms, such as Mac OS, or with certain players, such as Windows Media Player. Media not supported by major technology suppliers, such as Flash, should also be avoided. Restrict yourself to standard media types (Table 4.3).

TABLE 4.3 Standard recommended media formats

Media usage	Recommended format
Images	JPG for photos GIF or PNG line art and logos
Audio	MP3 While WAV gives perfect quality because it is uncompressed, it is 10 times larger than MP3
Video	MP4, DivX or FLV

Because of the sheer amount of information in media files, most media file formats are compressed. Some use *lossy* compression (such as JPG, MP3, MP4), which reduces file size by removing detail and reducing quality. The user can usually choose how much compression to apply: a little compression is often unnoticeable. The other compression method is *lossless* (GIF and TIFF images, FLAC music, Zip/RAR archives), which reduces file size by optimising data encoding to reduce repetition.

Not all computers, phones, tablets or browsers can read all media formats. Some players (such as Videolan's VLC player) can play almost anything; others will play only some media. Some older software was created before formats like APE, FLAC or MKV were invented, and cannot decode them. Be conservative in your media format choices.

Timeliness

Your MMOS should be media-rich, and multimodal files can be large and slow to load. For it to be timely, there should be no significant delays in retrieving the information you have provided. However, users are known to be impatient. They are likely to leave if each page or screen takes more than a few seconds to load.

It is difficult to make a media-rich online solution load quickly for users with slow internet connections. You must always balance media size against acceptable quality levels in various ways.

- Use small thumbnail images linked to full-size images. Let the user choose the images in which they wish to invest their time and bandwidth.
- Offer smaller (quicker to download) and larger (higher resolution) versions of media. Let users choose: do not force them to take *your* favourite.
- Split a large webpage into an index page linked to several smaller, quick-loading subpages.
- Reduce audio file size by reducing sampling rate, such as from 48Khz to 44.1KHz, or number of channels, such as from 7.1 channels down to stereo.
- Recorded speech has little audio range, so it can be compressed even more than music with no perceived loss of quality. It can often also be downmixed from stereo to monaural to halve the file size.
- Adjust the compression level for JPG pictures. The original image might look lovely with no compression, but may still be acceptable with compression turned up quite high.

As you develop your solution with a fast local hard disk or network, the large video that loads and plays instantly on your computer must be slowly and expensively downloaded by everybody else when it is put online.

Another meaning of timeliness for the purposes of your MMOS refers to the actual currency of your information. Refer to the 'Integrity of data' section in Chapter 2.

Make sure the information in your MMOS is still current (not outdated) otherwise your conclusions will be meaningless.

Completeness

Completeness means that the information you are presenting in your MMOS is just that: complete. You need to ensure you present all of your findings to substantiate the conclusion you reached in Unit 3, Outcome 2. If you do not do this, you will not be able to convince your audience successfully that your hypothesis was genuinely supported or refuted. Although you have a degree of latitude in how you present the information you have obtained from gathering and processing your data, you still must present all of it so that it can be considered properly. You can use some of the tactics referred to in the 'Integrity of data' section in Chapter 2 as a guide if you find yourself struggling.

Attractiveness

The interface is the only part of your solution that users will see. It must be appealing, attractive, and easy to use, regardless of how brilliant your information architecture is.

- Give your solution a consistent look and behaviour. Create your own template, as discussed in 'Heading styles and CSS', as a starting point for pages.
- Make interfaces clean, simple, easy to learn, easy to use and attractive to look at for hours on end. Limit the numbers of colours and typefaces used.
- Avoid showing off. No one will be impressed that you just discovered animated GIFs or have 200 typefaces to choose from – and that you are determined to use every one of them in your solution.



FIGURE 4.12 A vision-impaired user relies on the computer's speech synthesis to read the onscreen out loud

- Use subtle colours. Neon or overly bright colours will be difficult to tolerate for any length of time.
- Employ leading, kerning and white space to make text more readable.
- You should generally obey conventions. Doing things your own way just to be clever or different, without a substantive reason, will only annoy your audience.

A poor interface is confusing, hard to learn, will cause user errors. It will discourage your audience from staying long enough to receive the message you aim to communicate.

- Consider accessibility needs, as described in the accessibility section earlier in the chapter (Figure 4.12). For example, attach alt text to images so vision impaired users with screen readers can hear what the images represent.

- Anticipate common user errors and cope with them tolerantly, such as always providing a 'Back' or 'Cancel' facility so users are not locked into a path from which they cannot escape. This relates to design principles > functionality > accessibility > error tolerance.
- Be informative. Provide indications of how well a long operation is proceeding, such as how much has been downloaded or calculated so far, so that users do not think the computer is frozen or the website has crashed.

For discussion on information architecture, refer to *Computing VCE Units 1 & 2*, Chapter 5.

- Give help for tasks that might be a challenge, such as ‘To download this video, right-click it and select ‘Save Link As’.
- Let typical end-users beta test your interface and pay attention to their feedback.
- Use the most appropriate graphical user interface (GUI) controls. For example, to force a choice of a single item from a selection, use radio buttons. To allow zero or multiple choices, use checkboxes.
- Be flexible. Your audience should be able to choose whether they use menus, shortcut keys or buttons to perform an action. Everyone has their own preference.

Manipulating data

Explore and learn the features that the multimodal authoring software you have chosen offers. It will take you a very long time and you will likely produce a subpar solution if you do not take time to become familiar with your software before you start to build.

Remember, as discussed at the beginning of the chapter: you need to choose software that allows you to perform a number of tasks, and you need to perform those tasks in your solution. Figure 4.13 restates the requirements for your authoring software.

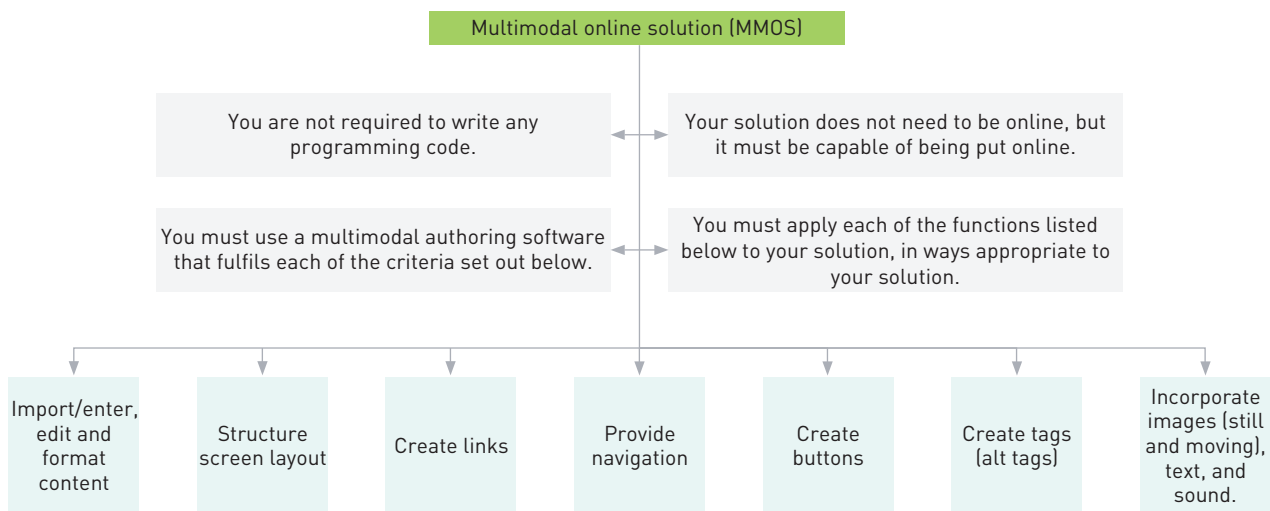


FIGURE 4.13 Requirements for the MMOS

You can use several types of solutions to satisfy these criteria, such as:

- websites, which are the classic multimodal medium, and your dominant choice for Unit 4, Outcome 1. They can be created with local software, such as Adobe Dreamweaver (expensive proprietary software), Archnophilia or KompoZer (freeware); or created online, with services like Google Sites (sites.google.com), Wikispaces or WordPress.



Archnophilia



KompoZer



Google Sites



Wikispaces



WordPress

- slideshow software applications, such as Microsoft PowerPoint, Keynote (for Mac) or Slideshow. There are services to upload and share PowerPoint presentations, such as Powershow.com and Microsoft's PowerPoint add-on Office Mix
- online slideshow presentation makers, such as Google Slides, Microsoft Sway and Prezi
- downloadable presentation makers like touchcast.com, which create online interactive videos on an iPad or a PC.

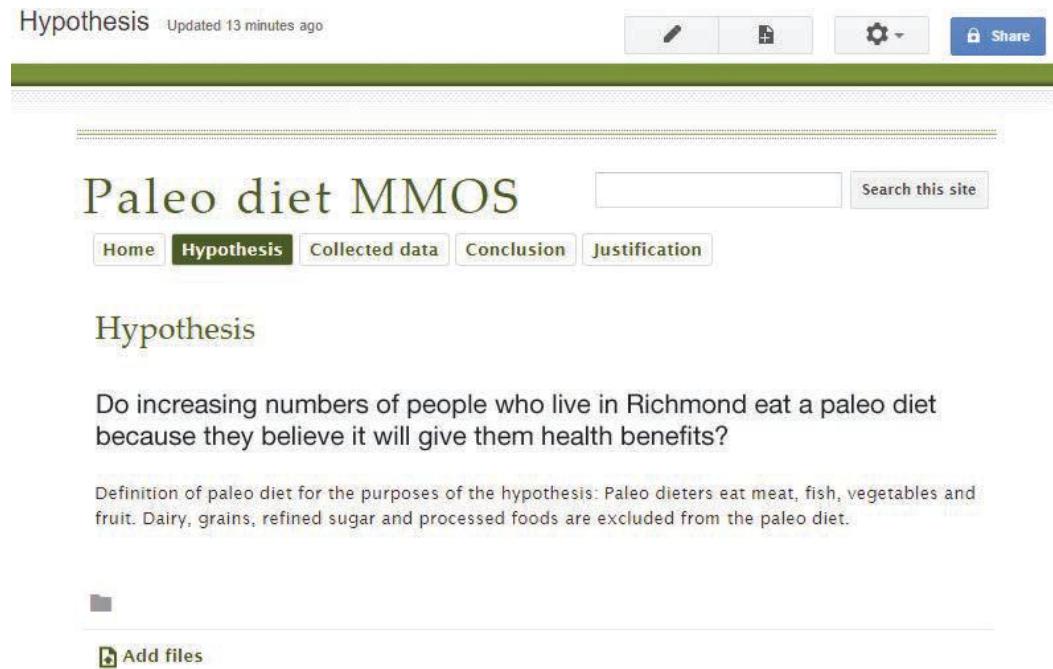


FIGURE 4.14 A website created using Google Sites

Heading styles and CSS

Both heading styles, such as those used in Microsoft Word, and cascading style sheets (CSS) in webpages allow authors to describe formatting styles once, use the style many times and change styles easily.



FIGURE 4.15 Heading styles in Microsoft Word showing some in-built styles (e.g. Heading 1) and some custom-made styles (e.g. code)

In a large document or website with many pages, headings and subheadings, it is easy to change your mind about formatting as the document grows. Going back to each piece of text and reformatting its text size, colour, alignment, typeface and style is exhausting. However, with heading styles or CSS you can redefine any heading or text style throughout the document instantly.

Heading styles also let you automatically create a table of contents (TOC) or an index with little effort. The TOC appears instantly and can easily be updated as sections are moved, added or removed and page numbers change.

Use the 'Outline View' in Microsoft Word to collapse body text so only headings are visible, so you can see and reorganise the structure of large documents by dragging section headings to new locations.

Cascading style sheets, like heading styles, format information that is defined once and used in many places. When the CSS changes, the appearance of every webpage using that information also changes. It makes site maintenance a breeze, it is not very hard to learn, and it promotes a consistent appearance for all pages in a website.

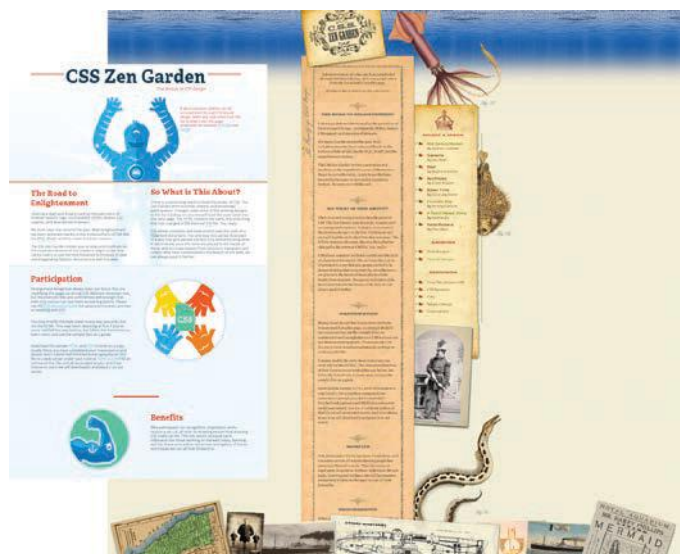


FIGURE 4.16 Two CSS definitions produce vastly different websites using the same HTML file

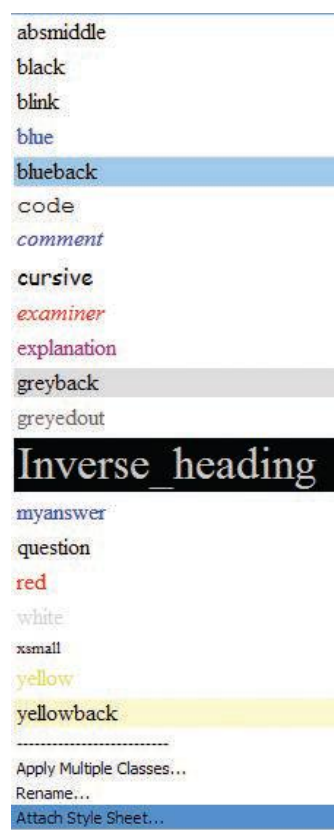


FIGURE 4.17 CSS styles defined in Adobe Dreamweaver's web editor

Customise your tools

Straight out of the box, multimodal authoring software is set up to suit the average user. However, a power user will customise tools to maximise efficiency. Most sophisticated software will allow you to add and change keyboard shortcuts, menus and default values, such as typefaces and line spacing, to suit your preferences.

Learn your tools

Spend some time browsing the submenus of your favourite software. You would be pleasantly surprised by the power and assistance that is available. Microsoft once surveyed users of Word, asking what features they would like to add: 90% of the requested features were already in the program, but users had not found them.

Templates

It is common to create many documents that follow the same pattern – using, for example, the same logo, links, margins, colour schemes, header/footer and copyright statement. Instead of re-creating all those settings for every new document, create a blank template document with those settings already entered, and save it. Whenever you need a document of a similar type, load the template document, add the new content and save with a new name. This saves time, ensures you do not forget to add necessary components, and promotes a consistent look and feel for all of your related documents.

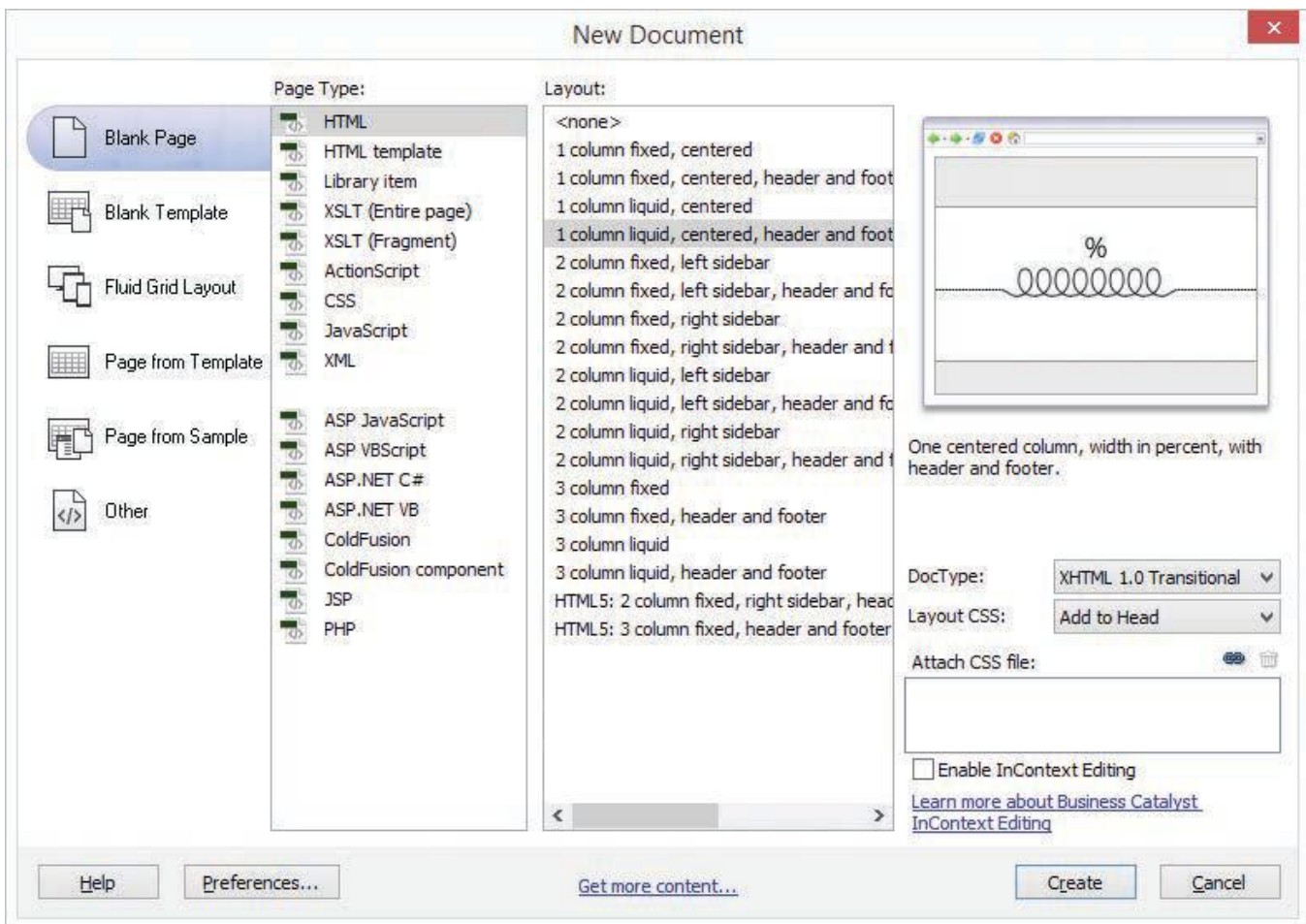


FIGURE 4.18 New document templates in Adobe Dreamweaver

A slideshow template may consist of a few special slides (a title slide, a slide for contents, one for your references and a concluding slide) plus one slide that will be the parent of all of the slides containing body text.

Pre-made templates of websites, slideshows or other MMOS documents may be available to you with pre-designed design elements. In your MMOS you should not use templates created by other people, since you will not be demonstrating your ability to choose and apply design principles.

Formats and conventions

Presenting your information effectively in a MMOS requires a knowledge of both formats and conventions.

A **format** is the form in which information is presented, such as a webpage, pie chart, text in paragraph, table, comic strip, limericks, pop song or newspaper article. There may often be several formats from which to choose, such as presenting data in either a table or a graph.

Conventions are the accepted techniques that an audience will expect to find when a format is used. Each format has its own conventions.

TABLE 4.4 Formats and conventions relevant to MMOSs

Format	Conventions for format
Pie chart	• Coloured or patterned slices • Largest slice starts at the 12 o'clock position, and slices proceed clockwise in order of decreasing size • Slices are labelled or the chart has a legend
Cartoon	• Speech bubbles have a pointed tail to the speaker's mouth • Thought bubbles have a trail of bubbles leading to the thinker's mind; other characters cannot hear these thoughts • Time moves in frames from left to right, top to bottom
Webpage	• Underlined or coloured links for URLs • Site logo acts as a link to the homepage • Landing page is named index.html, index.htm or index.php • One main topic per page • Large pictures are linked to small thumbnail versions • Image formats are JPG, GIF or PNG • A contacts page lets visitors contact the site owner • Sites have a search facility • Up to 4–5 clicks should take users to 80% of the documents they want to view • Sound should never play automatically when a page loads
Table	• Column headings are bold and in the top row • Numbers are right-justified, or centred on the decimal point • Australian currency has two decimal places • Border lines separate columns • Totals appear at the bottom row and/or rightmost column
Text	• Only centre short units of text, usually left justified • Sans-serif font for onscreen body text to ensure readability • Serif heading text for visual distinction • Consistent fonts, type sizes (usually 9–12 pts) • Lines are no longer than 60 characters (10–11 words) • Limit paragraph size (4–8 lines) • Use wide margins • Avoid use of all capitals • No underlining – use italics or bold for emphasis • Consistent heading size and styles • One space at the end of sentences

You should obey conventions because information is far quicker and easier to find and understand if it is presented in a predictable manner. You should not force your audience to learn new ways of information presentation for no reason. Audiences also expect conventions to be used. However, conventions also come in different strengths.

If you disregard **mandatory conventions**, you are breaking the law. For example, whether a country chooses to drive on the left side of the road or the right is quite arbitrary. However, once a nation has chosen a side, people who choose to drive in the incorrect lane will be punished for disregarding legal conventions.

Preferred conventions are those where most people have a distinct preference about how something should be done. For instance, in a large book, people expect to find page numbers. In reference books, readers would be upset if they found there was no index. Hyperlinks in a webpage are traditionally underlined.

Optional conventions offer you a real choice. Some individuals may prefer to see a book's page numbers centred at the bottom of the page, but would not write angry letters to the publisher if page numbers appeared at the top of the page. Similarly, some individuals prefer to see a link to the site map at the header of a MMOS rather than at the footer.

If you do not learn or do not follow conventions, you will hinder the communication of your message and possibly tarnish your credibility with your audience. When communicating, be considerate. To maximise the readability of your MMOS, and to better communicate its message, discover the conventions that apply to the format you choose to present your solution and use those conventions consistently and accurately.

Design principles

Design principles have been discussed in Chapter 1. They have also been discussed in *Computing VCE Units 1 & 2*. Design principles are factors that enhance the appearance and functionality of solutions. Online solutions need to be easily understood and accessed with minimum time and effort. To communicate effectively, your MMOS needs to be clear and functional. You need to ensure that facts are obvious and your message is unmistakable. Your solution must be carefully designed, taking into account an important set of design principles.

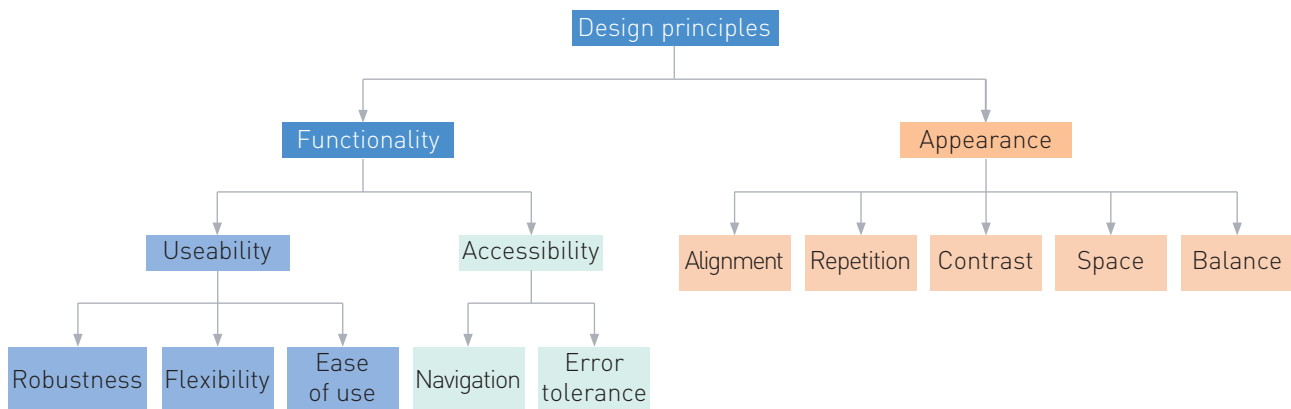


FIGURE 4.19 Design principles

If you sense that a page or screen is awkward to use, or looks odd or unattractive but cannot exactly say *why*, you are probably responding intuitively to your solution's use of design principles.

In Unit 4, Outcome 1, you are required to create two or three design ideas from which one is chosen to be developed into a fully detailed design. A design idea is a rough, general 'back of an envelope' plan of a solution that has little detail. Techniques for generating design ideas is discussed later in the chapter, beginning on page 209.

Functionality

Useability

The useability design principles are robustness, flexibility and ease of use (Table 4.5).

TABLE 4.5 Useability design principles

Principle	Description
Robustness	Your solution's ability to cope with errors during use. Your solution should resist crashes, failure and security threats, and function correctly in spite of invalid data or a stressful environment. It anticipates the threats it may face and takes action to deal with them. Robustness countermeasures include comprehensive data validation, preventing errors (such as by disabling certain buttons on an interface when they are irrelevant or harmful), and anticipating troublesome user actions (such as by choosing to print when no printer is connected, or saving to a device that is already full).
Flexibility	Your solution's ability to cope with multiple ways of performing tasks. For example, many websites were once designed with fixed-width dimensions, such as 800px. These websites only supported a single way of performing the task of viewing the website – using a browser that supported the fixed-width page size. With so many different browsers, operating systems and devices in use today, it is better to build websites that are flexible, without fixed dimensions, and instead let the user determine the page size.
Ease of use	How user-friendly your solution is. Some of the user-friendliness aspects overlap slightly with the appearance design principles. However, you should also ask questions when creating, planning and testing your solution. • Is it easy for users to intuit the design of your MMOS and perform basic tasks? • Can basic tasks be performed quickly? • Will users remember how your MMOS works when coming back after they have not visited for a long period of time?

Accessibility

The accessibility design principles are navigation and error tolerance (Table 4.6). They overlap heavily with the earlier accessibility material in this chapter, which discussed catering for people with disabilities or special needs.

TABLE 4.6 Accessibility design principles

Principle	Description
Navigation	The clarity, simplicity and intuitiveness of your solution's navigation system. You should ensure your MMOS can be navigated comfortably by multiple browsers and a touchscreen (ensure the buttons are not too small and close together, making this difficult). High-quality user interfaces are transparent, meaning that your users would not really notice the actual interface because it is so easy to use that they interact with it intuitively. The required information would be found quickly enough that users focus on it rather than the way they found it. The interface is a connection point or navigation tool rather than an end in itself, so it should be unobtrusive but clear. For example, the labels on your navigation buttons should be short and clear; every page should include a link back to the homepage and this link should always be on the same place on every page. Use linked forward and backward arrows to indicate when text continues on another page, with 'Next', and 'Back' or 'Previous' to help orientate users as to their current position. Do not hide the most important information under sub-menus or three screens down on the webpage. See Chapter 5 of <i>Computing VCE Units 1 & 2</i> for more detail on navigation.
Error tolerance	Your solution's ability to help users avoid and correct mistakes using clear instructions, and its ability to prevent them making errors in the first place by avoiding allowing them to perform actions that could lead to errors. Grey out non-selectable options. Ask for confirmation of major actions. This is also connected to robustness.

Appearance

The screen layout of your MMOS should not be unnecessarily elaborate or decorative, or contain superfluous animations. Too many buttons, bullets, icons, rulers and flashing graphics will confuse the eye and distract your audience.

Long or wide screens that require your audience to scroll should be avoided when possible, particularly if scrolling is required both across and down the screen. This makes the information harder to view. Short screens of information, with links to other screens, are more easily viewed and therefore more effective.

The design principles contributing to appearance are alignment, repetition, contrast, space and balance. They are discussed in the following section.

Alignment

The alignment of text can be left, right, centre or fully aligned (justified). Generally, choose one alignment for each page and stick to it. Left-aligned text is easier to read than centred text for paragraphs because the text begins on the left-hand side every time. The text is in a straight line and readers can follow the text with their eyes starting from the left edge. Centred text makes the eye work harder to locate the start of each line. Unlike left alignment where there is a consistent straight edge for the eye to follow, there is no consistent focus point for eyes to return to once each line is read.

The human eye can detect when an object is only a single pixel out of place vertically or horizontally compared to its neighbours. Text, images, and columns should all be aligned precisely.

Sloppy alignment looks careless and unprofessional, and ruins the impression that items are visually related to each other.

The following text is centred:

Many paleo-friendly restaurants and cafes have opened in the inner suburbs of Melbourne since 2012, such as Paleo Den, which opened in August 2014 in Chapel St, Prahran.

The following is the same text, shown right-justified:

Many paleo-friendly restaurants and cafes have opened in the inner suburbs of Melbourne since 2012, such as Paleo Den, which opened in August 2014 in Chapel St, Prahran.

Full justification looks formal and is traditional in novels and textbooks, but in magazines and newspapers, it can lead to unattractive ‘rivers’ of white space travelling down narrow columns as text is stretched out to create a straight right margin.

Headings are often shown centred.

Repetition

Your audience will be reassured by repetition in your solution. This is not repetition of content and words – rather, of design elements! Using the same logos, icons, typefaces, heading styles, colour scheme, margins, borders, menu positions and shortcut keys throughout your solution will help your audience to trust the predictability and consistency of your solution (Figure 4.20). This also ties into formats and conventions.

Essentially, your audience does not want to have to re-orient and learn the system themselves on every screen they visit, and they do not want to see a different design on every page. This is especially the case when you want them to be able to focus on the content, not the design.

REPETITION

Annotate the two pages from the W3C website in Figure 4.20 below to show how repetition has been used to achieve consistency.

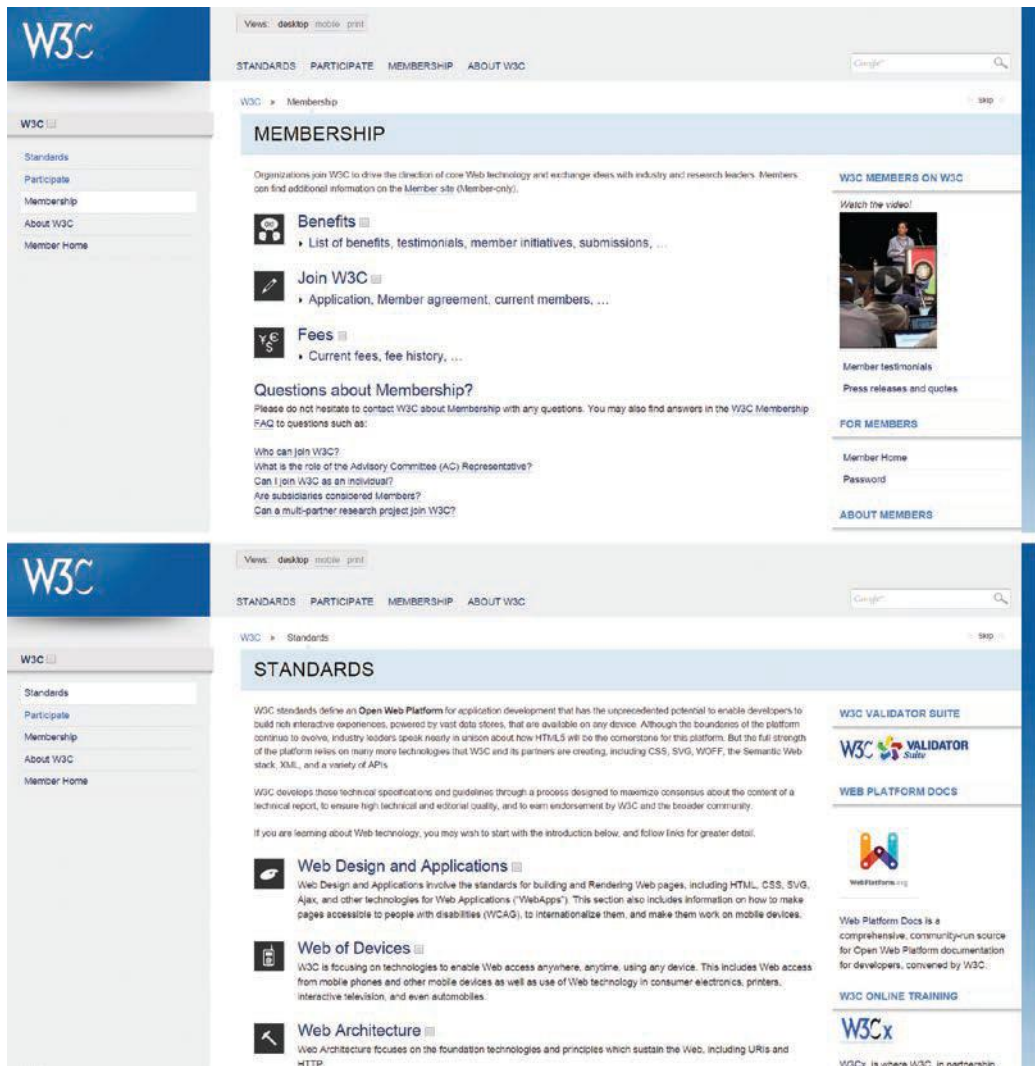


FIGURE 4.20 Two pages of the W3C website

Contrast

Contrast refers to the visual difference in colour or tone between objects (both text and images). Greater contrast will make objects appear to stand out more from one another. If there is not enough contrast between two objects, they may appear to blend into each other, making it difficult for the user to see each of them clearly. Contrast between the background and text should make the information clearly visible and legible.

Be careful about putting text on images. It may make the text unreadable. You should also avoid using red–green colour combinations. They are hard to read for everyone, and near impossible for the 10% of males who are colour-blind.

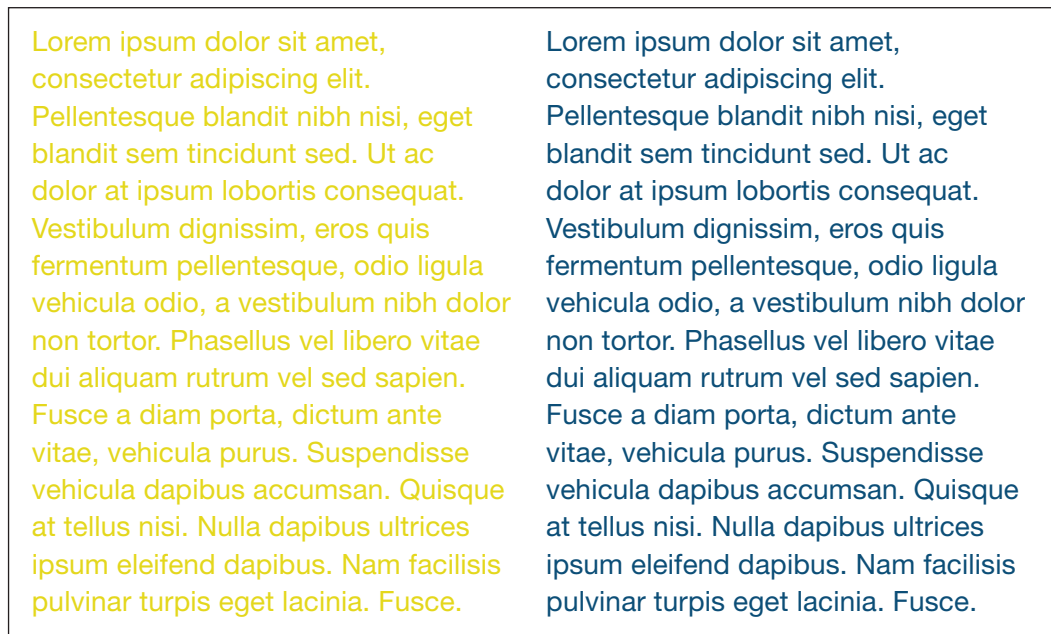


FIGURE 4.21 Avoid using light-coloured text on a white background because there is not enough contrast to make it readable

Space

Space refers to the areas around and between objects – text and images (still and moving). If your solution is cluttered it may be unpleasant to browse. You do need to include all of the information obtained for your Outcome on your MMOS, but you still need to space your objects so you they can be individually distinguished and navigated through correctly.

You need to space objects on your MMOS – columns of text, buttons on an interface, headings, and so on – so they are easy to perceive, but not overlapped and obscured. Each screen should not be so crowded with objects and features that the audience finds it difficult to see the information they need. You can use white space as a contrast to draw the user's eye (page 194). Within graphics, animations or videos, levels of colour and contrast should make the information clear and attractive. A common convention is also to avoid yellow or other light colours for text on a white background because this can be difficult to read on a screen.

A large area of white space may be used to balance a section that contains an equally large area of text, because it will be of equal visual 'weight'. Working with space means also working with balance.

Balance

A solution with a balanced design is visually appealing. Solutions with unbalanced designs can lack the appropriate emphasis, look untidy and may end up discouraging your intended audience from visiting them.

All elements of a layout have a visual weight. If the elements on either side or the top and bottom of the screen are of an equal weight, then visual balance is achieved.

There are two types of balance: symmetrical and asymmetrical.

With symmetrical balance, the visual elements on each side of an imaginary horizontal or vertical dividing line appear to be exactly the same in terms of visual weight, right down to the proportions and shading.

Asymmetrical balance occurs where visually matched weighting occurs through a combination of objects of differing sizes, shapes and colours. A large image slightly to the left

can be balanced by an object far on the right. Humans enjoy symmetry in nature and music, and they crave it in design. You can use asymmetrical balance to draw your audience's attention to an exceptional item.

Generating design ideas

A **design idea** is a brief, rough, 'back of the envelope' outline of a strategy for solving a problem. It lacks the detail and precision of a detailed design, but it points in the general direction of how a solution may be created. For example, if a storm is coming and a fence looks likely to be blown over, design ideas may include:

- using rope to support the fence
- driving in some stakes against the fence
- standing out in the storm and holding it up
- parking a truck beside the fence to hold it up
- convincing a herd of cows to lean against the fence
- nailing some wooden beams to the fence to act as props
- tearing down the fence and rebuilding it more strongly before the storm hits.

The design ideas are quite different, and not all are equally viable, but they cover a creative range of possibilities from which one could be developed into a detailed design for a solution.

Creative design techniques

There are several techniques for generating a range of creative and appropriate design ideas. The Study Design does not name specific design techniques that you must know, but these are the most common techniques. All of them aim to find the most effective and efficient solution to an information problem. Your techniques should take into account the functional and non-functional requirements of the MMOS.

FUNCTIONAL AND NON-FUNCTIONAL REQUIREMENTS

- 1 Using the definitions of both terms as set out in Chapter 2, page 79, as a guide to help you, identify:
 - a the functional requirements of the MMOS
 - b the non-functional requirements of the MMOS.
- 2 Justify your decision.
- 3 Discuss the functional and non-functional requirements of the MMOS in class with your teacher and classmates.
- 4 Suggest how a design technique could take into account functional and non-functional requirements.

Brainstorming

Try not to hamper your imagination by rejecting ideas too soon. Brainstorming is a process where ideas are presented in a non-judgemental, spontaneous, unstructured and admittedly somewhat haphazard process. The only rule is that no idea is criticised or rejected; every idea, no matter how outrageous or silly, goes onto a list of possible solutions.

Participants must have no fear of being ridiculed, judged, making mistakes or breaking rules. While some or many ideas *do* turn out to be ridiculous, sometimes a half-baked, half-comical

Do not be afraid to have fun with your MMOS. If you were using the paleo diet hypothesis, you could:

- create a Google Sites website as shown in Figure 4.14, which includes a page for each stage of the Outcome and begins by restating the hypothesis and defining the paleo diet
- use illustrations and photos of fossils from the paleolithic era
- have the report that substantiates your conclusion filmed and recorded as well as transcribed
- have a paleo chef from a restaurant perform your report as a dramatic script while you record it
- design the substantiation as an infographic, a video or an animation that uses HTML5, Keynote or Flash
- write your substantiation or another part of your Outcome in the form of a song, perform it as a singer-songwriter with a guitar and upload it as a video with a transcript
- perform your substantiation as a hard-hitting journalist in the style of an exposé and upload it as a podcast with a transcript

While you need to be inclusive, there is nothing to stop you being creative with your MMOS.

concept may in fact turn out to be an unexpected work of creative genius, or it may stimulate a related idea that would be perfect.

There are certain rules that you need to adhere to when you are running a brainstorming session. The most important one is that no one judges any contribution. You must accept all contributions. An idea that may seem slightly crazy to begin with can sometimes be workshopped into a great idea. In the 1970s, a brainstorming session came up with the idea of a pet rock. The idea was workshopped and before long you could buy not only a pet rock but a pet rock house and a training manual. Everyone had to have a pet rock in the 1970s and the idea made the company millions of dollars. It was the pre-technological version of a Tamagotchi.

Make sure that everyone listens to everyone else's ideas. Specify that only one person talks at a time and there is only one idea at a time. This not only ensures the shyest member of the group will contribute, but also makes it easier to record the ideas. Using these rules will assist you in getting a large amount of ideas with which you can work.



FIGURE 4.22 Whiteboards are popular brainstorming aids because they are visible to all, and are easy to edit.

Brainstorming example: NASA faces the problem of lifting people and equipment into space. It is hugely difficult, expensive, loud and dangerous. How can it be improved?

Brainstorms for this project include:

- helium balloons: float up, take off
- fire the rockets from the tops of mountains: reduce the distance to space

- a very, *very* tall ladder
- uhh, a slingshot, maybe
- a jet airliner carries the rocket ship as high as it can, then the rocket takes off from there
- antigravity.

While antigravity has no foundation in real science, some of the other design ideas could work, and would deserve more research. The team chuckles at the funny ‘very, very tall ladder’ idea until one person pauses and says:

Wait ... I wonder if we could somehow get a cable from the ground to low Earth orbit and anchor it in space, like a space elevator. You would ride up the cable to the end. The rocket can take off from there. You do not need all the fuel to achieve escape velocity... no need for ground launches.

‘And ... and re-entry,’ says someone else. ‘You could ride down the cable to get home. Easy. And cheap.’

From thinking that was whimsical, impromptu, unconventional and unconcerned with constraints comes a serious concept that can be further investigated.

Brainstorming is helped by inviting into the team people with different skills, experiences and areas of expertise. Sometimes, a group of specialists struggling for a solution may be inspired by an idea from someone who is not constrained by their shared assumptions, preconceptions and modes of thought.

Consult end users

Your solution, and all information solutions, will be used by real people. Thus, it makes sense to include real people in the design stage rather than wait for the **testing** and evaluation stages of the PSM to find out what they think of the solution. Manufacturers, political campaigns and film producers are known for their use of ‘focus groups’ of ordinary consumers whom they gather together and question about their likes, dislikes and reactions to design ideas.

A dedicated team of specialist designers may have their own ideas of what an end user wants, but you should value primary evidence of your audience’s requirements.

Mind mapping

Mind mapping is ideal for complementing the process of brainstorming.

Mind mapping is a technique for quickly generating and linking ideas. It is a creative and flexible tool that enables you to add, connect, organise and reorganise ideas. Mind-mapping software is generally flexible enough that you will not need to stop very often to learn how it works while mapping; in other words, when mapping, your creative flow will not often be interrupted.

Unlike old-fashioned sheets of butcher paper or whiteboards (Figure 4.22), electronic mind maps can stretch endlessly in any direction (Figure 4.23), easily add or remove links between items, allow entire branches of thought to be moved to new locations, and you will not face the laborious task of copying out all of the scribbled ideas at the end of the brainstorming session. The mind map can be saved for later development, printed, saved as an image, or transferred to a word processor.

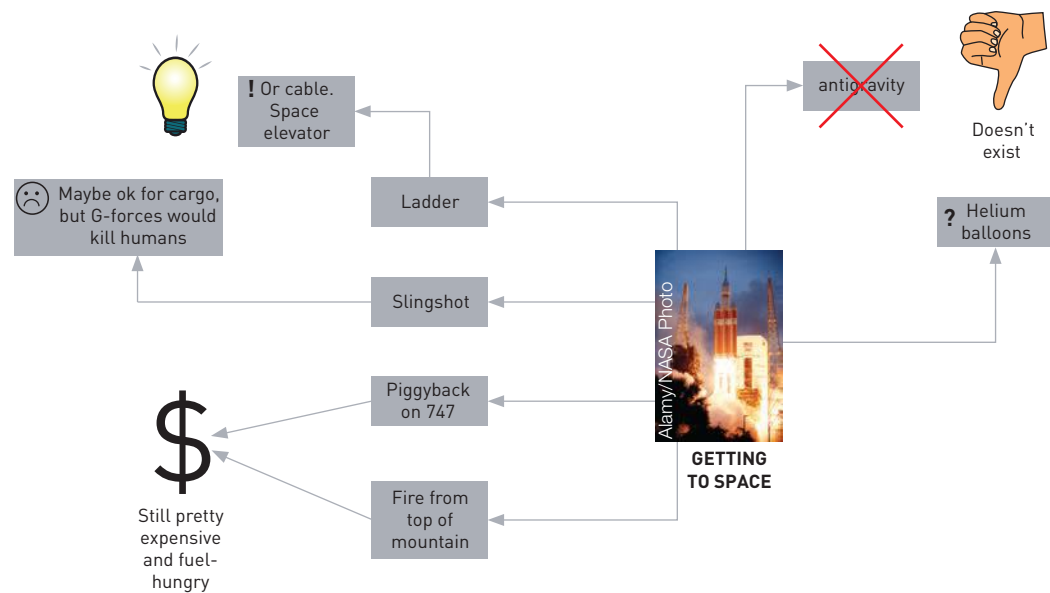


FIGURE 4.23 Mind mapping a project about how to get into space, using Inspiration software

Using the ‘getting to space’ problem, a mind map of the design process may look like Figure 4.23.

Graphic organisers

Graphic organisers are visual methods of organising ideas. A popular graphic organiser is a PMI. A PMI involves organising ideas into three columns: what has been successful (Plus), what was unsuccessful (Minus) and what needs more thought (Interesting) (Table 4.7). You can use a PMI to reflect and evaluate, or to brainstorm new ideas.

TABLE 4.7 Example of a PMI

Using helium balloons to reach space		
P	M	I
Quiet	Crashes if gas leaks	Can balloon go high enough?
Relatively cheap	Slow to reach stratosphere	How much does helium cost?
Limited payload weight		

A KWHL is another popular technique of structuring thought, focusing on the scope of an investigation, and distilling the results of the research. KWHL provides a specific structure for thinking about a topic.

- K – What do I know about the topic already?
- W – What do I want to know about the topic?
- H – How will I find out this information?
- L – What did I end up learning?

A Venn diagram is a way to represent the similarities and differences in a set of concepts or objects graphically. The intersection of two or more circles provides an opportunity to list any shared features (Figure 4.24).

A spider diagram (Figure 4.25) is a powerful tool to create an overview of a central idea. The body of the spider is the central idea and the branching legs radiate out to related ideas and sub-ideas.

There are dozens of variations of such visual tools to help organise and clarify ideas. Others include character maps, concept webs, POOCH (Problem – Options – Outcomes – Choice), ranking ladder (to prioritise or rank ideas, information or tasks), stair steps (to organise a process step-by-step), a chain of events, sequence charts (to put sequential factors in order), pie charts (to represent the relative sizes of components in a whole), bone charts, organising trees and even Gantt charts for managing project timelines. Gantt charts will be discussed later in the chapter regarding project management.

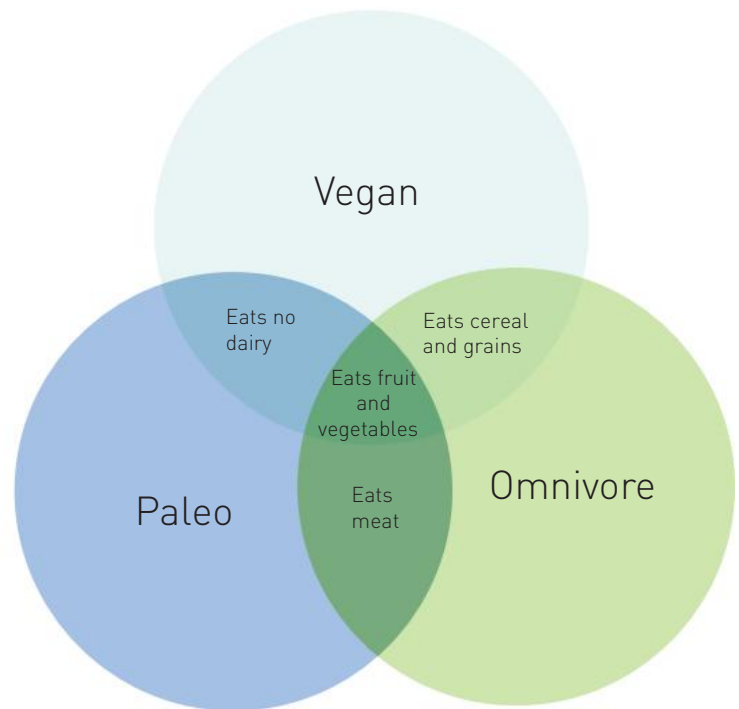


FIGURE 4.24 Paleo diet Venn diagram, showing relationship with the omnivorous and vegan diets

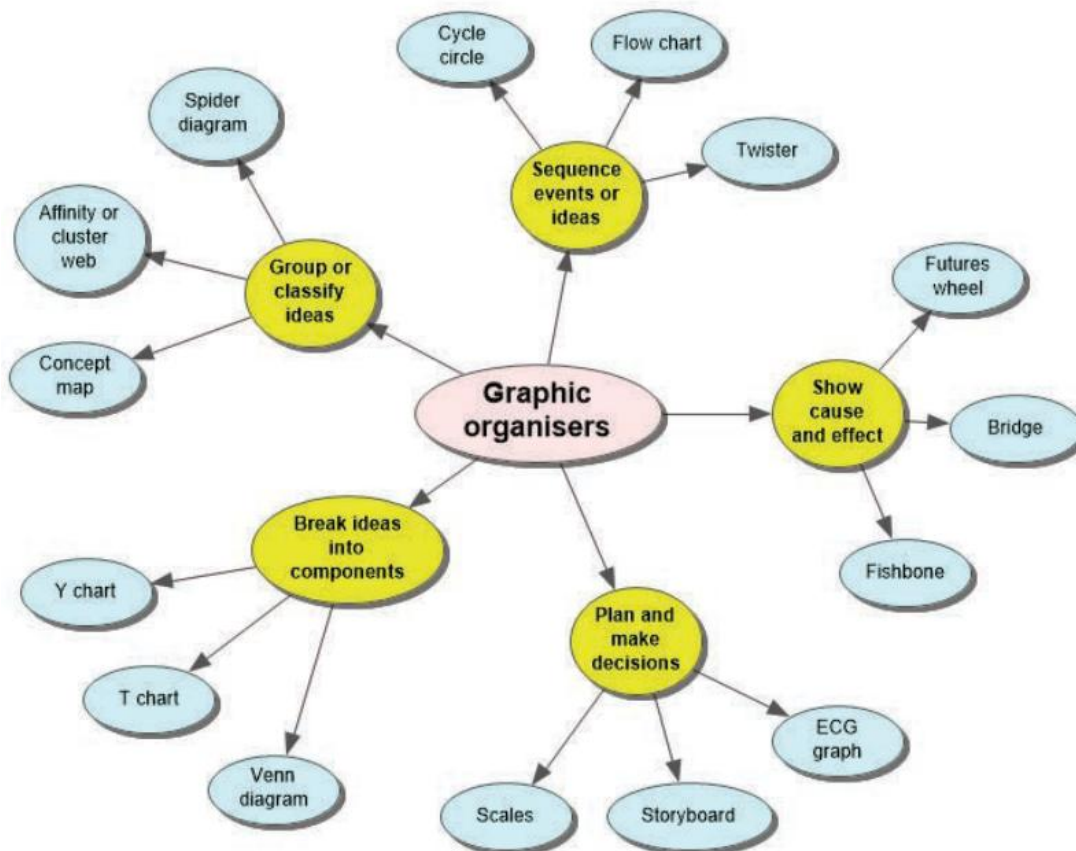


FIGURE 4.25 A spider diagram showing related concepts and sub-classes of concepts; this one was created with Inspiration software.

Attribute listing

Attribute listing helps break down pre-conceived notions about the nature of a product, and helps create fresh, new products.

Create a table with columns that make up a list of the attributes of a solution or strategy. In each column, fill in as many examples of that attribute as you can. Now mix and match one entry from each column to create the specifications for a brand new item. For example, you want to create a new game. You might create a table similar to Table 4.8.

TABLE 4.8 Game concepts

Genre	Age group	Platform	Gender target	Violence	Price range
Excitement	Toddlers	Phone	Male	None	Cheap
Problem solving	Early primary	Tablet	Female	Comic, cartoon-like	Medium
Comedy	Late primary	PC		Realistic, mild	Expensive
Education	Young teen			Ultra-violence	
Strategy	Late teen				
Gambling	Young adult				
Role-playing	Middle aged				
Simulation	Elderly				
First-person shooter					

From this, you may stimulate thinking about a cheap game about gambling for young female teens on a tablet with realistic, mild violence, or you may pursue an expensive, non-violent, role-playing game for retirees on a PC. Obviously, not all combinations seem feasible, but an odd mixture might just give rise to a brand new concept, such as the untapped market for a first-person shooter game for middle-aged women.

de Bono's Six Hats

Edward de Bono, renowned creative thinker and inventor of 'Lateral Thinking,' devised a powerful problem-solving strategy known as 'Six Hats'. It is based on the observation that problem-solving people can think in different ways, and change those ways deliberately, just as if they were changing from wearing one hat to another. When a problem needs a thorough investigation, people in the team are assigned to temporarily wearing one of the six hats, each of which has a style of thinking that balances the thinking of a different hat-wearer (Table 4.9).

TABLE 4.9 De Bono's six hats

Hat colour	Focus	Questions that could be asked
Blue	Managing	What is the subject under discussion? What are our goals? How can we achieve those goals?
White	Information	What are the facts? What do we know? What data do we need?
Red	Emotions	How does that make you feel? What do you like/dislike about it?
Black	Caution and critical thinking	Why may that not work? What could go wrong? What are the problems and dangers?
Yellow	Optimism	What are the positives? What value will that bring us? What benefits will it have?
Green	Creativity	Where else could that lead us? What about if we....?

Having representatives of each thinking type (hat colour) ensures that a team is not neglecting certain ways of thinking. For example, a group of naturally creative and gentle people may need the logical and antagonistic thinking of logical and confronting people if they are to investigate all aspects of a problem thoroughly.

Tips for creative thinking

Creative design can be learned. You do not need to be born with the talent. There are techniques that anyone can use to improve their design creativity.

Substitute

Replace part of the problem with something else. For example, if you are producing hundreds of certificates, do not use mail merge to take data from a spreadsheet and insert it into a word processor. Use a database instead.

Combine

Join unconnected things together, such as reducing the weight of camping supplies by combining a spoon and fork into a single utensil – the spork.

Adapt

Use an existing component in a different way, such as using mind-mapping software to create a site map for a website. The first spreadsheet was created using the concept of paper-based accounting books.

Strip

Reduce the problem right back to its most basic parts and see what is left. For example, the tiny and cheap computer, the Raspberry Pi, is a stripped down Linux PC with minimal components. Inspecting the basics may reveal the nature of a problem more clearly.



FIGURE 4.26 A Raspberry Pi

Compare

Ask yourself, ‘What other thing do I know that resembles *this* problem, and how does that other thing work?’ For example, when sending a number of print jobs to a single printer, how can they be handled? Like a group of waiting people at a gate, you could organise them into a queue and process them in the order of their arrival.

Sleep on it

Creators often reach a point where they can make no further progress. Rather than dwelling on the same failed ideas, it is often better to let them go and think of something else. While the front of your brain is enjoying a wrestling match with your sibling, or an episode of *Spongebob Squarepants*, the back of your brain will busily be pulling ideas together to create a solution.

Research

Thomas Edison said: ‘Through all the years of experimenting and research, I never once made a discovery. I start where the last man left off’. It is important to learn from your predecessors so you do not waste time re-inventing the wheel.

How have other people solved problems similar to the one you face? You are unlikely to be the first person in history to have faced a problem before. How have others coped? Their successes may lead you in the right direction, and their failures may prevent you wasting time.

Visualisation

Geniuses often make their thoughts visible because words cannot adequately convey the ideas they have. Einstein was famous for his non-verbal thought experiments. He visualised travel at the speed of light as travelling on a train. He said that written words and numbers did not play a significant role in his thinking process.

Leonardo da Vinci is renowned for his sketches of inventions. Galileo Galilei drew diagrams and maps of planetary orbits and phases of the Moon when others used mathematical formulas and words. Sigmund Freud, Alfred Hitchcock, Isaac Asimov, Beethoven and Mozart all reported the use of mental imagery in their creative processes. Dr Temple Grandin, famous for her work with livestock, said:

I think in pictures. Words are like a second language to me ... Language-based thinkers often find this phenomenon difficult to understand, but in my job as an equipment designer for the livestock industry, visual thinking is a tremendous advantage.

You may choose to use software simulations or models to help structure your thinking and construct knowledge.



Alamy/Stocktrek Images, Inc.

FIGURE 4.27 Galileo's drawings of phases of the Moon, based on observations through his telescope, from his 1610 manuscript, *Sidereus Nuncius*

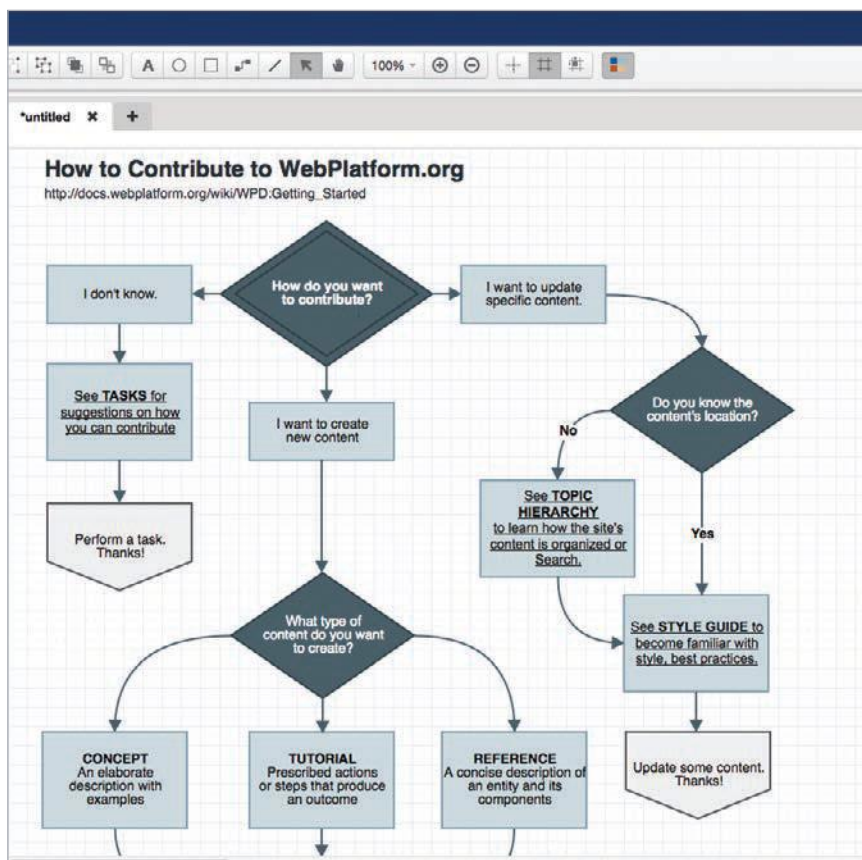


FIGURE 4.28 Creating a diagram with Gliffy

Be observant and prepared

Many inventions arose from people seeing things that were similar to a problem for which a solution needed to be designed. Can a blockage in a canal be similar to solving a blockage in blood vessels? How can thousands of ants travel safely and quickly through a small gap, while a crowd of human spectators take nearly an hour to leave a football stadium? Discoveries are often serendipitous.

Play Doh, Post-It notes, potato chips, Velcro, Teflon, Cellophane, insulin, Dynamite, stainless steel, Super Glue, Cornflakes and vulcanised rubber were all found by observant people after accidents or failed attempts to invent something else.

Research also suggests that creative people are typically hoarders – they keep lots of knick knacks, photos and articles around, and revisit these for stimulus at a later date.

Keep your eyes open and be receptive to connections between apparently dissimilar things. Revolutionary ideas often come from ‘ridiculous’ connections that no one previously let themselves consider. Physicists argued whether light was a wave or a particle, until someone stupidly (and correctly) proposed that it was both.

Someone with a solid knowledge of a topic and an ongoing curiosity about new findings is receptive and can recognise the importance of an observation to an existing idea. The uncreative observer will either not notice the idea, or fail to see its relevance to a developing design.

Take risks, persist and be brave

A creative design idea often needs to take the risk of being dismissed, mocked or rejected. Many of the greatest breakthroughs were rejected at first and took a lot of time and effort to be proven right.

The Germ theory, that diseases are caused by microorganisms, was put forward by Louis Pasteur in the 1860s. It superseded the miasma theory, that a poisonous vapour in the air caused diseases. This theory had endured for several centuries. Pasteur’s theory was initially mocked until further experimentation showed it to be most likely correct. In more recent times, Steve Wozniak combined the concepts of a typewriter, a calculator and a display. He was envisioning a whole new technological paradigm: the personal computer. His employer at the time, Hewlett-Packard, rejected Steve’s concept five times. This led Wozniak to team up with Steve Jobs, which then led to the creation of Apple Computers. The idea of a tablet computing device had been tried by Apple and Microsoft and ended in failure. Steve Jobs tried again when the technology was mature, and the iPad was an instant success. James Dyson (of Dyson vacuum cleaner fame) is believed to have created 5000 prototypes of his vacuum cleaner over five years before he got it right. These examples show that it is persistence, not genius, that is probably the greatest contributor to success.

Thomas Edison, developer of the light bulb, phonograph and electric power, famously said:

Genius is one percent inspiration; 99 per cent perspiration.

Many of life’s failures are people who did not realise how close they were to success when they gave up.

I have constructed 3 000 different theories in connection with the electric light, each one of them reasonable and apparently likely to be true. Yet only in two cases did my experiments prove the truth of my theory.

Evaluating design ideas

When designing the solution to a problem, the first design idea you have is rarely the best one. A different strategy might be cheaper, easier, faster, more effective, or may better meet the client’s demands. While one design idea may be attractive to the developer, the client may have

non-technical constraints or priorities that will make one strategy more attractive to another. Providing a range of design ideas lets the clients choose the solutions that best suit them.

You may have used a design idea successfully in the past, but it may not be appropriate in the current circumstances. Although previously proven strategies can be useful, you need to be willing to think outside the box. Old strategies will not work for you in every situation – it is lazy and unimaginative to assume that they will.

A successful problem-solver will consider current functional and non-functional requirements and relevant constraints to develop an imaginative range of options from which the best design idea can be chosen and developed into a detailed design.

The criteria for choosing the best design idea include:

- ease of use
- how long it will take to implement
- scalability (how easily the product can be increased in capacity)
- its scope for future modification and enhancement (for example, one design idea may include plug-ins so extra functionality may easily be added)
- the degree to which it satisfies all requirements
- the degree to which it copes with constraints
- ease of implementation
- development cost, future running and maintenance costs
- the amount of labour required to create it and keep it working
- the amount of disruption likely to be caused to the organisation
- compatibility with existing hardware, software, data and procedures
- the amount of training required for staff.

Some design decisions can be very difficult, and require careful balancing of competing needs – usually cost against quality. A design that is cheap and quick to produce may be barely competent, quickly wear out or be unpleasant to use. A superior design that would lead to a solution with a long life and happy users will probably take longer to produce and cost more.

Compare the likely differences in design philosophies and criteria between the pairs shown in Table 4.10.

TABLE 4.10 Pairs

A Formula One race car	A Smart Car
 Shutterstock.com/David Acosta Alley	 Alamy/Keith Erskine
A PC desktop tower	An all-in-one Mac
 Courtesy of Dell Inc.	 Copyright © 2015 Apple Inc. All rights reserved

TABLE 4.10 (continued)

A basic, inexpensive coffee table	An ornate, expensive antique table
 Shutterstock.com/Maryana M	 Shutterstock.com/bergamont

Design tools

Input-process-output (IPO) charts

IPO charts help to design **algorithms** in spreadsheets, databases and programs, which can be used to devise formulas, scripts and program code. An IPO chart is created using the following steps.

- 1 Identify the information required, such as a person's age. List it in the Output column.
- 2 Determine what input data is needed to calculate that output. For an age, that means a date of birth (DOB) and the current date. The data goes into the Input column.
- 3 What processing needs to be done on the input to calculate the desired output? This algorithm goes in the Processing column, expressed in pseudocode: a mix of English and actual formula language to outline a way of calculating the information. Do not write full formulas; only the basic logic of the calculation strategy is required. Turn it into a proper formula later, during development.

Notice that in Table 4.11 there is one row for each output, and that output from previous calculations can be used as input in later calculations.

TABLE 4.11 Example IPO chart

Input (data)	Process (algorithm)	Output (information)
• DOB • Current date	Days between DOB and current date $\div$ 365	Age in years
• Quantity • Cost per item	Quantity $\times$ Cost per item	Subtotal
• Is tax payable? • Tax rate %	Subtotal + (If tax is payable, then: Subtotal $\times$ Tax rate %)	Total cost
• Age in years • Total cost • Discount rate %	If age in years $\geq$ 60, Total cost $-$ (Total cost $\times$ discount rate %)	Senior citizen cost

Mock ups or annotated diagrams

Mock ups or annotated diagrams show the intended appearance of printed output, onscreen information and interfaces.

Mock ups show features such as:

- positions and relative sizes of controls (buttons, scroll bars, status bars, scrollbars)
- positions, sizes, colours and styles of text (headings and labels, body text)
- menu positions and contents
- borders, frames, lines, shapes, images, decoration and colour schemes
- object alignments (vertical, horizontal, diagonal)
- contents of headers and footers.

A mock up can be considered successful if you can give it to another person and they can create the interface without needing to ask you questions about it.

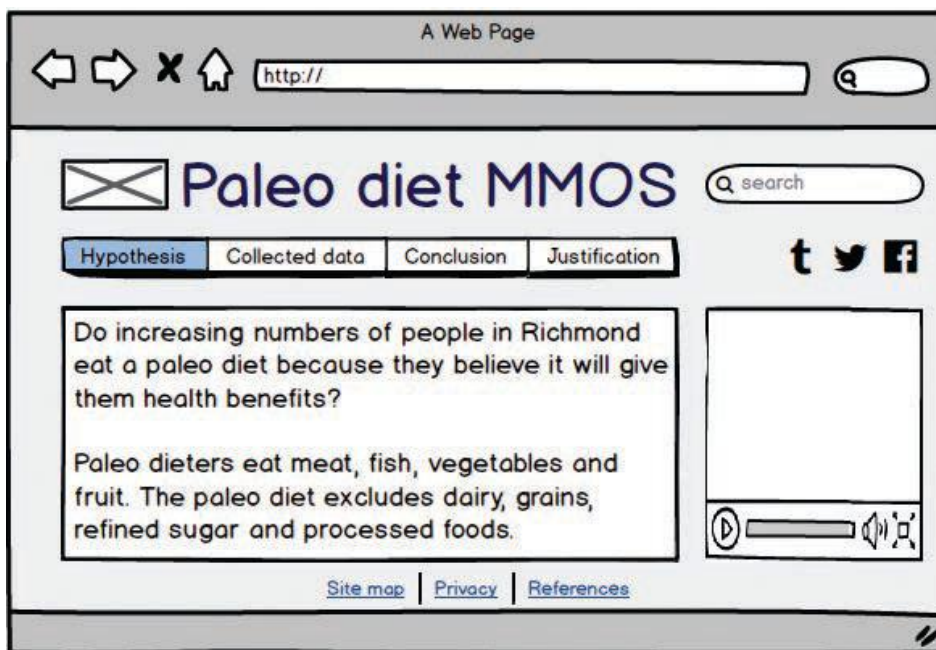


FIGURE 4.29 Paleo diet MMOS mock up created in Balsamiq

Site map

A site map is a hierarchical diagram that shows the pages and links of a website, screens of a slideshow, or forms in a program. Use it to summarise your MMOS's structure, ensure all relevant topics are included, and give your audience an overview of the scope of the site.

Figure 4.30 is an example of a site map.

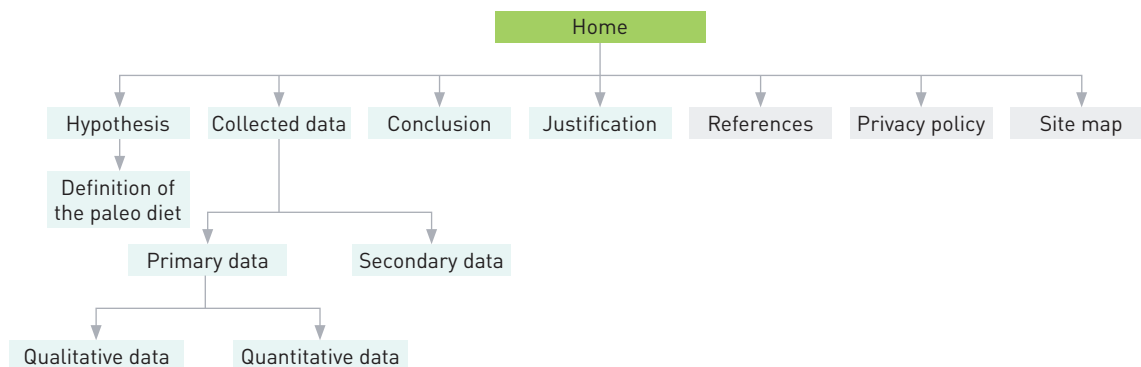
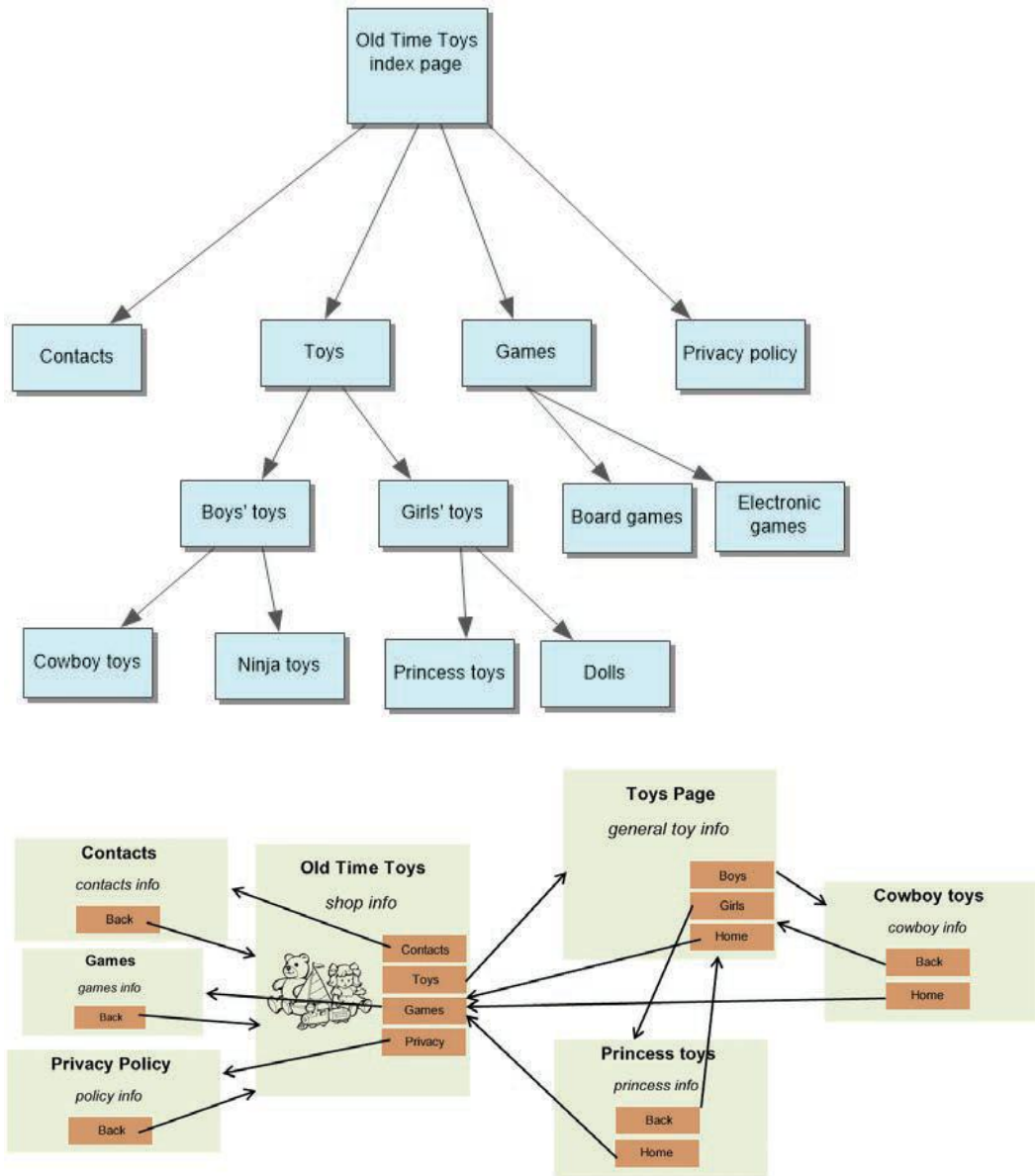


FIGURE 4.30 Paleo diet MMOS site map

Storyboard

A website's storyboard focuses on site navigation. Like a site map, it shows all pages in the site, but also adds information about how objects within those pages are used as navigational links. Figure 4.31 shows a site map for a toy website on the top, and its storyboard on the bottom.



THINK ABOUT INFORMATICS 4.5

Figure 4.31 opposite represents a not-uncommon model for a site map and storyboard for a toy website. Referring to the earlier parts of Chapter 4 and inclusive language, suggest how these documents could be modified to be more inclusive and less sexist.

FIGURE 4.31 The site map for a toy website (top) and its storyboard (bottom)

STORYBOARDING

- 1 Redraw the Paleo diet MMOS site map (Figure 4.30). You can do this on a sheet of A3 paper, or electronically.
- 2 Annotate it to add links on each page, and then draw arrows from the link to other pages, to convert it into a storyboard. You can use Figure 4.31 as a model for a storyboard.
- 3 Write a short summary (2–3 sentences) on how you might apply this to your own site map and storyboard for Unit 4, Outcome 1.

Organisational and hierarchy charts

These charts map out the relative positions of items, or people, in a hierarchy. An organisational ('org') chart particularly indicates who has authority, and who reports to whom.

Layout diagrams

A layout diagram shows the components and their relationships in a system. A restaurant planning a software management system might use a diagram like this to visualise their solution (Figure 4.32).

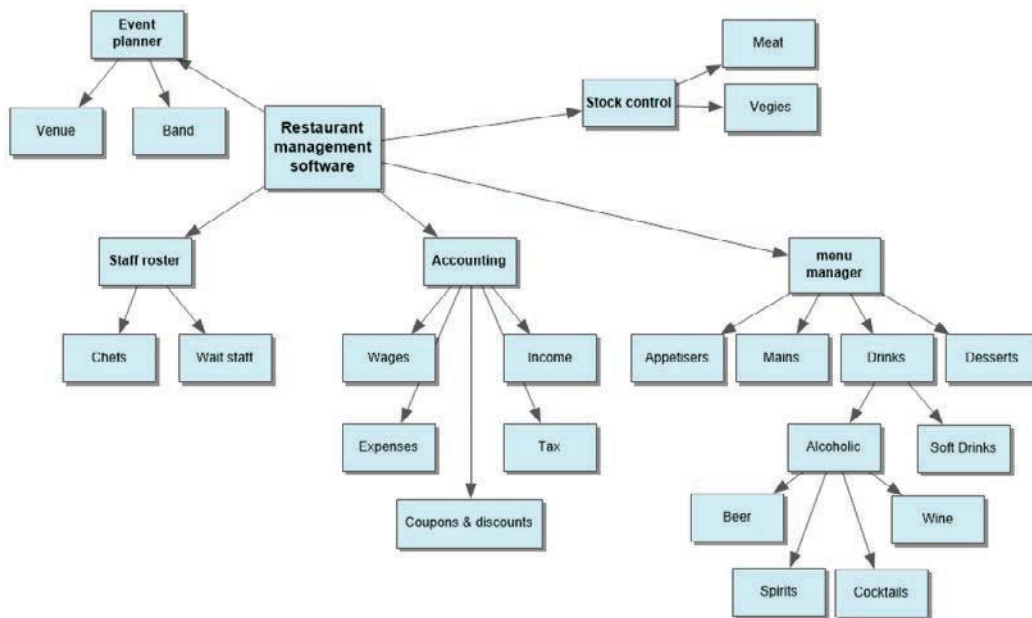


FIGURE 4.32 Example of a layout diagram

The software Inspiration is useful for creating quick and easy design diagrams. Dia is another free and useful tool for diagramming. In Microsoft Word, you can use Insert > Shapes. The drawing tools in PowerPoint are another easy way of combining images, shapes and text to form a complex diagram. Export the slide as an image (Save As) or take a screenshot (PrintScreen key in Windows. Command+Shift+3 in Mac OS) and paste the image into a graphics editor for cropping and saving.

Documenting the progress of projects

In the first half of this SAT (Unit 3, Outcome 2), we discussed how to manage a project, and why project management is important. In Unit 4, you will continue to use these project management techniques to complete your hypothesis research by creating a multimodal online solution.

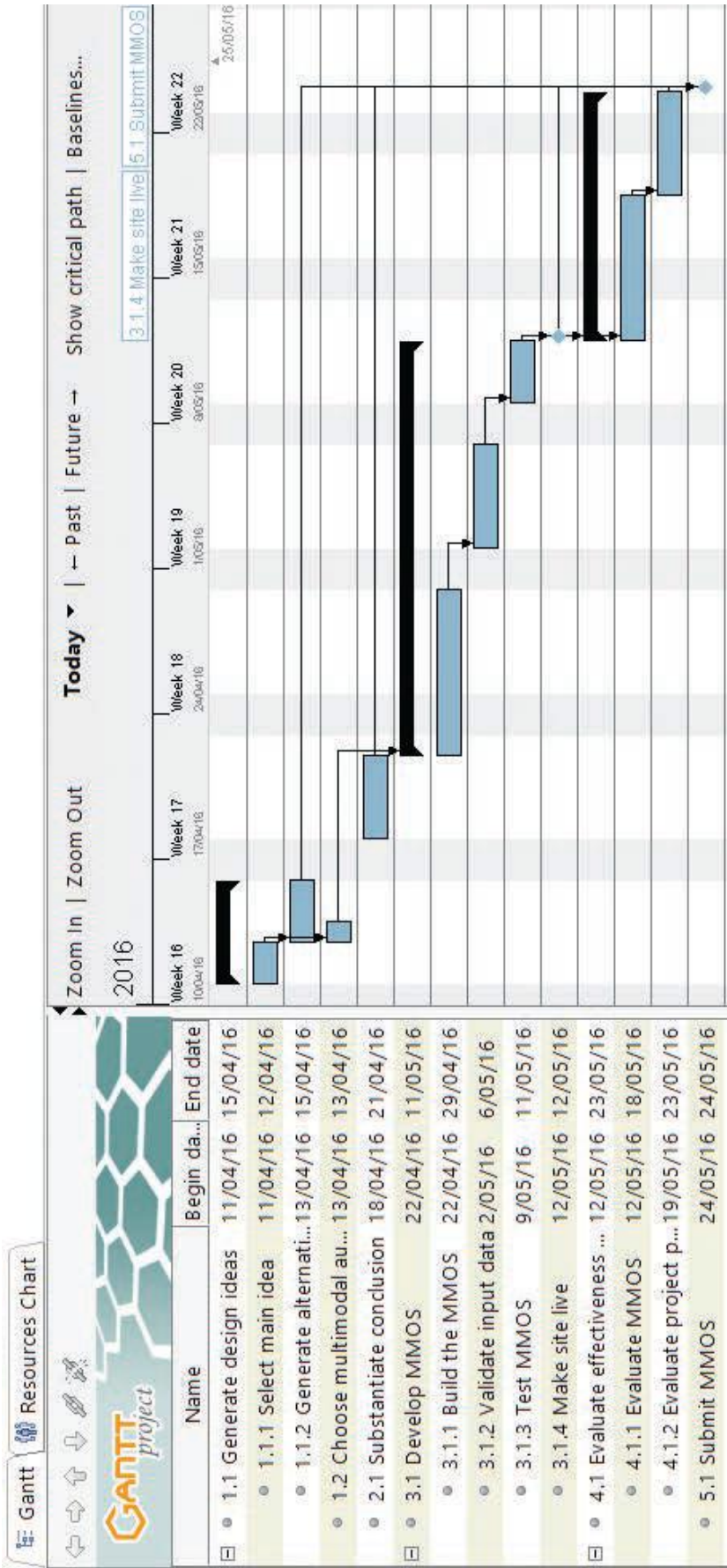
It is unlikely that any project ever proceeded perfectly in line with the project plan. One unexpected rainy day, a hard disk crash or a sick day for a key worker can slow down a work team enough to affect tasks, other teams and deadlines. Project plans are not written in stone and obeyed regardless of real-world events. Gantt charts are modified on an ongoing basis to reflect reality.

Project plans are living documents. Tasks that run overtime may have resources added to them or be modified so they finish earlier. Bad weather may force changes to scheduling, so indoor tasks may be done instead of outdoor work. Late deliveries of equipment may cause a project manager to take people from one task and assign them to another one that can proceed without the deliveries.



NelsonNet additional resource: Figure 4.33 Sample GanttProject chart, p. 224

FIGURE 4.33 Gantt chart: a project plan for the MMOS



GANTT CHART

- 1 Redraw Figure 4.33 by hand or electronically, or modify the Gantt chart in Chapter 4 of the student website (The Figure 4.33 'Additional resource' icon linked on the previous page) to add two days to the '2.1 Substantiate conclusion' stage.
- 2 You will need to adjust the rest of the project so that the final date on the rest of the project does not move forward by more than one day.
- 3 Justify why you have chosen to reduce the length of at least one stage by one day.

A project plan may be annotated to explain reasons for changes to task schedules or resourcing priorities. When the project is later assessed, these annotations might serve as valuable lessons when undertaking the next project. Annotations might also be added by other project leaders to advise the team of significant news or concerns. Annotations could be handwritten, or added as notes in the Gantt chart itself.

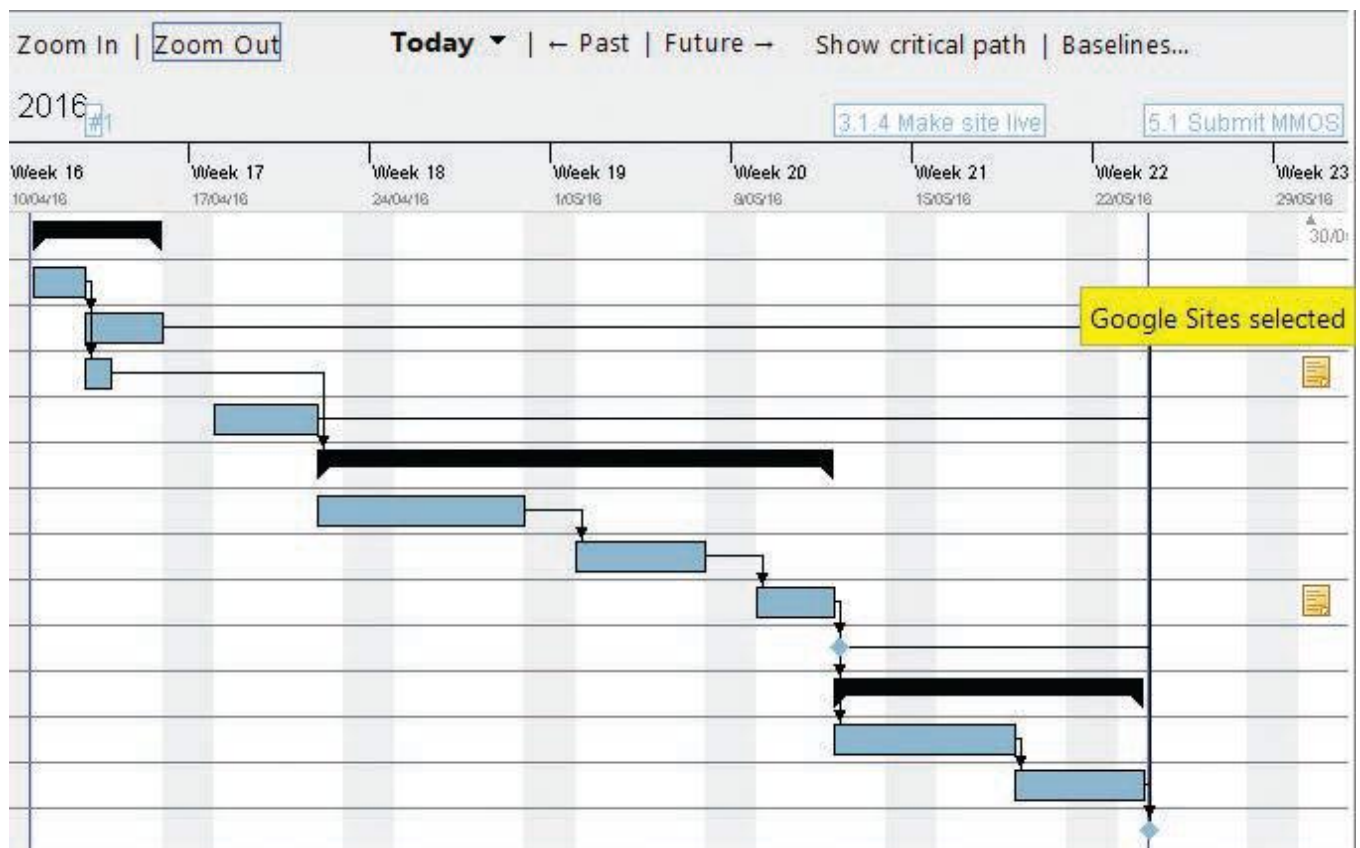


FIGURE 4.34 Gantt chart with notes added

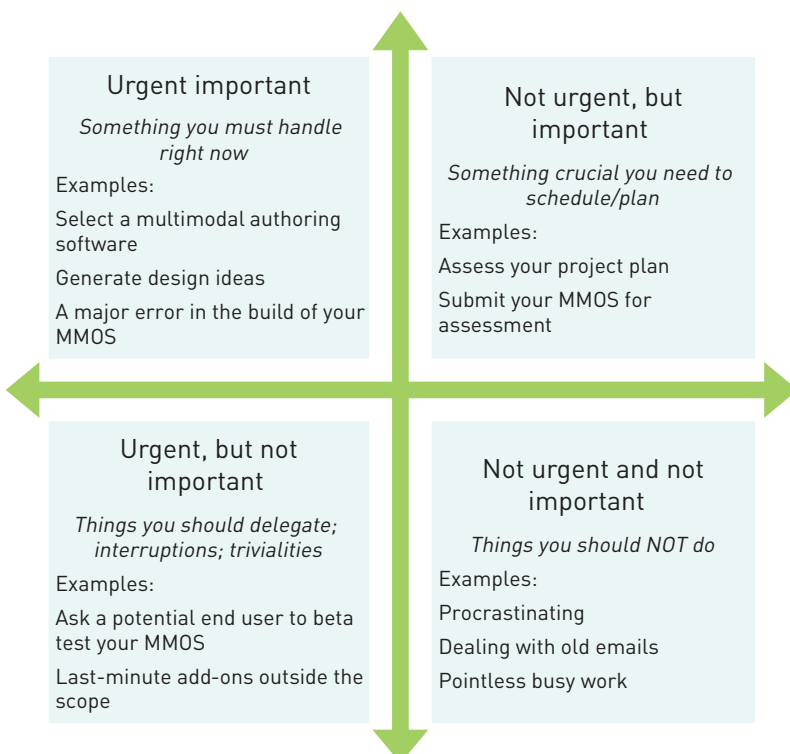


NelsonNet
additional resource:
Figure 4.35 Project
log template, p. 226

Project logs are a record of all the small and large steps a project takes on its way to completion. It is usually in electronic form and shared with all project leaders. It may be created with specialist log software, or with Microsoft Word or Excel. It can be shared online using Google Drive or a similar technology.

	A	B	C	D	E	F	G	H	I	J
1	Issue #	Date	Reported by	Page/functionality	Description	Issue type	Priority	Impact	Suggested fix	Status
2	1	2016-05-09	Me (Developer)	Hypothesis https://sites.google.com/.../paleo-diet-mmoss/hypothesis	Video describing the paleo diet is not embedding correctly.	Functionality	High	High	Check the embedding instructions in the backend code to make sure I haven't added the video incorrectly.	Resolved – re-test
3	2	2016-05-09	Me (Developer)	Hypothesis https://sites.google.com/.../paleo-diet-mmoss/hypothesis	The font size of the hypothesis body text is too small.	Design	Medium	Low	Enlarge the font size to 14pt for the hypothesis itself to distinguish from the description of paleo diet for the purpose of the Outcome.	Resolved and confirmed
4	3	2016-05-09	Me (Developer)	Conclusion https://sites.google.com/.../paleo-diet-mmoss/conclusion	The audio transcript of the conclusion is taking too long to download, and is also taking too long when streaming.	Useability	High	High	Re-upload after converting wav file to mp3	Unresolved
5	4	2016-05-09	Me (Developer)	Justification https://sites.google.com/.../paleo-diet-mmoss/justification	Figure 1.2 and 1.4 on Justification are coming up with red X's and no alt text.	Data	High	High	Re-upload missing images. Check that img src tags link to correct image files. Check that alt text is in place for all images for accessibility.	Fix in progress
6	5	2016-05-10	End user (beta tester)	Site map	Site map does not include privacy page.	Data	Low	Low	Add privacy page to site map image and links	New

FIGURE 4.35 Project log template created in Microsoft Excel. Columns C, F, G, H and J are data validation dropdown lists. Columns G and H also have conditional formatting applied to indicate priority and impact by colour: low, medium, high, critical and showstopper. The template for this project log is available on the NelsonNet student website.



A project log helps a team to keep track of project tasks, which team member is responsible for the tasks, when deadlines and milestones are due and the status of tasks. It can also help you keep track of the work you are doing on your MMOS alone. In conjunction with your Gantt chart, it can be a valuable tool to help you manage your project efficiently. A project log can include time and date stamps, comments on progress, project risks, issues that arise, ideas for solutions, actual solutions, explanations of decisions, results of testing, forecasts and warnings, task changes, photos and future action needed. A project log is like a diary that describes the complete history of a project.

If you find yourself struggling to divide up the tasks by priority in general, creating a priorities quadrant may also be of use (Figure 4.36).

FIGURE 4.36 Example of a priorities quadrant

Efficiency tips

You have a limited amount of time to complete the project so you need to make sure that you use your time as efficiently as possible. Your first priority is to make the solution that you are proposing work. While the attractiveness of your solution is an assessment criterion, appearance is worth much less than the quality of its information. You need to make sure that you allocate your time appropriately to reflect this priority.

One way of doing this is to use a plain text heading and then insert all of the key information. If you have time later, return and design a graphic heading and modify the choices of fonts. Remember, a beautiful solution with incomplete content is a demonstration of wasted time.

It is easy to waste time ‘fiddling’ with small details. It is not a good use of time to spend half an hour in Photoshop cutting a figure from its background when there are more important things to be done. You may enjoy using Photoshop in this way, but it is an inefficient use of your time and this will be reflected in your assessment. You need to complete the less exciting jobs, such as typing data and testing that all links work.

It may help to make an agreement with yourself to do just a little of the task. It will not be finished, but at least you have made some progress. This way your self-confidence will be boosted by your courage, and you may even decide that since you have started you might do a little bit more while you are at it. The hardest part of most jobs is taking the first step. You will find, however, once the first step has been taken, it is usually much easier to continue.

Managing files

Managing files safely and efficiently is a crucial skill when you are managing your project, especially when there are large quantities of valuable data on a computer, network or website.

You will be managing many files as you develop your MMOS, and will not want to lose valuable time and effort through ineffective file management practices.

File system basics

To store, retrieve or link to information, it is necessary to be able to describe file locations in a complex storage environment. This is done by describing a **path**, which is a list of the folders and subfolders that lead to a file.

The **root** of a drive is the starting point of a path, and refers to the part of a file system that is not subdivided. It is like a building’s foyer: the single entry point from which many destinations branch off.

Imagine we are finding Mr Patterson in a three-storey office block.

In this building, the first main paths are the three floors rented by companies. From the foyer, one of those floors is chosen. We know Mr Patterson works for ABC Enterprises so we follow that path.

Within one of those companies are further logical subdivisions: PR, Accounts, and Sales. Mr Patterson works in Sales, so we enter that subsection. Sales is divided into Australia and Foreign. Mr Patterson works in Australian Sales so we enter that office.

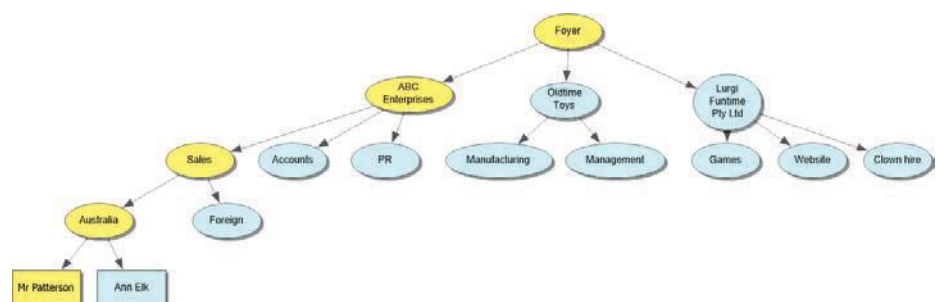


FIGURE 4.37 A path through a building

Finally, we find Mr Patterson at his desk. We could describe the path to Mr Patterson's desk hierarchically like this: Foyer/ABC Enterprises/Sales/Australia/Mr Patterson.

File systems and websites work the same way.

The root of a disk drive is its 'foyer'. In Windows, disk drives are given letters such as C: where the colon marks C as a disk drive. Macintosh uses names such as LionHD.

Each folder may be further divided into subfolders, they in turn may be subdivided, and they may contain subfolders.

Therefore, if the building were a disk drive and we were searching for the document named JackPatterson.doc, the path may be described as:

`C:\ABC Enterprises\Sales\Australia\JackPatterson.doc`

On a network, we use Universal (or Uniform) Naming Convention (UNC) syntax:

`\\buildingserver\ABC Enterprises\Sales\Australia\JackPatterson.doc`

On a website, we use a Uniform Resource Locator (URL):

`http://building.com.au/ABC_Enterprises/Sales/Australia/JackPatterson.html`

While the directions of the slashes may vary, the same basic method of describing hierarchical paths is universally used in Windows, Mac OS, Linux, Android and web servers.

Planning a disk structure

It is important to plan a directory structure from the beginning to make storage and retrieval quick, easy and intuitive later.

In the root of the drive, create the most logical major subdivisions of data. If they are well chosen, there will not be a single document that cannot easily be categorised under one of those major headings. These major subheadings can now be subdivided into logical subsections. If necessary, these subfolders may be further subdivided. Avoid creating so many levels of subdirectories that they end up being so specific that there's only one file in each folder; for example:

`Q:\letters\personal\holidays\2015\fiji\hotel\complaints\soap\complaint-about-soap.doc`

On the other hand, avoid using multi-purpose miscellaneous folders that contain hundreds of unrelated files that you could not (due to confusion) or would not (due to laziness) categorise logically.

File management terms

The folder you are currently working in is the **current directory**.

The folder that contains the current directory is called the **parent**. The parent directory is shown as two dots in a path.

An **absolute file path** describes the location of a file, starting at the root.

A **relative path** describes the location of a file in relation to your current directory. For example, the relative path `..\file.txt` means 'in the parent directory, find file.txt.'

Relative paths are commonly used in webpages. Links to other pages within the site can stay unchanged if an entire site moves to a new web server. If ABC Enterprises moved to a new building in a new city, you could still find Mr Patterson easily because his *relative* position within ABC Enterprises has not changed even though his *absolute* position has changed to a new city.



FIGURE 4.38 The parent folder shown as two dots in a file manager

Version control

When working on large and complex tasks, especially in a team, management of document versions is vital. If you just keep pressing **Save** for a document you have been developing, sooner or later:

- you will accidentally save an empty document with the same name and lose everything
- you realise the last change you saved was a big mistake, and it will take hours to revert the document to its previous state
- you find copies of the same document on your desktop, laptop, phone, and website and cannot work out which is the most recent
- you and your co-worker each have different versions of a file called `AnnualReport.doc` that you need to exchange and you lose track of which is the authoritative version.

A solution is to add version numbers to document names. You would **Save** the file first as `MyPattersonEssay-v01.doc`. The next day you load it, make changes and **Save As** `MyPattersonEssay-v02.doc`. Put it on a flash drive to take to school, edit it again, **Save As** `MyPattersonEssay-v03.doc`. Take it home, work on it, **Save As** `MyPattersonEssay-v04.doc`. When the work is finished, you may delete all of the preliminary versions – or **archive** them, just in case you may need them later. Disk space is so huge and cheap that keeping even dozens of previous versions is not going to be a problem.

File management utilities

There are many free and commercial types of utilities dedicated to streamlining your file handling.

- Batch renaming utilities rename thousands of files, add prefixes or index numbers, insert leading zeroes into numerals, replace text and alter file extensions.
- Duplicate file searchers scan massive numbers of files to find space-wasting duplicates.
- Disk management utilities such as TreeSize help manage huge numbers of files in many ways, and locate where disk space is being used or wasted.
- Replacement file managers that come with operating systems are usually barely adequate. Replacements can offer more functionality, speed or ease of use, such as FreeCommander for Windows or the XtraFinder plugin for Mac OS X. Figure 4.39 shows a dual-pane file manager. It provides a search box (in the bottom right corners of panes) and allows zipped archives and online FTP sites to be treated as if they were local folders on disk. It also offers ‘flat’ view, showing all files in all subfolders, making it easy to find files.

- Partition managers, such as Easeus, organise physical disk structures, such as dividing a single physical disk into two 'logical' drives: drive C: for the operating system and programs, and D: for user documents. They can often also back up an entire drive into a single-file backup that can be copied to another drive. Handle programs like this with great care since you can erase an entire disk in seconds.

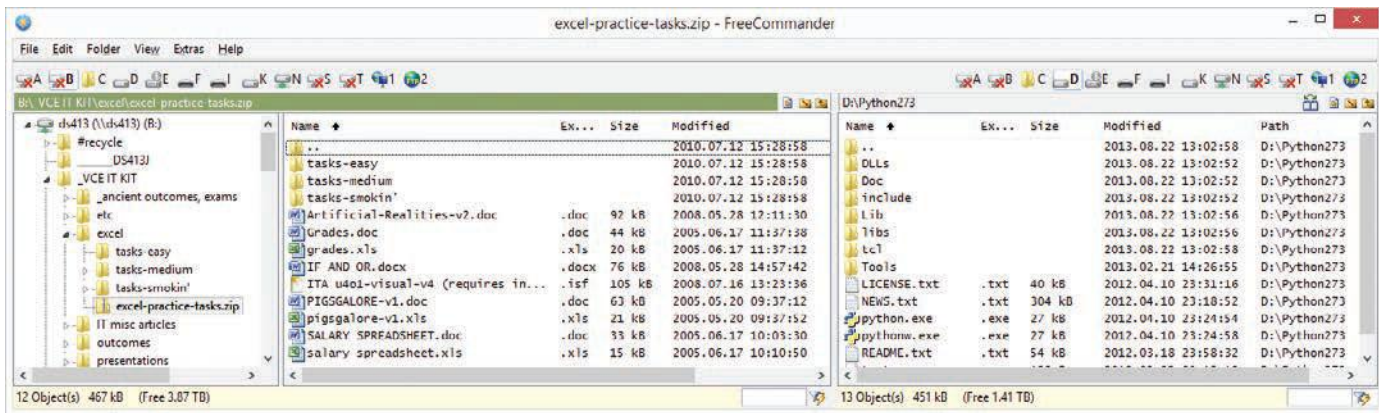


FIGURE 4.39 A dual-pane file manager showing a hierarchical structure

- ISO imaging tools create a perfect, compressed copy of an entire optical disk that can be stored as a single backup file or copied to a new disk.
- File searchers find those files you know must be somewhere.
- Undelete tools such as recycle bins and trash cans allow a limited ability to recover deleted files. Dedicated undelete tools such as Recuva are more capable.
- File shredders or wipers delete files so effectively that no one is ever likely to be able to recover them.
- Utilities, such as Teracopy, act as file copiers to replace operating system components.
- Specialist file managers, such as IrfanView and Adobe Bridge, manage and manipulate large numbers of images.
- File synchronisers compare the contents of two folders, determine which files are newer and copy files back and forth between the folders to sync them, bringing each one up to date.
- Backup software will identify new or changed files and automatically copy them to a second location.
- Compression software (such as PKZip, WinRAR, 7-zip and many others) will compress and/or group large numbers of files into smaller, more manageable bundles that are quicker and easier to archive or communicate.
- Encryption utilities such as PGP can make files, folders or entire disks unreadable to unauthorised people.
- Macro recorders such as AutoHotKey let you press key combinations to insert recorded keystrokes and mouse clicks into any application as if you had done them manually. With a single keystroke they can automate repetitive tasks (for example, entering your email address) or simplify complex tasks you rarely carry out (for example, searching for all occurrences of
 in a webpage's HTML and changing them to spaces).

Manipulating data with software

To a computer, everything is data: text, music, programs, movies, photos, games and webpages. They all need processing to produce a MMOS. In Unit 3, Outcome 2 you needed to be competent in data processing to manipulate the data to generate information to test your hypothesis. In Unit 4, Outcome 1 you need to manipulate **multimodal data** such as images, video, charts, text and audio effectively.

Multimodal authoring

Even the most modest of smartphones now offer facilities to create and edit documents incorporating images, video, audio, animations and text. However, you should still approach multimodal authoring with care and thought.

Media editors that may assist you with your MMOS are listed in Table 4.12.

TABLE 4.12 Media editors to assist with multimodal authoring

Media editor	Description
Image editors	Adobe Photoshop (expensive, powerful), GIMP (free), Inkscape (free) Enable users to crop , convert to other formats, resize, distort, repair, combine images, add text and shapes, and add special effects to images.
Video editors	Windows Movie Maker (free), Avidemux, Videoredo, AVS video converter, or Adobe Premiere Pro (very complex, powerful and expensive) Can read video files in several formats and join or cut scenes; add soundtracks; add special effects like credits, fades and dissolves; distort colours; and export to different formats.
Audio editors	Audacity (free) or Adobe Audition (expensive, powerful) Can crop, cut and join clips; repair clicks, hums and background noise; change volume; merge and mix multiple tracks; save in a variety of formats
Animation editors	Animatron (free, online), Keynote (expensive, powerful) or Adobe Edge (expensive, powerful) Create a timeline in which users can add and edit video, audio and text to create engaging videos
Text editors	Microsoft Word, WordPerfect, Microsoft Publisher, PowerPoint, NotePad Fundamentally important, whether they are high-end, like Microsoft Word, or a basic text editor, like Notepad. Advanced text editors, such as Editpad and Notepad++, offer multiple tabs, regular expressions, macros and code colouring for programmers.



GIMP



Inkscape



Windows Movie Maker



Audacity



Animatron



Notepad++

Validation techniques

Validation checks that input data are *reasonable*. Validation does not and cannot check that inputs are accurate. How could validation tell whether a person is being honest when entering their age? However, validation can detect problems when a person enters their age as 178 years, or 'fish', or nothing at all. You can perform validation manually (yourself), or allow software to do it for you.

Computers are particularly good at conducting validation checks.

- 1 **Range checks** ensure that data is within acceptable limits (for example, children enrolling in kindergarten must be 3–6 years old) or comes from a list of acceptable values (for example, small, medium or large).
- 2 **Type checks** ensure data is of the right type; for example, the age that has been entered is actually a number.
- 3 **Existence checks** ensure that a value has been entered.

THINK ABOUT INFORMATICS 4.6

In Unit 4, Outcome 1, what data will you mainly need to validate, and how will you achieve this?

FIGURE 4.40 Validation rules in FileMaker database. Here an ID field is made compulsory ('Not Empty') and unique, and within a defined range of values. The database is also told what error message to display if validation fails.

Testing

If a software solution fails, it could annoy, disadvantage, or kill users, so thorough and careful testing is necessary whether the solution is a game, a website shopping cart, or an airliner's autopilot.

If your MMOS fails because of undiscovered faults, it may become unpleasant to use, or completely unreadable.

Testing checks that a solution produces correct output and does what it should do. Testing is not easy, quick or cheap – especially for a product such as an operating system with megabytes of code in thousands of files created by hundreds of people.

The typical steps involved in testing are as follows.

- 1 Decide which tests will be conducted.
- 2 Create suitable test data.
- 3 Determine expected results.
- 4 Conduct the test.
- 5 Record the actual results.
- 6 Correct any errors.

There are many testing types, each intended to uncover different kinds of errors at different times during development. The types of testing relevant to your solution are listed in Table 4.13.

People can perform manual validation, especially proofreading for sense, clarity, relevance and appropriateness. In addition, unlike spreadsheets, people tend to smell a rat when values entered would pass electronic validation checks, but are inaccurate because they are ridiculous.

Similarly, Microsoft Word can find words that are not in its dictionary, but it cannot advise writers that a paragraph is boring or that the previous page was pretentious, sexist and irrelevant.

Remember:

- validation checks the reasonableness of data inputs
- **testing** checks the accuracy of information outputs.

In your SAT, ensure that all of your data is thoroughly and appropriately validated.

TABLE 4.13 Testing types

Type	What is tested?
Informal (alpha)	The part of the solution that has just been finished
User acceptance (beta)	Typical end users use their own equipment to check that the finished solution is acceptable in real-world conditions
Component	A single part of a system works properly by itself (for example, a web form applies the correct delivery cost for a given destination)
Integration	Individual parts of a system work together (for example, the web form sends correct data to a separate database)
System	All components in the solution work properly as a single unit
Installation	The multimodal authoring software is installed correctly and working on your system, server or domain
Compatibility	The multimodal authoring software and its components are compatible with a variety of browsers and OS
Useability	Whether users can operate the solution quickly and easily
Accessibility	Whether users with special needs or disabilities can use the solution

Test data

To prove the accuracy of a solution's output, give it some test data to work on, and compare the solution's answer with one known to be correct.

Good test data includes:

- valid data – **data** that is perfectly acceptable, reasonable and fit to be processed.
- valid but unusual data – data that should not be rejected even though it seems odd. A 10-year-old might, once a century, enrol in university. Validation that rejected the young genius' enrolment would cause embarrassment.
- invalid data – to test the code's validation routines. For example, if people must be 18 years old to be given a credit card, test data should include people *under* 18 so they can be seen to be rejected.
- boundary condition data – data that is on the borderline of some critical value where the behaviour of the code should change. These 'tipping point' errors are a frequent cause of logical errors in programming.

Testing a multimodal solution

After designing and building your multimodal solution you need to demonstrate that it has been thoroughly tested. You need to know what to test in your MMOS. The following sections discuss these.

Media and plug-ins

Each picture, audio clip, video, graph and animation (that is, any non-textual information), must be inspected to see that it is displaying in the right place, at the right time, at the right speed and volume in a variety of common environments, meaning different browsers and devices.

Hyperlinks

Every internal and external link in the solution needs to be manually clicked and the result noted. Create a list of links and tick off each one as it passes testing.

Links to external services

You should be able to test all parts of the MMOS under your control completely, but many modern solutions rely on external entities to provide services such as site search boxes, advertisements, feedback forms, forums, shopping carts and payment facilities. You need to test the operation of all external connections to your product.

Readability

Checklist	Tick ✓
Is the text large enough to read comfortably on a small device?	☐
Is contrast optimal, or at least satisfactory?	☐
Is the typeface a readable size?	☐
Are lines or paragraphs of a good length?	☐
Is text alignment attractive and readable on the page?	☐
Is the spelling, punctuation and grammar accurate?	☐
Is the vocabulary appropriate and inoffensive?	☐
Is expression clear and unambiguous?	☐
Are headings clear and do they divide content into logical sections?	☐

Calculations

If your solution calculates any information, its answers need to be verified by manual recalculation in a testing table. For example, you might insert a webpage containing JavaScript code to display a countdown timer to the next Informatics exam. To prove that you have tested the accuracy of its output, take a screenshot. Annotate the screenshot with whatever manual calculations will demonstrate that it is true, based on the time the screenshot was taken and the time of the exam.

Loading times

If the site is online, clear your browser's cache to remove pre-loaded copies of files and media and try loading the site via cable and wi-fi. Any page that takes more than a few seconds to load should be inspected and optimised. Another method is to use one of many online services that can measure the loading times for your pages.

Browser compatibility

A multimodal solution must rely on its browser, plug-ins and installed players and codecs (coder/decoder or compressor/decompressor) to read and display its media. Browsers differ in their ability to interpret different media, and some systems may not have the right technology, such as HTML5, or the necessary plug-ins installed, such as Flash player or Silverlight. Every piece of media must be checked on the dominant browsers to verify that they appear as they should.

You can test most site functionality yourself manually, but if your MMOS is online there are many services that can perform automated cross-browser compatibility checks using many versions of new and previous browser versions.

CSS

Test that CSS is consistently formatting your solution so items like text, headings, links, frames and page backgrounds look right throughout the product.

Accessibility

Does your solution create unnecessary difficulty for users with poor eyesight or muscular control, or weak language skills, or other common disabilities? Is alt text applied to images? Are colour combinations considerate of colour-blind people?

There are several places online to test the accessibility of your MMOS. Try the W3C website.

Dynamic features

Every menu item must be selected and its behaviour documented in a testing table (Table 4.14). If data entry forms are expected to work, data should be entered and its successful arrival at its destination should be documented. Simulated functionality, such as a faked login box, should, as far as is practical, appear to work genuinely. All coding, such as JavaScript, Java, PERL, macros and Python, should be run using a variety of test data and the behaviour of the code recorded.

Load capacity

You might hope otherwise, but your MMOS is unlikely to receive so much attention that it cannot cope with the huge amount of global interest it receives.

However, some sites need to be load tested to verify that they can stand up to enormous workloads. There is software and services that will simulate and evaluate stress applied to web servers, bandwidth, storage capacity, databases and server-side programming code.

Testing table

A testing table is a commonly used way to record evidence of functionality testing. A testing table for a website may look like this.

TABLE 4.14 A testing table

What was tested	How it was tested	Expected result	Actual result	How it was fixed, if relevant
Embedded YouTube video	Loaded page and clicked play button	Play cow video smoothly	✓	N/A
Links to pages for contacts.html, bibliography.html, conclusions.html, and data.html	Clicked each link on 'index.html' (links on other pages were copied from here, so if they work on homepage, they will work everywhere else)	Go to linked pages	All worked except for data.html	Changed link text in source code from 'data.htm' to 'data.html'
Readability	Asked two volunteers to read sample pages and report on text size, contrast, alignment, spelling, vocabulary, offensiveness, and headings	Reports that pages were easy to read, accurate and inoffensive	One reader suggested a paragraph in italics was hard to read.	Changed paragraph's text style from italics to bold

TABLE 4.14 (continued)

What was tested	How it was tested	Expected result	Actual result	How it was fixed, if relevant
JavaScript calculation of number of days since the page was created	Set computer's clock to 7 days after the page creation date	Index page should display 'This page was created 7 days ago, so some external links may be no longer accurate.'	Displayed '... created 6.89586 days ago ...'	Rounded up the age calculation in JavaScript
Loading times	Performed website speed test on each page	Performance grade of at least 75	All pages were graded 77/100 – 'faster than 87% of all tested websites'!	N/A
Browser compatibility	Load all pages in: • Chrome, Firefox and IE (all on PC and a tablet) • Safari (on iPhone)	All pages appear as intended in the site's design	• Chrome v.41: all pages good • Firefox v.35: all pages good • IE v.11: index.html containers not aligned as they should be • Safari: pages OK, but controls too small	• IE: Had to change HTML to float the <i>li</i> elements. This didn't upset the other browsers during re-testing • Safari: increased control sizes 20% to make them more finger-sized
CSS	On all pages, inspect the elements defined in maincss.css	Same formatting on each page	✓	N/A
Feedback form	Enter sample feedback and click 'Submit'	Feedback to come to my email address	Feedback arrived (jumbled together, but it arrived)	Can't fix the jumbled data. May use a Google Form instead to make life easier.
Alt text	Used Windows Narrator to read each page	Alt text should be read by Narrator	All alt text was read OK	N/A

Classroom constraints

Your MMOS may, because of constraints in classrooms and networks, not actually be online, although it must be capable of being online. However, it should have the look, feel and apparent functionality of a real online solution, even though some features may have to be simulated because it is unreasonable to expect them to function under your working conditions.

How to document your testing

- Use a testing table.
- Seek a subjective report from a fellow student who tried out your solution's readability and useability.
- Capture screenshots of features that are not normally visible, such as dropdown menus and warning messages, showing that they work when needed.
- Make handwritten calculations annotating printouts of screenshots of your solution's calculations to verify that the output has been checked for accuracy.
- Capture screenshots of the solution's validation rules responding properly to invalid data.

Evaluating your solution

Evaluation is the final stage of the problem-solving methodology. It checks how well the solution is satisfying the needs of the user for which it was originally created.

Remember you will be evaluating your MMOS and your project plan as part of your assessment in Unit 4, Outcome 1. When evaluating your MMOS, you need to refer to the evaluation criteria you developed during the design phase. For each criterion, you will choose a method to use for its evaluation.

Evaluation is not the same as testing; its purpose is distinctly different. By the time evaluation begins, the solution has already been proved to work properly and its functionality is no longer in question.

Evaluation can best be understood by saying what it does not do.

- Evaluation does *not* test that a solution is working properly. That should have been done during testing.
- Evaluation does *not* enter test data to check that output is accurate. That should have been done during testing.
- Evaluation does *not* use a stopwatch to time how long a process takes. That should have been done during testing.
- Evaluation does *not* perform checks with immediate results, such as pulling out the power plug to see if a system loses data. That should have been done during testing.

Evaluation looks at a solution's performance *over time* in terms of the **evaluation criteria**.

What to evaluate

Evaluation criteria are determined during the design phase of the problem-solving methodology and are based on the most important qualities that the solution is expected to have when it is designed. For example, for your solution, essential criteria include appeal to a global audience and adequate substantiation of the conclusion of the results of your hypothesis. For an accounting package, security and accuracy might be most important. You should evaluate the features that would, if they were not achieved, render your MMOS unsatisfactory.

Evaluation criteria fall under two headings: efficiency and effectiveness:

- 1 Efficiency can be measured in terms of speed or productivity (work produced in a given time), profitability (income generated versus running costs) and labour requirements (how much labour is required to achieve its productivity levels).
- 2 Effectiveness includes completeness, readability, attractiveness, clarity, accuracy, accessibility, timeliness, communication of message, relevance and useability (Figure 4.2).

You will probably not have access to months of error logs, site visitor statistics or financial records related to your solution, but you may conduct interviews or create questionnaires to seek feedback from your peers that is relevant to the criteria you were expecting your MMOS to achieve, such as communication of message, readability, accuracy, clarity and so on. This feedback could be offered as evidence of evaluation.

You may decide to use annotated screenshots to demonstrate how key criteria were satisfied, such as readability, or consistency of formatting. If recordings of animated screen activity are needed to prove you satisfied an evaluation criterion, you could use screen recording software such as Camtasia or Camstudio. To demonstrate that you achieved appropriate readability levels, you could include a screenshot of your text's readability score, as calculated by Microsoft Word.

Evaluation methods

For each evaluation criterion for your MMOS, there must be a corresponding evaluation method that can measure the degree to which the criterion has been achieved.

- **Objective** (fact-based, measurable) results are solid facts that are hard to argue with. Measure whenever you can.
- **Subjective** results (emotions, opinions, personal judgements) can be gained from interviews, questionnaires and surveys. These should only be used when objective measurement is not possible or practical, such as to evaluate how comfortable users feel when using a multimodal solution.

TABLE 4.15 Typical evaluation methods

Criterion	Method
Accuracy (effectiveness)	Check the complaints log and count the complaints from staff or customers about inaccurate information received over three months from the system.
Reliability (effectiveness)	Count the number of faults in the system's error log.
Security (effectiveness)	Count the number of successful and thwarted attempts made to penetrate system security.
Attractiveness, pleasure, comfort, confidence (effectiveness)	Interview users.
Productivity (efficiency)	Refer to system logs and count how many transactions the system handled over three months compared to the previous system.
Profitability (efficiency)	Ask the accountants to tally the new system's running costs over time. Check organisational profit figures and see if it has increased.
Labour requirements (efficiency)	Count the number of staff hours spent operating and maintaining the system compared with the previous system.
Ease of use, useability (effectiveness)	Count the number of times online help was used (indicating that the solution may not have been intuitive). Add up how many errors were made by users. (A solution that is hard to use tends to cause users to make mistakes.) Check the help desk records to see how often users asked for help or complained about the solution. Give users a questionnaire asking about their feelings regarding the system's useability.

Remember: Evaluation assesses your solution's performance over time. It is not instantaneous the way that testing is. Any emotional or judgemental feedback is only gathered on appropriate criteria. For example, it is pointless to ask interviewees questions such as, 'Is the new system faster than the old one?' Even if you received an answer to this question, you would not be able to trust its accuracy.

When to evaluate

Evaluation occurs after the solution has been in regular use for some time so it is well 'bedded in' and its users are familiar and comfortable with it. A few months of regular, daily use is typical.

Evaluating a solution too soon can lead to negative feedback if users are not yet used to it and are slow and prone to making errors. Later, when they are comfortable and skilled with the solution, their feedback may be much more positive.

In cases when a system is used infrequently, but its success is critical to the organisation (such as creating school reports, or managing a flood of tax returns at the end of the financial year), evaluation may be done immediately after the system is used.

Waiting is not a feasible option for your solution. You will not have the luxury of waiting for a long period of time for users to explore your MMOS. Instead, you will need to ensure that you evaluate your own solution especially thoroughly and take into account all of the feedback you received during beta testing, which should have included asking potential end users to look at your MMOS. (Potential end users could have included classmates, friends, your teachers, your parents and other family members.)

Assessing your project plan

Organisations invest a great deal of time, money and labour into projects, so they tend to look back at their project plans to evaluate how project planning went. Organisations often need to undertake further projects, so they need to keep lessons in mind from earlier projects to prevent repeating mistakes. Evaluating an organisational project plan can help answer the following questions.

- 1 Did the project finish on time?
- 2 What tasks delayed your project? Why were these delays not anticipated?
- 3 Could lessons be learned to help the next project finish on time?
- 4 Did the project finish on budget?
- 5 What assumptions did we get wrong?
- 6 Why did this task cost far more than expected? How can we avoid that next time?
- 7 Why were new requirements being added just weeks before the system was due to go online? Was our analysis a failure?
- 8 Why did the first three prototypes blow up? Was the design team under-skilled, overworked, under-equipped or working to an impossible deadline?

Even a failed project can be a valuable learning experience if it prevents the same mistakes happening again:

- In November 2012, the US Air Force scrapped a major enterprise resource planning software project called the Expeditionary Combat Support System (ECSS) after it cost \$1 billion in expenses but failed to create any significant military capability. ECSS was supposed to replace more than 200 **legacy systems**. The project dated to 2005 and its ballooning costs clearly suggest that Air Force officials and systems contractors conducted an overwhelming amount of additional custom coding and integration work. An Air Force spokesman said the project would require *another* \$1.1 billion just to complete 25% of the original scope, and that the project would not be complete until 2020.

On the other hand, it can also be an expensive mistake to *not* undertake projects to improve systems:

- As of 2012, Long Beach, California was owed \$18 million in parking fines as a result of antiquated software used by the local government. Fines had gone uncollected because of the age of the systems in use. Staff were stuck using manual processes that took up so much time that there was no time left to undertake collection efforts – leading to \$18 million in unpaid fines. Even worse: Long Beach first knew about the problem as of an audit in 2009, but still failed to address it (LA Times).

Your project may be on a more modest scale than these scenarios, but there are lessons that you can learn from them. The purpose of you assessing your project plan to is judge how your plan and the techniques you used to make adjustments along the way helped you to manage your project. For instance, how much notice did you ultimately take of your plans? How did

THINK ABOUT INFORMATICS 4.7

Identify by number the questions from the list opposite that evaluate an organisational project plan that could also be used to evaluate *your* project plan.

your plan assist you when things did not go as expected? Did the quality of your annotations or other tools you may have used keep you on track? These and other questions are of great use to you now, and they are a good idea to keep in mind for your future studies and later career, because they will help you to understand how planning functions alongside real-world projects.

You need to fully think through the scope of the tasks so you do not discover a forgotten task when your time has nearly elapsed and it is too late to finish it. Make sure your WBS is complete. Keep referring to your project plan so you do not forget tasks, do them in the wrong order, or fail to observe milestones.

You also need to think carefully about your software choices. It would be unwise to choose a multimodal authoring tool, only to discover halfway through development that it is unsuitable for the task. If you need to learn and practise new software skills, do it before the project begins. Learning new software while you work on a critical solution is very inefficient and you will be prone to making errors.

You also need to be realistic when you are estimating time requirements. A good thing to do is think back to times when you had similar tasks to complete. Remember what problems caused delays in completing those tasks and add time to your schedule to deal with them this time. If a task looks like it will run overtime, refer to your Gantt chart to see what consequences it will have on the project. If the task has slack built into it already, it may have no effect on following tasks. If it is a critical task, you will need to save time on other tasks, and you will need to find where and how this can be done.

When evaluating your project plan, you first need to establish the evaluation criteria that will indicate how successful it was in managing your activities and getting the project completed on time.

Such criteria may include the following.

- **Completeness:** Were any significant tasks omitted from the WBS? Were resources included? Was it annotated when required?
- **Maintainability:** How easy was it to modify the Gantt chart to keep it up to date with reality?
- **Accuracy:** Were tasks correctly identified and marked as dependent or concurrent? Were tasks in the right sequence? Were time estimates realistic?
- **Readability:** Was it easy to see all tasks and their dependencies? Was the chart and its text of a readable size? Were colour choices appropriate?

Once complete and relevant criteria have been chosen, you will again need a method to evaluate each one. Annotated printouts highlighting key features of the solutions may be useful. You might take screenshots before and after changing a task's duration to show how easy it was to maintain the chart. You can describe how well the chart worked to keep you on schedule to finish the solutions on time. You might offer lessons you learned from the projects that will make later projects even more successful.

Next steps

You should work your way through the chapter summary material, including 'Preparing for Unit 4, Outcome 1', and, in line with your teacher's instructions, begin to prepare your solution for submission.

ESSENTIAL TERMS

- algorithms** a description of a calculation strategy, such as a method of finding whether a year is a leap year
- annotate** add comments to a document
- archive** moving old files to offline storage instead of deleting them so they do not consume disk space but are available if necessary in future
- condescending** showing an attitude of superiority
- crop** to cut away parts at the edges of an image, making it smaller
- data** raw facts and figures
- design idea** brief, rough strategies for solving a problem; the best design idea will satisfy all functional and non-functional requirements, work within the defined constraints and be efficient
- evaluation criteria** rules set out during design that include effectiveness and efficiency criteria; based on the solution's requirements that were defined during analysis

IMPORTANT FACTS

- Multimodal solutions** use more than one type of data, such as text, images (still or moving), audio, animation and gesture.
- A successful **solution** must convey its message clearly and be easy to understand.
- Your multimodal online solution (MMOS) need not actually be uploaded to the internet, but it must be appropriate for a global audience and capable of being online.
- To improve commonality of language, use culturally neutral vocabulary and references in your MMOS that may be understood by other cultures. Use universally understood terms and units of measurement, such as date formats. Aim for clear, concise language that non-native English speakers can easily understand.
- A global audience will include readers of all ages. Try to write in a way that is approachable by as many of them as possible.
- A detailed design is produced by the design idea that best solves the problem.
- Design ideas may come from brainstorming, using outside consultants, talking to end users, mind mapping, graphic organisers, attribute listing, de Bono's Six Hats, research and looking at parts of the problem in different ways.
- Persistence and being observant increase the chances of devising the optimal design idea.
- Formats** are chosen methods of presenting information (e.g. tables, graphs, webpages).
- Conventions** are the standard ways of using formats (e.g. underlined web links) that facilitate information absorption.
- Design principles** affect the functionality and appearance of solutions.
- Solutions are tested with **test data** to fully exercise all conceivable types of data inputs.
- Good test data especially checks **boundary conditions** where the behaviour of a solution should change.
- The Gantt chart you created in Unit 3, Outcome 2 is continued in Unit 4, Outcome 1 to monitor progress and reflect changes in schedules.
- Gantt charts and project plans should be updated on an ongoing basis to reflect real events as projects proceed.
- Each evaluation criterion has a corresponding method.
- Evaluate by objective measurement whenever possible. Use interviews only to discover people's feelings, opinions and judgements.
- Do not evaluate too soon, or the results may be negative.
- Evaluate project plans to find ways to improve projects in the future.
- Poorly managed projects can be disastrously expensive, long running, disruptive or futile.



TEST YOUR KNOWLEDGE



Review quiz

AN EFFECTIVE MMOS

- 1 What is *colloquial* language, and why should it be avoided in an MMOS?
- 2 What techniques might you use to appeal to younger and senior audiences?
- 3 What types of media can be used in an MMOS?
- 4 List three ways in which an MMOS may be changed to improve its accessibility.
- 5 How can compatibility problems prevent an MMOS from conveying its message to a global audience?
- 6 Suggest two ways to make a webpage more responsive.
- 7 Choose the three factors that contribute most to a usable interface. Justify your choices.
- 8 List three types of software that could create an MMOS for your Informatics SAT. Choose the best one and explain why it is best for your circumstances.

MANIPULATING DATA

- 9 Explain the main benefits of using heading styles or CSS.

FORMATS AND CONVENTIONS

- 10 Explain the difference between formats and conventions.

DOCUMENTING AND THE PROGRESS OF PROJECTS

- 11 Why should Gantt charts be updated as projects proceed?

EVALUATING YOUR SOLUTION

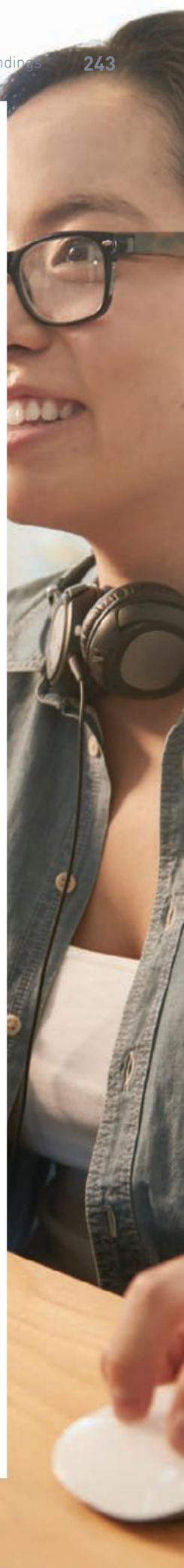
- 12 How is evaluation different to testing?
- 13 Give three examples each of effectiveness and efficiency evaluation criteria. Give an appropriate evaluation method for each criterion.
- 14 When should interviews be used during evaluation?
- 15 Why should project plans be evaluated after a project ends?

APPLY YOUR KNOWLEDGE

Create a multimodal, online solution (MMOS) that presents the information you generated about your hypothesis, and the conclusion this information led you to regarding whether the hypothesis is supported or not.

NOTES

- 1 While your MMOS may not actually be put online, it must be suitable for a global audience.
- 2 It should use appropriate formats for presenting data and information, and use appropriate conventions for each format.
- 3 It should apply relevant design principles to enhance its functionality and appearance.
- 4 Present your two or three hand written design ideas of the structure and appearance of the MMOS.
- 5 Present the details of your preferred MMOS design's structure and appearance.
- 6 Describe how you applied manual and electronic validation techniques.
- 7 Show evidence of testing, such as testing tables, screenshots of functionality testing, and test-user feedback.
- 8 Using a written report or an annotated visual plan, assess how effectively the Gantt chart helped you manage the development of the MMOS, and explain how the timeline was modified as the project proceeded.
- 9 Write a self-evaluation of your work during the development of the MMOS, describing how well the finished product satisfied its original aims, and what could be improved if you repeated the project.
- 10 Describe the procedures you used to manage all files relating to this SAT.



PREPARING FOR

UNIT 4 OUTCOME 1

Design, develop and evaluate a multimodal online solution that confirms or refutes a hypothesis, and assess the effectiveness of the project plan in managing progress

To achieve this Outcome you will draw on key knowledge and key skills outlined in Area of Study 1. This Outcome concludes the data acquisition, processing and analysis of the hypothesis formulated in Unit 3, Outcome 2.

OUTCOME MILESTONES

- 1 Design the MMOS, including two or three feasible alternative design ideas, which are rough sketches or mockups of development strategies, without much detail. Choose one design idea and add detail to create a complete design of your preferred design.
- 2 Develop an educational MMOS intended for a global audience that communicates the confirmation or refutation of the hypothesis you developed and researched in Unit 3, Outcome 2.
- 3 Evaluate the effectiveness of the solution.
- 4 Assess the effectiveness of your project plan in monitoring your project's progress.

STEPS TO FOLLOW

- 1 Use appropriate software and manual design techniques to design the MMOS. Develop two or three feasible alternative designs, and a detailed design of your preferred design: a total of three or four design strategies. Design should include information about:
 - a appearance
 - b architecture (how it will be built)
 - c object naming
 - d data structure
 - e data validation
 - f data manipulation
 - g evaluation criteria that will later be used to evaluate your MMOS and your Gantt chart.

The alternative design ideas need not be as detailed as your preferred design.
- 2 Develop the MMOS, including:
 - a building the actual MMOS
 - b performing a manual and/or electronic validation of the input data
 - c thoroughly testing that the MMOS works as intended.
- 3 Evaluate the effectiveness of the MMOS using the criteria established during design. Note that this is *not* the same as testing: you should judge how well the finished MMOS succeeds in satisfying the requirements of the evaluation criteria. In a written report or an annotated visual plan, assess the effectiveness of your project plan (Gantt chart) in monitoring your progress during the entire project from the beginning of Unit 3, Outcome 2.

DOCUMENTS REQUIRED FOR ASSESSMENT

- 1** A folio of two or three alternative design ideas and the detailed design specifications of the preferred design
- 2** A multimodal online solution that communicates the confirmation or refutation of a hypothesis as detailed in Unit 3, Outcome 2
- 3** An evaluation of the effectiveness of the MMOS
- 4** An assessment of the effectiveness of the project plan (Gantt chart) in monitoring project progress in a written report or an annotated visual plan

ASSESSMENT

Your teacher will provide you with a more detailed set of assessment criteria before you begin this assessment.

The SAT (comprising Unit 3, Outcome 2 and Unit 4, Outcome 1) will contribute 30% to your Informatics study score.

INFORMATION MANAGEMENT

Key knowledge

After completing this chapter, you will be able to demonstrate knowledge of:

Interactions and impact

- reasons why data and information are important to organisations, including meeting the goals and objectives of both organisations and information systems
- reasons why information management strategies are important to organisations, including maximising opportunities, minimising risks and fulfilling legal requirements
- key legislation that affects how organisations control the storage and disposal of their data and information: the *Privacy Act 1988*, the *Privacy and Data Protection Act 2014* and the *Health Records Act 2001*
- ethical dilemmas arising from information management practices
- strategies for resolving legal and ethical tensions between stakeholders arising from information management practices.

For the student

With so much information now in an electronic format, there is growing interest in how it can be stored securely and protected. There are also many legal and ethical considerations that affect how organisations store and dispose of information. Individuals and organisations have legal obligations they need to follow in Australia, such as the *Privacy Act 1988*, *Privacy and Data Protection Act 2014* and *Health Records Act 2001*. In this chapter, we will consider how information systems can be used to meet organisational goals, and provide an overview of legal considerations that have an impact on how organisations manage information.

For the teacher

This chapter provides part of the background information required to complete Unit 4, Outcome 2. Chapter 6 covers the rest of the concepts for this Outcome. The focus on Area of Study 2 is on information management and its importance to organisations. Students investigate how different organisations store and dispose of their data and information, and information management strategies to protect the integrity and security of data and information, taking into account key legal requirements of organisations and any ethical dilemmas faced by organisations and individuals.

Importance of data and information to organisations

From data to information

Data and information are terms frequently used in information systems. Indeed, the main purpose in using computers to solve information problems hinges on the difference between data and information. The term data refers to the raw, unorganised facts, figures and symbols fed to a computer during the input process. Data can also mean ideas or concepts before they have been refined.

As discussed in Chapter 2, information is obtained when data is manipulated by the computer's processor into a meaningful and useful form (also becoming its output). This can be achieved by organising the data and presenting it in a way that suits the needs of the intended recipient. Information is then used to assist in decision-making.

Characteristics of information

A raw piece of data, for example, might be the temperature measurement 22°C. On its own, this data does not tell us much. We do not know if this is a measurement of the current temperature, the predicted maximum, the minimum recorded for the day or simply the temperature of a cup of coffee. If, however, the data is added to other data and processed by a computer, we can produce a chart showing the range of temperatures experienced during the day. The chart provides information on how the temperature has varied during the day. A user can read the chart to find the highest and lowest recorded measurements and the range of temperatures experienced during the day. This information can then inform or persuade the user to follow a particular course of action. For example, a farmer may decide to harvest a crop over the next few days based on favorable weather conditions, a person may be persuaded to go surfing or a storekeeper may be influenced to stock up on winter clothing.

Information can be presented in a variety of forms. It may be paper-based (hard copy) or digital (soft copy). Information must conform to various criteria. It certainly must be comprehensible to the user. In fact, it can only be considered to be information if the user can understand it.

Information needs to be more than just comprehensible. Effective decision-making is predicated on information possessing certain qualities.

Complete information

It is also important to include all the relevant information. A report is not complete if the intended user has to find additional information. Sometimes it can be difficult to determine exactly what information is required.

Timely information

Information must be timely to serve its purpose. There is no point, for example, in receiving notification of a family gathering the night after the event has been held.

Accurate information

Information must be accurate. Incorrect information will lead to the wrong decisions being made or problems remaining unsolved. Even if the source data is correct, information produced by a computer may be inaccurate.

Information in a document or report needs to be relevant or the document can be misleading.

a

Shutterstock.com/
Andrey_Popov

b



Alamy/PAF

FIGURE 5.1 Payroll must use accurate information, or wages will be incorrect; **a** today, information systems that include payroll staff and appropriate hardware and software are used, but **b** throughout history, different systems, including the paymaster machine, have been used.

Any decisions based on flawed information will have serious consequences for the organisation or person that makes them. In many cases, the person who receives the faulty information is unlikely to know that it contains errors. For example, think about the payroll staff member who uses a spreadsheet to calculate the pay of sessional staff members. If the information he or she enters into the spreadsheet is inaccurate (for example, not taking into account that a sessional staff member is on a different rate due to their seniority), the information produced by the computer will be inaccurate. This issue will not only impact the pay of the individuals affected, but will also impact on the overall budget allocation for sessional staff.

Unbiased information

It is vital that the output from an information system remains **unbiased**. This means that the process must be carefully designed. **Biased** information will adversely distort the decision-making process and inhibit the effectiveness of any solution implemented by the user. When designing output, it is important to avoid situations that could lead to a biased interpretation, and efforts should be made to make users aware of any possible bias in the information that they receive.

Bias through sorting

The way in which lists are sorted can introduce bias. In many cases, this is unavoidable, as any list must have a beginning and an end. For example, in a chat program with 250 personal contacts, the contacts can be sorted in order of popularity rather than alphabetical order. That is, the list is sorted based on how often contact is made with each party. Therefore, the contacts that are most popular are listed at the top of the list, and those that are not often used fall towards the bottom of the list.

Bias through graphics

Bias can occur in selecting the graphic, the scale used and the size chosen. Graphics should be sized proportionally, to avoid overstating or trivialising the importance of any of the variables involved. In fact, representing certain aspects rather than others in graphic form can produce a biased effect. In Figure 5.2, both charts have the same identical data. However, the change of scaling on the units would suggest that the diagram on the right has an increased difference compared to the chart on the left.

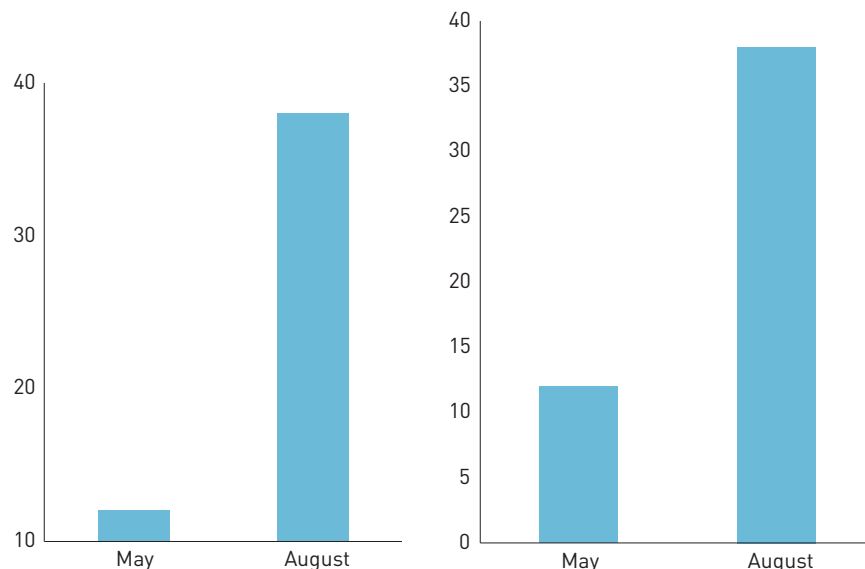


FIGURE 5.2 The two graphs are using exactly the same data. The graph on the left starts the scale at 10 and uses larger intervals. The graph on the right starts the scale at 0 and uses a finer scale.

Clarity of the message

If the intended message or conclusion of a report is not clear, the effectiveness of the message will be diminished, even if all the information is included. If that information is hard to find, the report is not effective. For example, financial reports issued by companies tend to be lengthy, complex and unwieldy. Investors and other stakeholders have become critical of these corporate reports as they lack clarity and conciseness. Even savvy readers of these reports can find it difficult to interpret due to the complex accounting jargon used. They can find it difficult to locate information. Given that it takes a lot of time and effort to prepare and verify the information in these reports, the criticism is that they are unresponsive and the clarity of the message is not clear.

We live in an information-driven society in which individuals and organisations have an insatiable thirst for information. Information is essential to businesses, governments, non-profit organisations and individual users. Organisations use information to support (persuade, inform and educate) their decision-making. Most decision-making in organisations is undertaken in order to solve problems whose solution will allow the achievement of stated goals. Combined with technology, the collection of vast amounts of data and information make it possible to do many things that previously could not be done, such as identify business trends, prevent diseases and assist with reducing crime. Data and information assist with connecting the dots and make better business decisions. Data and information are very powerful to organisations. It could be said that knowledge is power and that it can provide new business opportunities, increase profits and provide new insights into science and medicine.

Goals and objectives of organisations and information systems

Organisations tend to go through many changes over time. These changes are generally the result of a strategic plan. A strategic plan is a process for identifying long-term goals within an organisation. For example, a school is an organisation. A school needs to establish a strategic plan on how it intends to maintain or increase the current student enrolment level, introduce new courses and perhaps erect a new building. This type of planning looks beyond the day-to-day running of the organisation and concentrates on future developments that could range anywhere from two to 25 years.

Once an organisation has developed a strategic plan, a mission statement is developed based on the organisation's purpose, visions and values. The **mission statement** is the basis for establishing a set of common goals that will help accomplish the organisation's aims. These are known as organisational goals. See Figure 5.3 for an example of a mission statement.

Organisational goals explain how an organisation intends to go about achieving its mission. For example, a car manufacturer might identify its mission as increasing its market share and making a profit. Establishing goals of introducing a new model of car each year and providing the highest quality spare parts to customers will enable it to achieve that mission.

THINK ABOUT INFORMATICS 5.1

- 1 Can you think of any other type of organisation that would need a strategic plan and for what purpose?
- 2 Using the internet, search for two examples of a 'mission statement'.

Identifying organisational goals is important to understanding how and why organisations operate.

Mission statements concentrate on the present, whereas a vision statement focuses on the future. A mission statement explains the company's reason for existence. It describes the company, what it does and its overall intention. A vision statement describes the organisation as it would appear in a future successful position.

Mission:
'Inspire and develop the builders of tomorrow'
Our ultimate purpose is to inspire and develop children to think creatively, reason systematically and release their potential to shape their own future – experiencing the endless human possibility.

Vision:
'Inventing the future of play'
We want to pioneer new ways of playing, play materials and the business models of play – leveraging globalisation and digitalisation...it is not just about products, it is about realising the human possibility.

Statements reproduced courtesy of the LEGO Group

FIGURE 5.3 An example of a mission statement from Lego®

A school, for example, may establish targeted goals, such as the introduction of a vocational education course within two years, or the completion of a new science and technology centre in the next five years. To achieve these goals, the organisation needs to develop a list of **objectives**. Objectives are small achievable tasks undertaken to accomplish a big task. They have a set target to achieve; that is, they are measurable. For example, an objective may be to increase student enrolments by 10% within 12 months.

Every business has a different set of organisational goals. Some have financial goals, such as making large profits, while others want to be more competitive by increasing their market share. Depending on the type of organisation, the goals will be different. For example, non-profit organisations such as charities have goals that focus on providing services to members or people who may be disadvantaged in the community.

For-profit public organisations want to provide shareholders with maximum returns in the form of dividends and growth in share prices.

Achieving the goals of an organisation can often be feasible only by purchasing or modifying an information system. However, objectives determine the size and type of information system that is needed. Depending on the organisation's goals and objectives, one or more types of information systems (enterprise computing system, transaction processing system, business-support system, knowledge-management system and user-productivity system) may need to be developed.

If an organisational goal is not being reached, the organisation has a problem. Developing or improving an information system can help solve the problem and allow the organisation to achieve its goal.

Consider the needs of a simple organisation such as a sporting club. A small club might have only one or two goals. They might be as simple as keeping an accurate record of members' names and addresses and whether they have paid their subscriptions. As the club expands, it is likely that the goals and objectives will grow. Table 5.1 illustrates how the goals and objectives of an organisation can influence the type of information system that needs to be developed.

TABLE 5.1 Common goals of an organisation

Goal	Explanation
Increase the company's profit margin	Businesses exist largely to make money. To provide value to the owners (shareholders or owner/operator), allow for further growth and the realisation of opportunities, the business needs to increase its profits.
Expand the company	As businesses want to increase their profit margins, they may find that they need to grow. They may need to employ more people and build larger premises so that their production levels meet customer demand.

THINK ABOUT INFORMATICS 5.2

Locate mission statements from two competing companies, such as Pepsi Co and Coca-Cola Amatil.

1 How do their missions differ?

2 How are they similar?

Goal	Explanation
Provide quality service	In particular, non-profit organisations such as charities would perceive this to be one of their most important goals. They exist to provide service to people who are disadvantaged in the community. Department stores such as Myer would also regard this as an important goal, as excellent customer service is paramount to its existence.
Maintain confidentiality	Information stored about customers, products and the workings of a company needs to be protected by an organisation. Organisations need to ensure privacy and that all information is treated with confidentiality.

TABLE 5.2 The goals and objectives of an organisation can influence the type of information system that needs to be developed

Goal	Explanation
Improve the communication of events to members	• send letters to members and past players • develop a website • produce a regular newsletter
Maintain accurate and detailed membership records	• regularly check all current members' names and contact details • promptly add new members' details to the club's records • pursue overdue subscriptions • inform all members of subscription renewal procedures

Organisational goals and objectives will often relate to improving the efficiency or effectiveness of operations.

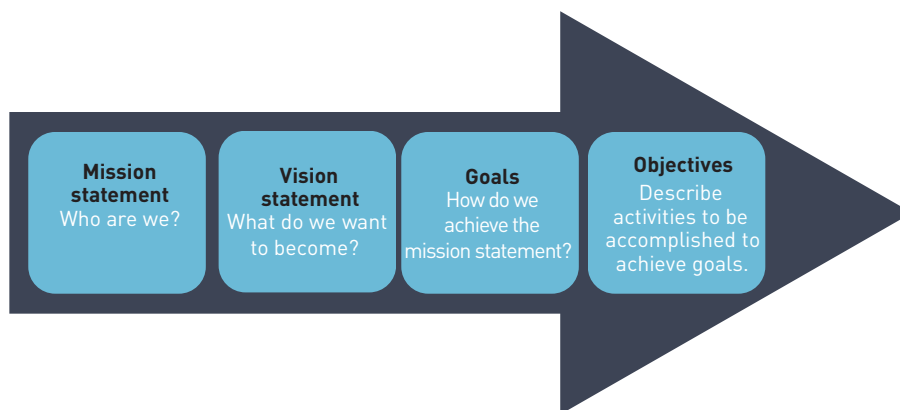


FIGURE 5.4 Relationship between mission statement, vision statement, goals and objectives

Information systems are often created to support the organisational goals. When planning the system, the systems analyst will identify a **system goal**. The system goal explains the specific role of the information system in achieving the organisational goal, and ultimately the company's mission. Setting up the right type of information system can help an organisation make improvements in efficiency, effectiveness and decision-making.

It is important that you are able to identify the mission and organisational and system goals in any organisation that you study. In particular, if a new information system is to be introduced, or an existing information system is to be modified, it must be designed to meet the system goal. If no system goal has been specified, you need to identify one. A new information system cannot be implemented if the system goal has not been previously articulated. You must ensure,

A systems analyst performs both the analysis and the design of a system. As well as being a good planner, a systems analyst must be a good communicator. As part of the analysis, a systems analyst is responsible for talking with users and translating their needs into a design that is passed on to programmers to build.

however, that the system goal reflects the aim of the organisational goal. If an information system is not designed to meet the goals of an organisation, it is not worth implementing.

Information management strategies

Information can be considered as a key asset in the same way as property or finance. The main function of information management is to support the business needs of an organisation by ensuring that vital information is available to support decision making. Organisations and governments recognise this and have formulated legislation and principles to guide the use of information.

Some of the issues that arise from using information include the following.

- How do we protect information?
- How do we store it?
- How do we destroy it?
- What are the correct uses of it?
- How much information is necessary and what might be surplus?
- What information do we own and how much can be shared?

Since 1998, in Australia, there has been legislation developed to manage information effectively. At the federal level, the *Privacy Act 1988* and, at state level in Victoria, the *Privacy and Data Protection Act 2014* and the *Health Records Act 2001* have been enacted. Periodically, the legislation is updated to reflect the technological advances that impact on information.

As organisations conclude that information should be managed as a company asset, they also recognise that managing information provides opportunities and minimises business risks. An organisation that facilitates good information management strategies has many advantages that can maximise opportunities, such as increased revenue and profitability. Advantages can also reduce costs by assisting organisations in optimising and streamlining their information assets.

Such information management strategies may also support decision-making, and identify appropriate organisational structures and governance that assist with effective information management. We can summarise these strategies into categories, such as those intended to maximise opportunities, minimise risks and to fulfil legal requirements.

Maximising opportunities

The large amount of data that is being gathered and analysed is very valuable to organisations. Large data sets (or 'big data') promote competition, productivity growth and innovation. The vast amounts of data and information are equally as important as labour and capital in an organisation. Organisations can collect detailed and large amounts of information on just about anything – from inventories and customer-buying patterns to product preferences.

Data analysis can boost organisational performance, because it encourages organisations to make smarter decisions based on the data. Organisations can forecast in the short- and long-term and adjust their business activities as needed. This can help them to achieve their organisational goals. They can also maximise opportunities so that they can tailor products or services to their customers. For example, Woolworths collects data through their loyalty card. Every time a customer scans their loyalty card during a purchase, Woolworths takes note of the items and this provides insight into customer information and data. All of the information collected *about* the customer is then used to create a one-to-one personalised experience *for* the customer (Figure 5.5).

The ability to find information easily can save an organisation time and resources. It can also reduce costs through removing duplicate information, which can lead to less searching time and updating of records, and removing unnecessary work.

The information Woolworths collects also provides the organisation with ample intelligence to pick the best locations to build its supermarkets in.

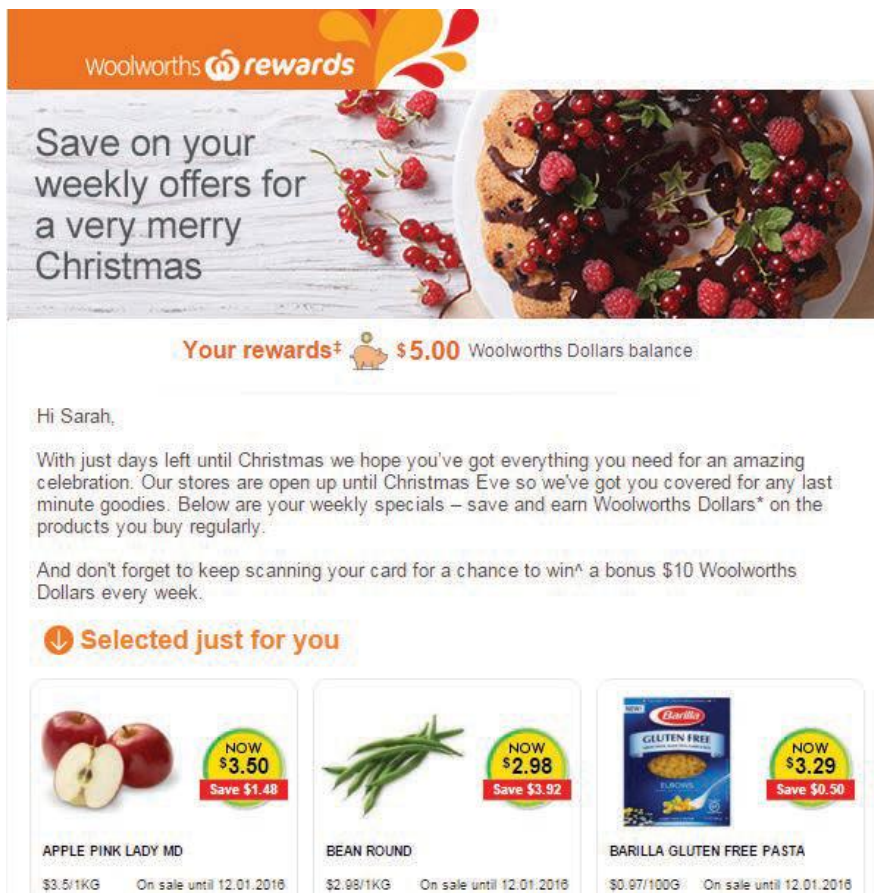


FIGURE 5.5 This email was sent specifically to a Woolworths loyalty cardholder because she has purchased these items in the past



Shutterstock.com/Evgeniia Trushkova

FIGURE 5.6 Data that you supply through loyalty cards is used by organisations to specifically target products based on your shopping habits

PREGNANT? BIG DATA IS WATCHING YOU

by Karl S. Kruszelnicki

Wherever we travel in our brave new digital world, we leave behind a rich and fertile trail of breadcrumbs of data. Our mobile phones continually tell the phone company where we are, our credit cards reveal how much we spent, and where, and on what – and so on.

Welcome to the world of 'big data', where the chain store Target has data-mined its way into... the uterus – and can now accurately predict when your baby will be born.

Back in the old days, a bookstore might slowly build up a base of loyal customers by offering discounts, and mailing out a brochure every month or so. But 'big data' is very different. So let's look at the three 'V's: volume, velocity, and variety.

First, volume. In the USA, Walmart, the giant chain store, harvests some 2.5 million gigabytes of data each hour – just from customer transactions.

Second, velocity. Today, we can analyse the data in real time. For example, you can quite accurately predict the sales figures of big shopping centres by monitoring the activity of mobile phones in the parking lot – even before customers have opened their wallets.

And thirdly, variety. The data can be GPS location data from your mobile phone, or it can be the entire contents of your address book with the phone numbers and birthdays of all your friends and their families. Big data also includes everything you have ever done on Facebook, Twitter and all the other social media – and let's not forget the good old credit card.

Back in 2002, Target in the USA wanted to know which of its shoppers had just become pregnant – because it wanted more customers. You see, shoppers tend to shop at the same place that they've always gone to – and it's hard to get them to shift. But when women get pregnant, it's a whole new world – and the pregnant woman might be lured into changing to a new shop.

So Target hired Andrew Pole (who has master's degrees in both statistics and economics) to see if he could data-mine his way into freshly pregnant uteruses. His starting point was the so-called 'guest ID number'. This is a number that Target assigns to each customer, and is linked to their name, email address or credit card.

It soon became clear that women (who had voluntarily put themselves on the Target Baby Registry) were buying some 25 products around the beginning of their second trimester. These included mineral supplements such as calcium, magnesium and zinc, large quantities of unscented lotions, and so on. And if they bought a little bright blue rug, the new baby was probably a boy.

And when they started buying scent-free soap, hand sanitisers, washcloths, a handbag large enough to carry a few nappies and extra-large bags of cotton balls – well obviously, they were getting close to their delivery date. Indeed, Target found it could predict their delivery date with some 83 per cent accuracy – just from their shopping habits.

So Target began sending discount coupons for various baby items to customers, according to how they rated on Target's own 'pregnancy prediction algorithm'.

But then it got messy.

An angry father walked into a Target store just outside of Minneapolis, clutching a brochure that had just come through the mail, demanding to speak to the manager. Angrily he said: 'My daughter got this in the mail! She's still in high school, and you're sending her coupons for baby clothes and cribs? Are you trying to encourage her to get pregnant?'

Sure enough, the brochure did carry the daughter's name and address, and had advertisements only for maternity clothing, nursery furniture, and so on. The manager apologised.

A few days later the manager called the father to apologise again. But this time, the father was a little embarrassed. He said: 'It turns out there's been some activities in my house I haven't been completely aware of. She's due in August. I owe you an apology.'

Now some people like their privacy. Some Target customers were annoyed by receiving brochures packed only with baby merchandise, especially when they had not told anyone they were pregnant. So Target produced new brochures that had some non-baby merchandise sprinkled around the baby goodies, so the newly pregnant mothers didn't realise they had been data-mined.

Target is happy. Their revenue jumped from \$44 billion in 2002, to \$67 billion in 2010 – helped mightily by the fact that with regard to pregnancy, Target knows, before it shows.

Now it's early days for big data. But one step might be to have more transparency and openness about what we (the customer) are giving away – and what we get back in return.

THINK ABOUT INFORMATICS 5.3

- 1 What are the advantages of data mining?
- 2 What opportunities can open up by using big data?
- 3 Why would a customer want to be tracked down based on what they buy at a particular store?
- 4 What are the privacy issues associated with data mining?

Minimising risk

Threats to data and information occur every day. Data can be compromised through activities such as deliberate attacks of theft (disgruntled workers or criminals that will make money selling the data), loss of devices (such as accidentally leaving laptops at airports or in a taxi), neglect (not erasing data when recycling computer hardware) and not following appropriate data handling procedures and policies. Collecting, storing, sending, encrypting, gathering and disposing of data is fraught with risk. To protect data and information, organisations need to manage such risks by putting measures into place, including:

- securing computers, servers and wireless networks
- utilising anti-virus and anti-spyware protection, and firewalls
- storing data backups off-site and ensuring that backups occur routinely
- securing passwords
- ensuring staff are familiar with digital systems policies and procedures
- becoming familiar with legal obligations pertinent to their needs.

Legal obligations

Every citizen of a country has a responsibility to follow its laws. A law is a rule that must be followed. Organisations have legal responsibilities as well. While Australia has had laws concerning the transfer of data via cable since the *Post & Telegraph Act* was passed in 1901, it has only been in recent years that significant changes have been made to the handling and release of data by government or independent agencies. Several laws have been introduced since 1988 as a reaction to the rapid spread of electronic communications and storage systems.

The popularity of technology has also led to the necessity of introducing legislation to protect data, particularly personal information. The definition of 'personal information' covers both information and/or an opinion about an individual who can be identified. There are no restrictions as to the quantity of personal information an organisation may collect, but organisations are prohibited from collecting and keeping personal information unless the information is essential for the organisation's activities. Additionally, criminal penalties exist for unauthorised access to, and disclosure of, credit reporting of personal information.

Under the 'Five Eyes' agreement, the US, UK, NZ, Canada and Australia share intelligence data.

ISSUE

Data retention laws pass Federal Parliament as Coalition and Labor vote together

Emma Griffiths, 26 Mar 2015, 8:19 p.m.

Contentious data retention laws have passed Federal Parliament, with both major parties voting for the legislation in the Senate.

The laws will force telecommunications providers to keep records of phone and internet use for two years and allow security agencies to access the records.

Companies already retain the data but for varying durations and in an unregulated environment.

The Coalition and Labor have argued the laws were necessary to help authorities in counter-terrorism and serious crime investigations.

Both major parties knocked back several amendments put forward by the Greens and concerned crossbenchers during Senate deliberations.

Labor announced last week that it would vote with the Coalition after the two parties agreed to several amendments, including specific protections for the phone and internet records of journalists, in a bid to protect anonymous sources and whistleblowers.

According to Dr Clinton Fernandes, there is no formal definition of metadata in Australian telecommunications law. A useful way to think about metadata is to describe it as machine-produced data. You control what you say and type, but not what the machine (e.g., a phone or computer) does to enable you to do so. When you make a phone call, your words are the content. What the phone does, to let you say it, is recorded as metadata. However, when you browse a webpage, the distinction between content and metadata disappears because it is all machine-produced (and machine-understandable) information about web resources, and therefore all metadata. Alternatively, it is all content! Examples include the email address, phone number, VoIP (Voice over Internet Protocol) number, the time and date of the communication, general location information (e.g. cell tower data), information about the duration of the communication, and the names and addresses (home, postal, billing) of the parties.

The Greens argued strongly against the law, saying it would entrench 'passive, mass surveillance'. Senators Nick Xenophon and David Leyonhjelm had also sought to change the legislation to increase privacy protections.

But Attorney-General George Brandis said the legislation, which passed with 43 votes to 16, would strike the right balance.

'It does contain safeguards that were not there before, it is in the Government's view, shared I'm pleased to say by the Opposition, a measured and proportionate response,' he said.

The cost of retaining the information is set to be partly covered by the taxpayer in what the Government described as a 'significant' contribution.

There are concerns telecommunications companies will pass on the rest of the cost to consumers. Telecommunications (Interception and Access) Amendment (Data Retention) Bill 2014

Introduced to Parliament on October 30 and redrafted on the advice of the Joint Committee on Intelligence and Security (which tabled its report on February 27). The legislation will:

- Require telecommunications companies to retain customer's phone and computer metadata for 2 years
- Define which types of data must be retained, such as phone numbers, length of phone calls, email addresses and the time a message was sent, but not the content of phone calls or emails and explicitly exclude internet browsing
- Detail which agencies are able to access the data
- Give security agencies access to the data when they can make a case that it is 'reasonably necessary' to an investigation
- Still require security agencies to obtain a warrant before accessing the actual content of messages or conversations
- Introduce an independent oversight mechanism, allowing the Commonwealth Ombudsman access to agency records, in a bid to boost privacy protections
- Give the Parliamentary Joint Committee on Intelligence and Security oversight of the use of metadata by the AFP and ASIO.

The government is negotiating with Telcos about who will pay for the new system.

ABC News, Dr Clinton Fernandes, 19 March 2015, Reproduced by permission of the Australian Broadcasting Corporation and ABC Online. (c) ABC. All rights reserved.

Privacy policies

An organisation that collects data on individuals or other organisations usually has a privacy policy about how it uses the data collected and to whom it will disclose it (Figure 5.7). By law, privacy policies must be located on a company's website so that they are easy to find. The policy must include details about what data is gathered and how it is used. As a consumer, it is important for you to read privacy policies, because some organisations may use your personal details in ways you might not expect.

THINK ABOUT INFORMATICS 5.4

- 1 Why is this legislation contentious?
- 2 Why is privacy protection important?

THINK ABOUT INFORMATICS 5.5

Can you locate three privacy policies from different industries and make a list of the differences and the similarities?

Key legislation for storage and disposal of data and information

There are several key laws relating to the information systems and telecommunications industries. At a federal level, the law concerned with how information about people can be used is the *Privacy Act 1988*. In Victoria, we are especially concerned with the *Privacy and Data Protection Act 2014* and the *Health Records Act 2001*. Combined, these laws govern the collection and use of private information by both government and non-government organisations at both state and federal levels. Employers and government agencies have a legal responsibility to ensure that these laws are implemented within their organisations. In addition, organisations must make employees and customers aware of their rights, as well as their responsibilities, in relation to these laws.

CLICK & COLLECT › Order online before 12PM for same day pickup in-store [find out more](#)

MYER Search     login | join

women men miss shop beauty kids & toys home electrical sale giftorium myer one

home - content - customer service - privacy policy

customer service

FAQs
Contact Us
Product Recalls
Customer Notices
Myer Stores Competitive
Pricing Policy
Returns Policy
Privacy

Privacy Policy

Myer Privacy Policy
Updated March 2014

[Print this](#)

1. Your privacy
Myer Pty Ltd (ABN 83 004 143 239) (Myer) is committed to protecting your personal information in accordance with the Australian Privacy Principles (APPs) in the *Privacy Act* as well as other applicable legislation, such as the *Spam Act*, the *Do Not Call Register Act*, and health privacy legislation.

This Privacy Policy is available on the Myer website at www.myer.com.au, or you can request a copy by contacting our Customer Service Centre. **MYER one** members are covered by a separate **MYER one** Privacy Policy, which includes the MYER one App.

2. What we collect and how we collect it
The personal information we collect about you depends on the dealings you have with us. Personal Information includes information about an individual who is reasonably identifiable. For example, if you:

- **make an in-store purchase**, we may collect your credit card, identity and transaction details in person and, if you use our delivery service, your name, contact details and delivery address. Certain departments such as intimate apparel or cosmetics etc may collect certain sensitive information about you in order to provide the goods or services
- **pre-order goods or services, put goods on hold, use our lay-by service, leave goods for repair or purchase certain electronic devices from us**, we may collect your name and contact details
- **provide details to join special mailing lists or databases** if you'd like to receive information about, for example, certain brands, products or designers, we may collect your name, contact details, date of birth and information about your shopping or product preferences
- **attend a store**, we may capture images and/or sound recordings of you through CCTV cameras
- **make a purchase online or using an electronic device such as an iPad when in-store or when you otherwise set up an online account**, we will collect your name, contact details, order details, credit card details, delivery address or store pick up details and we may also ask you to provide your date of birth and gender

Related links

- [Customer Notices](#)
- [Privacy Policy](#)
- [Website terms](#)
- [Team member sign in](#)
- [Supplier sign in](#)
- [Supplier portal sign in](#)
- [Contact Us](#)

FIGURE 5.7 Myer has its privacy policy on its website

This section examines the key laws affecting the storage and disposal of data and information held by organisations.

Privacy Act 1988

In the mid-1980s, the federal government attempted to introduce an 'Australia Card'. Due to overwhelming resistance from the public, this proposal was dropped. In its place, the pre-existing tax file number (TFN) system was updated. There were many concerns about how the federal government might use tax file numbers, especially regarding the release of confidential tax information about individuals, or matching data from different government departments. At the same time, the government was required by international laws to provide adequate protection for personal data being sent to other countries and to ensure that civil rights with regard to privacy were not being ignored.

‘These are the most significant changes to privacy laws in over 25 years and affect a large section of the community. The world has changed remarkably since the late 1980s when the Privacy Act was first introduced, and so the changes were required to bring our laws up-to-date with contemporary information handling practices, including global data flows,’ said Australian Privacy Commissioner Timothy Pilgrim.

Australian Government,
Office of the Australian
Information Commissioner

Changes made to the Privacy Act were enacted to increase consumer protection and benefit society, and for businesses and the Australian Government to be transparent on how they handle personal information, due to the modern and adaptable technologies available, such as cloud technology.

Originally, the *Privacy Act 1988* only dealt with the handling of data by government agencies. Many people criticised these limitations because it seemed that private organisations were not required to apply even the most basic of safeguards on data they collected. Even worse, there were no regulations preventing non-government organisations from collecting data by any method and using it for any purpose without the consent of the people whose private details it concerned. In particular, the rapid growth of electronic transactions, especially over the internet, led many people to demand some sort of legal protection from those who might gather data about internet browsing habits. The government was keen to encourage the development of electronic commerce, while protecting the confidentiality of consumers and increasing public confidence in electronic transactions. These amendments have now been incorporated into the *Privacy Act 1988* and are the most significant changes to privacy laws since the inception of the legislation.

There have been several additional powers included within this Act since 1988, but its main purpose has remained unchanged. The *Privacy Act 1988* was amended by the *Privacy Amendment (Enhancing Privacy Protection) Bill 2012*, which came into effect on 12 March 2014.

CASE STUDY

AUSTRALIA CARD

During 1985–86, the Federal Government tried to introduce a single identification number for taxation, national health insurance and welfare payment purposes. Each Australian was to carry an identity card bearing this number. Its purpose was to cut down on the number of people avoiding tax or claiming payments to which they were not entitled.

There was public concern over invasion of privacy and civil liberties. People believed that the card system would easily be avoided by criminals and that its only outcome would be to increase government power and control over the population. Rallies, petitions, and problems with passing the Bill through the Senate eventually forced the government to drop its controversial legislation. Interestingly, the prospect of a national ID card was again raised by the Federal Government in early 2006, although at that stage there was less initial support for the proposal than there was in the mid-1980s.

Tax File Number system

The TFN system existed before the Australia Card proposal. The TFN was used to identify individuals on their tax returns.

In 1988, after the failure of the Australia Card, the government updated the TFN system, but promised that it would still only be used for the purpose of processing taxation. Nevertheless, within a few years people had to supply their TFN in order to receive any payment or pension from the government. In effect, the TFN has fulfilled the role of the Australia Card, in that it is now an identity number, but without the card.



FIGURE 5.8 Measures to protect individuals and organisations from privacy and data security breaches are being taken very seriously

What is included in the Privacy Act?

The Privacy Act includes the following.

- Thirteen Australian Privacy Principles (APPs) that apply to the handling of personal information by most Australian and Norfolk Island Government agencies and some private sector organisations.

- Credit reporting provisions that apply to the handling of credit-related personal information that credit providers are permitted to disclose to credit reporting bodies for inclusion on individuals' credit reports.
- The collection, storage, use, disclosure, security and disposal of individuals' tax file numbers.
- The handling of health information for health and medical research purposes in certain circumstances, where researchers are unable to seek individuals' consent.
- The Information Commissioner to approve and register enforceable APPs codes that have been developed.
- Providing a small business operator, who would otherwise not be subject to the Australian Privacy Principles (APPs) to opt-in to being covered by the APPs.

Who is covered under the Privacy Act?

For an individual, the Privacy Act provides more control over the way their personal information is handled. The Privacy Act allows individuals to:

- know why personal information is being collected, how it will be used and who it will be disclosed to
- have the option of not being identified, or the use of a pseudonym in certain circumstances
- ask for access to personal information (including health information)
- discontinue receiving unwanted direct marketing
- ask for personal information that is incorrect to be corrected
- make a complaint about an entity covered by the Privacy Act, if personal information has been mishandled.

Australian Privacy Principles

As part of the Privacy Act, the Australian Privacy Principles (APPs) were devised to set out the standards, rights and obligations for collecting, handling, holding, accessing, using, disclosing and correcting personal information.

The Australian Privacy Principles generally apply to federal government agencies. They do not apply to local councils, or state or territory governments. Some states have their own privacy laws, such as Victoria's *Privacy and Data Protection Act 2014*.

The APPs oversee the handling of personal information by:

- Australian and Norfolk Island government agencies
- all private health service providers and
- businesses that have an annual turnover of \$3 million or those that trade personal information.

TABLE 5.3 The Australian Privacy Principles

APP 1	Open and transparent management of personal information Ensures that APP entities manage personal information in an open and transparent way. This includes having a clearly expressed and up-to-date APP privacy policy.
APP 2	Anonymity and pseudonymity Requires APP entities to give individuals the option of not identifying themselves, or of using a pseudonym. Limited exceptions apply.
APP 3	Collection of solicited personal information Outlines when an APP entity can collect personal information that is solicited. It applies higher standards to the collection of 'sensitive' information.
APP 4	Dealing with unsolicited personal information Outlines how APP entities must deal with unsolicited personal information.

APP 5	Notification of the collection of personal information Outlines when and in what circumstances an APP entity that collects personal information must notify an individual of certain matters.
APP 6	Use or disclosure of personal information Outlines the circumstances in which an APP entity may use or disclose personal information that it holds.
APP 7	Direct marketing An organisation may only use or disclose personal information for direct marketing purposes if certain conditions are met.
APP 8	Cross-border disclosure of personal information Outlines the steps an APP entity must take to protect personal information before it is disclosed overseas.
APP 9	Adoption, use or disclosure of government related identifiers Outlines the limited circumstances when an organisation may adopt a government related identifier of an individual as its own identifier, or use or disclose a government related identifier of an individual.
APP 10	Quality of personal information An APP entity must take reasonable steps to ensure the personal information it collects is accurate, up-to-date and complete. An entity must also take reasonable steps to ensure the personal information it uses or discloses is accurate, up-to-date, complete and relevant, having regard to the purpose of the use or disclosure.
APP 11	Security of personal information An APP entity must take reasonable steps to protect personal information it holds from misuse, interference and loss, and from unauthorised access, modification or disclosure. An entity has obligations to destroy or de-identify personal information in certain circumstances.
APP 12	Access to personal information Outlines an APP entity's obligations when an individual requests to be given access to personal information held about them by the entity. This includes a requirement to provide access unless a specific exception applies.
APP 13	Correction of personal information Outlines an APP entity's obligations in relation to correcting the personal information it holds about individuals.

Australian Government, Office of the Australian Information Commissioner

Credit reporting provisions

There have been changes to the credit-reporting provisions of the *Privacy Act 1988* and how credit-related personal information is collected. The Privacy Act also encompasses a code of practice for credit reporting. The credit-reporting provisions for consumer credit include the simplification of the language used in reports and improved privacy protections. Additionally, the process to lodge a complaint has been simplified.

Application of the Privacy Act

The Privacy Act applies to both electronic and manual or conventional forms of data gathering and handling by private organisations. The Act also has provisions specifically addressing the use of personal data for direct marketing via email, which can only be used with the consent of the individual concerned. It also extends to general privacy issues regarding workplace email. The Act encompasses businesses with an annual turnover of \$3 million, all private health services that store health records, businesses that trade in personal information and those organisations that choose to opt-in.

Individuals also have rights under the Act, which makes for provisions on how their personal information is collected. The Act defines personal information as being:

information or an opinion about an identified individual, or an individual who is reasonably identifiable: whether the information or opinion is true or not; and whether the information or opinion is recorded in a material form or not.

The amended Act defines personal information as including an individual's:

- name and address
- signature
- telephone number
- date of birth
- medical records and health information
- bank account details
- photos and videos
- biometric and genetic information
- philosophical beliefs
- likes and dislikes
- opinions or commentary about a person
- racial or ethnic origin
- memberships of political associations
- professional or trade associations or trade unions
- religious beliefs or affiliations
- criminal record
- sexual orientation or practices.

Since the inception of the updated *Privacy Act 1988*, organisations have had to consider and review how they handle customer information by updating their technologies and their security processes to ensure they comply with the new legislation.

The Act prescribes severe penalties for serious and repeated interferences with privacy which can result in criminal prosecution and or fines of up to \$340 000 for individuals and \$1 700 000 for public and private organisations.

ISSUE

Privacy breach: Medical records kept in garden shed

Tuesday, 15 July 2014

The Australian Privacy Commissioner, Timothy Pilgrim, has found a medical centre in Melbourne in breach of the *Privacy Act 1988* by failing to take reasonable steps to secure sensitive medical records.

The Office of the Australian Information Commissioner's (OAIC) investigation established that Pound Road Medical Centre (PRMC) stored medical records of approximately 960 patients in a locked garden shed at premises no longer operated or staffed by them. In November 2013, the shed was broken into and the medical records were compromised.

The Privacy Commissioner noted the seriousness of the case particularly as the records contained sensitive personal information such as full name, address, date of birth, Medicare number, treatment details including results of medical investigations and discharge summaries.

'The Privacy Act requires organisations to take reasonable steps to protect the personal information of their customers. I can't think of any circumstances in which it would be reasonable to store health records, or any sensitive information, in an insecure temporary structure such as a garden shed,' Mr Pilgrim said.

The Privacy Commissioner also warned organisations about the importance of secure document storage.

'Physical security of hard copy documents is just as important as digital security. There is no point in converting paper records to a secure digital system, and then leaving the paper files unsecured. If paper records are no longer needed, they should be disposed of securely,' Mr Pilgrim said.

Incorrect disposal of digital systems equipment and data storage media is a major source of privacy breach.

THINK ABOUT INFORMATICS 5.6

Many organisations have a privacy policy listed on their website. Find out what is covered by your school's privacy policy. What information might the school have about students that should not be made publicly accessible?

The amount specified in the Privacy Act is not variable. That is, the fine of \$340 000 for individuals and \$1 700 000 for public and private organisations does not change. The amount will only change if the Act is changed. In Victoria, however, the fines are variable and tied to penalty units with an amount that is adjusted on 1 July each year.

THINK ABOUT INFORMATICS 5.7

- 1 Why is it important to have processes in place to dispose of personal information securely?
- 2 What would the thieves who broke into the shed have to gain from looking at the medical records?
- 3 Describe appropriate methods to dispose of paper-based and digital medical records.

When a local council uses contracted service providers, including sub-contractors such as plumbers to repair public toilets, they too must comply with the Act.

The Commissioner for Privacy and Data Protection was established on 17 September 2014 under the *Privacy and Data Protection Act 2014*.

Schools have a legal obligation to ensure that they store sensitive data safely. This includes credit card information of parents.

These APPs replaced the two sets of principles, which, since 2001, have applied to Commonwealth public sector and private sector organisations. They were known as the Commonwealth Information Privacy Principles and the National Privacy Principles.

Non-compliance will result in a maximum penalty for a body corporate of 3 000 penalty units and 600 penalty units for an individual.

Penalty units define the amount that needs to be paid for offences in Victoria. Generally, the legislation does not specify the monetary amount, but does however specify the penalty unit. Each year, the penalty unit is specified, such as from 1 July 2015 to 30 June 2016; one penalty unit is worth \$151.67. The rate for penalty units is indexed each financial year so that is raised in line with inflation. Changes to the value of a penalty unit take effect on 1 July each year.

The majority of the medical records related to individuals who stopped being patients of PRMC before 2004. The Privacy Act requires organisations to securely destroy or de-identify personal information that they no longer require.

‘If organisations don’t need to keep personal information for a legal purpose, then they must have a system in place to dispose of it securely. Get out the shredder or hire a secure document destruction service. If you don’t, you’re putting your clients at risk of identity theft or fraud, and your company at risk of enforcement action.’

- The Commissioner recommended that PRMC:
- undertake a risk assessment in relation to records management and privacy practices
 - organise privacy training for all staff, and
 - develop a data breach response plan to assist with future incidents.
- PRMC are in the process of implementing these recommendations.

Australian Government, Office of the Australian Information Commissioner

Privacy and Data Protection Act 2014

The *Privacy and Data Protection Act 2014* (PDPA) was introduced by the Victorian Government. It replaced the *Information Privacy Act 2000* and the *Commissioner for Law Enforcement Security Act 2005*. The PDPA is intended to strengthen the protection of personal information and other data held by Victorian government agencies including local councils and contractors working for the State Government.

Under the PDPA there is a single privacy and data protection framework. The PDPA uses its own Information Privacy Principles (IPPs) and organisations are obliged to act in accordance with the IPPs. As a result of the PDPA, a Privacy and Data Protection Commissioner has been established.

Information Privacy Principles

As discussed in the previous section, the amendments to the *Privacy Act 1988* in 2014 introduced new Australian Privacy Principles (APPs). It was anticipated that the new Act would replace the current Victorian Information Privacy Principles (IPPs) with the new principles based on the Australian Privacy Principles (APPs). However, this has not happened, so the Victorian *Privacy and Data Protection Act 2014* continues to use the IPPs.

TABLE 5.4 The 10 Information Privacy Principles (IPPs)

IPP 1	Collection of personal information	When an organisation collects information, it should only collect the information it needs. The organisation should inform people that their information is being collected.
IPP 2	Use and disclosure of personal information	When an organisation uses and discloses personal information it is only for the purpose that it was collected for, or for secondary purpose that you would reasonably expect.
IPP 3	Data quality	Ensure that the information collected is accurate, complete and up-to-date.
IPP 4	Data security	Information must be protected from misuse, loss, unauthorised access, modification or disclosure. Reasonable steps must be taken to destroy or de-identify personal information that is no longer needed.
IPP 5	Openness	The organisation needs to be transparent about what it does with information.

IPP 6	Access and correction	When an organisation collects information it should allow people to see the information it collects about them and provide them with the opportunity to correct it if it is inaccurate.
IPP 7	Unique identifiers	Use of unique identifiers, usually a number, is only allowed where an organisation can demonstrate that the assignment is necessary to carry out its functions efficiently.
IPP 8	Anonymity	Where possible, people supplying information should be given the option of not identifying themselves.
IPP 9	Transborder data flows	If your personal information travels outside Victoria, your privacy protections must travel with it.
IPP 10	Sensitive information	Organisations need to ensure that they do not collect sensitive information about people, such as their religion, political views or criminal record, without checking the applicable laws.

Health Records Act 2001

The Victorian *Health Records Act 2001* was created to provide direction with the collection and handling of health information in both the public and private sector. It is anticipated that patients will use both private and public health services at various stages of their life. The Health Records Act allows people to access their own medical information, as well as establishing the health record privacy principles for both public and private medical services. The Health Records Act established 11 Health Privacy Principles to provide rights to both living and deceased people. These principles apply to the collection, use and storage of personal health information in Victoria.

The Act protects the confidentiality of patients' health care information by allowing the information to be used only for the primary purpose for which it was gathered. This means that information about medical test results and your medical history may be used by your doctor, the hospital and any other health professionals only for the purpose of your immediate or ongoing care. This information would not be disclosed to a third party for a 'secondary' purpose (for example, your medical insurance company or another hospital), without your consent. Health information may, however, be provided to third parties without your consent under certain, and strictly limited, circumstances that include requests by family members in an emergency when you cannot give your consent and your life is threatened, where there is a serious threat to public health and welfare, research in the public interest, investigation of unlawful activity and as part of a legal claim.

An individual who believes that the Health Records Act has been breached can make a complaint to the Health Services Commissioner, who will try to achieve a resolution by discussion between the parties. If a satisfactory resolution cannot be reached, the Commissioner may then serve a compliance notice on the organisation that has breached the Act to inform the organisation which area of the Act has been breached and that it must correct its procedures. The maximum penalty for an organisation is currently 3000 penalty units and 600 penalty units for non-corporate cases.

Health Privacy Principles

Table 5.5 sets out a summary version of the Health Privacy Principles.

THINK ABOUT INFORMATICS 5.8

Stevie is a student support officer at a Victorian government school. He has access to student and parent personal details. He has been approached by an external organisation to 'sell' these details in exchange for new computer equipment for the school.

- 1 Identify key legislation that Stevie should consider before providing the information to the external company.
- 2 What are Stevie's ethical responsibilities to the students, parents and the school?

The Health Records Act 2001 applies to a deceased individual who has been dead for 30 years or less.

From the age of 16, teenagers can consent to medical and dental treatment with the same authority as an adult. Therefore, teenagers can see a doctor by themselves without their parents. At 18 years of age, they have the legal capacity to give consent to, and refuse, medical treatment.

The Act appointed the Victorian Health Services Commissioner to perform a range of functions administered under the Act, including conciliation, investigation and resolution of complaints.

This does not set out the full set or form of the Principles, and is intended for quick reference only. The principles in full can be found in the Act.

TABLE 5.5 A summary of the Health Privacy Principles

Health Privacy Principles summary	
1 Collection	Only collect health information if necessary for the performance of a function or activity and with consent (or if it falls within HPP1). Notify individuals about what you do with the information and that they can gain access to it.
2 Use and disclosure	Only use or disclose health information for the primary purpose for which it was collected or a directly related secondary purpose the person would reasonably expect. Otherwise, you generally need consent.
3 Data quality	Take reasonable steps to ensure health information you hold is accurate, complete, up-to-date and relevant to the functions you perform.
4 Data security and retention	Safeguard the health information you hold against misuse, loss, unauthorised access and modification. Only destroy or delete health information in accordance with HPP4.
5 Openness	Document clearly expressed policies on your management of health information and make this statement available to anyone who asks for it.
6 Access and correction	Individuals have a right to seek access to health information held about them in the private sector, and to correct it if it is inaccurate, incomplete, misleading or not up-to-date.*
7 Identifiers	Only assign a number to identify a person if the assignment is reasonably necessary to carry out your functions efficiently.
8 Anonymity	Give individuals the option of not identifying themselves when entering transactions with organisations where this is lawful and practicable.
9 Transborder data flows	Only transfer health information outside Victoria if the organisation receiving it is subject to laws substantially similar to the HPPs.
10 Transfer/closure of practice health service provider	If you are a health service provider, and your business or practice is being sold, transferred or closed down, without you continuing to provide services, you must give notice of the transfer or closure to past service users.
11 Making information available to another health service provider	If you are a health service provider, you must make health information relating to an individual available to another health service provider if requested by the individual.

*In the public sector, individuals already have this right under Freedom of Information.

Ethics and information systems

Despite the various laws designed to protect our information privacy, there are still many activities that people can perform that, while not illegal, are certainly questionable or morally frowned upon. **Ethics** refers to behaving in ways that are based on our morals and accepted standards. These standards may be common in a particular society or specific to a single organisation. They apply to questionable activities over and above any legal requirements. Ethics often provide us with a set of guidelines of appropriate behaviour. If we choose to ignore these guidelines, we may not be committing a crime, but we may be sacked by an employer or shunned by society.

For example, there has been a great deal of concern over the release of violent video games. Some individuals believe that video-game writers should not include animated violence in their games because it has a negative impact on children. A system of classification exists for games, similar to television and film classifications, but there is no legal restriction preventing these games from being created. These examples demonstrate how ethics hinge on society's values and standards. In this example there are two competing principles. On the one hand some would argue that protecting children from possibly harmful video games is the right thing to do. On the other hand, others would argue that it is more important to maintain freedom of expression. Often, then, questions of ethics become debates over which of two principles is more important. This is especially the case when the consequences of action are open to debate or interpretation.

Where the concept that there are two opposing sides when it comes to ethical **dilemmas** is most clearly shown is in the area of freedom of expression. While any attempt to curtail freedom of expression can be seen as censorship, most would acknowledge that there are limits with respect to privacy and fair comment. The question is just where do those limits lie?

When we apply ethics to the use of information systems within organisations, we can see several important areas affecting employers, employees and clients, or customers. These key areas are:

- the responsibilities of employers and employees to each other and clients
- codes of conduct
- company computer-use policies
- employee monitoring
- free speech on the internet.

Workplace responsibilities

Within any organisation, employees and employers have certain responsibilities towards one another as a duty of care, as well as to the customer or client. In particular, the employer must pay staff for the work they carry out and provide a suitable work environment in which that work can take place. In return, the employee is expected to work in the interests of the organisation for the duration of time they are being paid. In addition, the organisation is expected to provide good-quality products or high levels of service to customers. If an employee is not working well for the organisation – perhaps they are rude to customers – it is understood that the manager has the right to dismiss the employee. The employee is not breaking a law, but they are going against the standards of service insisted upon by the particular company for which they work.

Codes of conduct and computer-use policies

A **code of conduct** is a set of conventional principles and expectations that is considered binding on any member of a particular group. A code of conduct for a lawyer would apply to the work he or she does on behalf of a client; a doctor prescribes treatment in the interests

Ethics are the principles of right and wrong that are accepted by an individual or a social group. Ethical behaviour often guides policy makers within organisations. For example, a network policy within an organisation governs both the legal and ethical use of the information system.

A dilemma occurs when there is a choice between two options of equal desirability. For example, if doctors are requested to send patient medical details to a central authority for a nationwide health study, parents of young children are placed in a dilemma in that they must weigh up their child's right to privacy against the benefits to society.

THINK ABOUT INFORMATICS 5.9

Identify five key behaviour and work standards expected of an IT professional.

Although an employer can dismiss a worker who is operating at below the required standard, the worker can access unfair dismissal laws. If this happens, the employer needs to be able to demonstrate that due diligence has been followed.

The due diligence process needs to include evidence of meetings between the employer and the worker to discuss improvements needed to the worker's performance and to document goals and strategies to redress the problem. The employer needs to show that they have tried to help the worker to fix any failings.

The Australian Computer Society has a code of professional conduct for its members. This code includes all aspects of behaviour and work standards expected of an industry professional. A serious breach of a code of conduct may result in the termination of an employee.

THINK ABOUT INFORMATICS 5.10

Locate your school's computer-use policy.

- 1 What key ethical areas does it cover?
- 2 How effective is the policy in your opinion?

Employee monitoring advantages

- It ensures that employees are doing company work.
- It ensures that employees maintain target performance levels.
- It saves time and money.

Employee monitoring disadvantages

- It is intrusive and may impact on employee privacy.
- Mistrust may develop between employee and employer.

of a patient. However, it may seem that a computer programmer's position is unclear: is the programmer working for their manager or the client? A code of conduct might say they are responsible to both, even though it may seem harder to relate workplace responsibilities in the information systems sense because conventional roles seem unclear.

What happens when there is a conflict? For example, the programmer might know that there is a bug in a piece of software that they are working on. This will have an impact on the accuracy of the software's data manipulation. Due to a strict deadline, the manager tells the programmer to ignore the bug, because it is more important to have the package delivered on time. In this situation, it is difficult for the programmer to know if they owe a greater obligation to their company or to the client, or even to the general public who will be the end-users.

All organisations should have their own company-wide computer-use policy. This document would explain clearly to staff what management believes should and should not be done on the computers or peripheral equipment. Employees will then be in no doubt as to what is permitted and when.

Some companies may allow employees to use company computers for their own purposes, such as network games, personal email or general internet use in their own time. Other companies may not allow employees to use computers for any purpose other than work, even during a lunch break.

Students, too, are often required to follow school-based codes of conduct. For example, students may be allowed access to the internet while at school, but they may have to sign an undertaking that they will not use it for inappropriate purposes, such as downloading pornography, using chat rooms or playing online games.

Employee monitoring

If codes of conduct and computer-use policies exist in an organisation, it is in the company's interest to enforce them. Often, managers will use monitoring systems to check on what their employees are up to. Although these systems are not illegal, depending on the extent of their intrusiveness, they may be considered unethical because they invade the employee's privacy. It might be considered acceptable for supervisors to monitor some telephone calls in a call centre to ensure quality of service, but it is unethical to monitor and then possibly restrict employees' toilet breaks to certain times of the day.

The most common monitoring occurs on email, PC use and web browsing. Email messages can be automatically redirected to a manager if they contain certain words or attachments, especially .jpg or .exe, or if they come from an unrecognised email address.

Computer monitoring allows a supervisor to see what an employee is keying, as well as the files on their computer and even the desktop layout. The employer may believe it is important to be able to check that employees are doing what they are paid to do, especially with regard to data entry operators who are required to perform to a high standard of keystrokes per hour. If staff know the boss may be watching what they are doing on the computer, they will be less likely to slacken off or do things that are not work-related.

In theory, this may save time and money for the employer. However, employees may argue that the employer is being too intrusive. In extreme cases, employees may see words start to change in a report they are writing if managers make changes to their files. They may feel an increase in stress levels because they may not know when they are being 'spied on' electronically. Computer monitoring may create mistrust between supervisors and staff. While monitoring is not illegal, in some cases it may be seen as unethical because employees would consider that they are entitled to some privacy while at work.

Although the internet might be invaluable to an organisation for e-commerce, research or as part of a corporate intranet, managers are also wary of its misuse for personal purposes. It is difficult to restrict the sites that users may access without severely limiting the access of other internet users in the organisation. Web browsing is, therefore, usually monitored by way of automatic logs. These are stored either on the user's own machine in the form of **cookies** or on a history of sites. The server will automatically maintain a list of every site and file accessed, cross-referenced to the relevant username and the date and time. While the log file may be huge, it is very easy to search it for key words or phrases.

Again, internet usage within an organisation might be regulated by clauses of the computer-use policy or, in the case of many schools, an internet acceptable-use policy.

The organisation may not be too concerned about an employee looking up destinations for an upcoming holiday during their lunch break, but the manager may become upset if the employee looks up pornography or gambling-related sites, or even job advertisements. From the employer's point of view, it is considered unethical to be 'misusing' the internet connection, even if it is outside working hours. Similarly, employees might expect some leeway if the alleged misuse takes place out of necessity. For example, if you were planning to book airline tickets at a discount rate, but they were only available through the internet at a particular time, then your supervisor might allow you to do so.

Ethics and the internet

A final area where ethics are under review is the internet itself. There is no central server or network administrator policing the internet. Many people see the internet as a true medium for free speech. Anyone with access to a service provider can set up a home page and publish any material they want. This might be something as harmless as photographs of a family birthday party or a research thesis on the breeding habits of moths. Or it could be hardcore pornography or extremist political views.

'Netiquette' is short for 'network etiquette'. It refers to the conventions of politeness observed in internet communications, such as newsgroups or mailing lists. It evolved as a means of regulating what people publish on the internet. Many people believe that others in the global community should have the right to say whatever they want on the internet, as long as they do so responsibly. As more families explore the internet, it is often in the interests of service providers such as AOL and Telstra, or larger search engines such as Google, to tame the internet wilderness by imposing controls on what types of material are stored on their servers or can be found from their search pages. Often, it is public pressure or the desire to chase the large market of family-based customers that effectively forces these organisations to censor their content and links.

Cookies were discussed in Chapter 1.

Netiquette guidelines include:

- posting only to appropriate groups
- refraining from commercial advertising
- keeping personal messages (between one or two individuals) private and not posting them to newsgroups or forums
- quoting only the minimum necessary context when responding to an earlier post
- keeping subject lines to fewer than 70 characters in length
- reading the existing messages in a group's FAQs before asking a new question

ISSUE

New legislation to be put forward to block torrenting websites, but will it work?

A Bill is set to be introduced to the Senate next week that could be the nail in the coffin for online piracy but its detractors say it curbs internet freedom.

The Copyright Amendment (Online Infringement) Bill is a part of the government's ongoing crackdown on copyright breaches in Australia and comes in the wake of a proposed three strikes scheme that could see illegal downloaders face harsh financial penalties.

The Bill will force ISPs to block overseas sites that allow people to share copyrighted material, at the request of a judge

In 2010, Stephen Conroy from the then-Labor Government proposed to filter the internet. His proposal was met with disdain by many groups of people, including internet service providers, civil libertarians, Google, Yahoo and the US Government. Although the ALP had honourable intentions to filter the internet of undesirable material, many people were concerned about providing the government with too much power. In addition, many people believe that the flow of information should be free and that freedom should not be violated by anyone, including government. In 2014, Malcolm Turnbull, of the incumbent Coalition Government, attempted to enforce internet filtering by putting pressure on ISPs to block piracy sites, such as The Pirate Bay.

THINK ABOUT INFORMATICS 5.11

- 1 Do you think the government should censor the internet?
- 2 Do you think it is appropriate for copyright holders to have their work 'shared' without any form of payment?
- 3 Do you think that the new legislation will be successful and prevent illegal content from being accessed? Discuss your views.

Cybersmart is a project run by the Australian Communications and Media Authority (ACMA). Cybersmart provides advice and resources to Australians about protecting themselves and their children while using the internet. Issues covered by this site include spam, safe internet use, cyber bullying, chatroom safety and racial hatred.

A spokesman for Senator Brandis' office said the bill will be presented to the Senate Legal and Constitutional Affairs Legislation Committee for review next week.

If the bill is to eventually pass, the expected outcome would see content rights holders moving to lodge court orders to have all known piracy sites blocked in Australia.

Rights holders can take direct action against Australian based websites, so the legislation could effectively see a blanket removal of torrent sharing sites across Australian internet providers.

The text of the bill has not yet been released and even ISPs have been kept in the dark.

John Stanton, CEO of the Communications Alliance, who worked on behalf of major ISPs to develop the three strikes program said he is 'disappointed' that his group hadn't been consulted.

The site-blocking scheme has been likened to online censorship by critics including consumer advocate group Choice and Pirate Party Australia, who argue that it will set a dangerous precedent.

Pirate Party Deputy President Simon Frew said the legislation 'will mean we are subjected to a censorship regimen. The Government has opted for a long and pointless game of whack-a-mole –as soon as a site is blocked it will pop up in several new places and copyright infringement will continue.'

So will it actually work?

The same type of system has been in the UK since 2011 with little success. Some of the most popular torrenting sites in the world, including The Pirate Bay and Kick Ass Torrents have all been blocked, but locals keep finding a way to download content.

Typically, within hours of a site being put on the block list by an internet provider, hundreds of mirror sites that offer the same content pop up for users to access. Then of course, by the time those are blocked or taken down, there are just more to replace them. It's a never ending circle.

Other methods UK locals have been using to get around the block include using proxy websites, which stop the ISPs from thinking you're actually visiting those websites.

As a result, these proxy sites have also been blocked.

However, The Pirate Bay, the world's biggest torrenting site now uses a new service to host the site which stops most ISPs from being able to block it. The new service effectively hides the information about The Pirate Bay from internet providers, making it harder for them to block access to their main site.

James Brandes from ORGZine, a UK digital rights magazine, says: 'Not only is the block policy fundamentally failing, but it raises important censorship.'

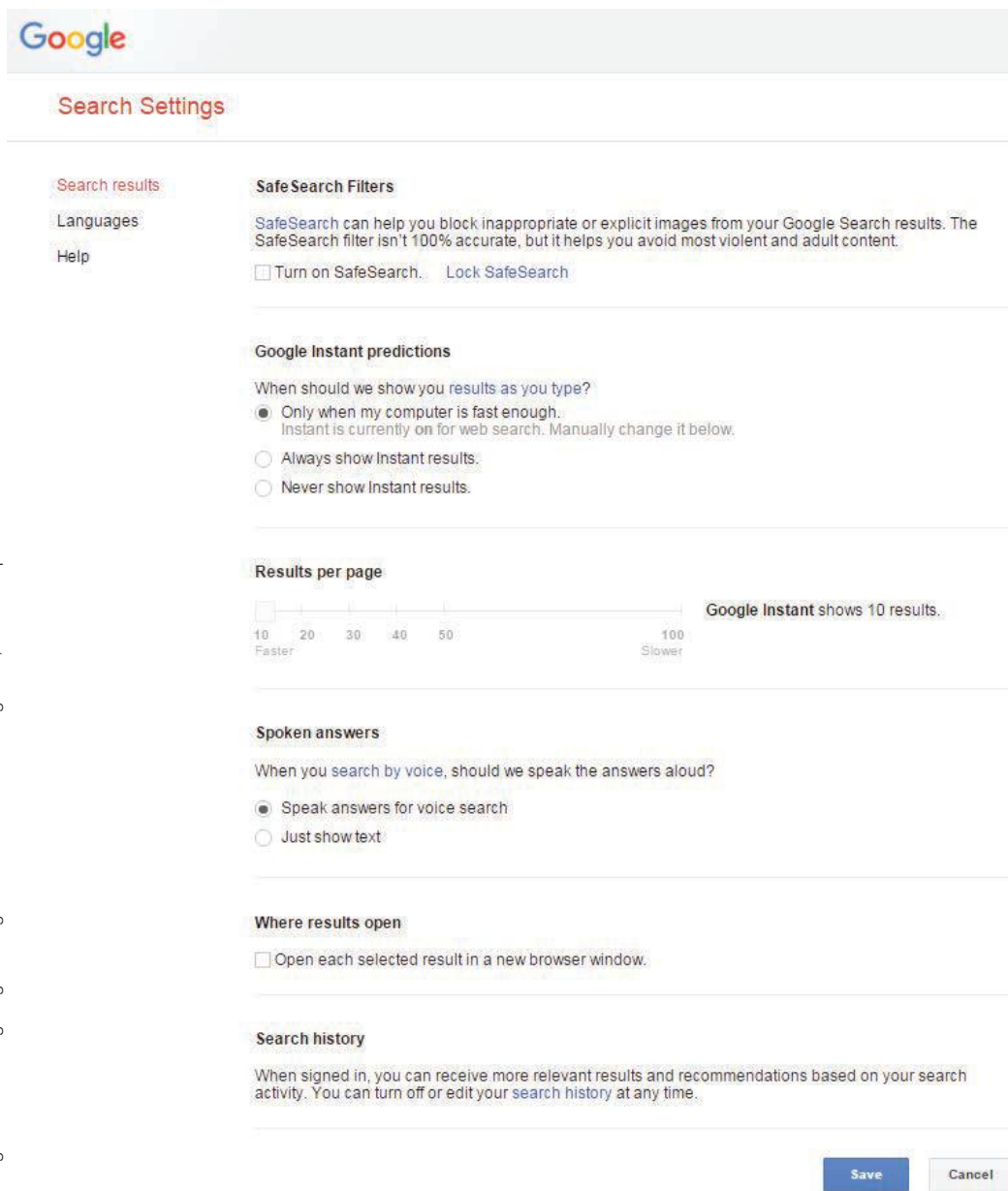
With the proposed metadata laws due to be passed by the end of this month, new piracy laws could be the nail in the coffin for Australia's freedom on the internet.

News.com.au, Nick Whigham & Harry Tucker, March 19, 2015

Ideally, ISPs should self-regulate the files and information that they allow on their server. In late 2000, however, a French court ruled that Yahoo! had to restrict any French internet users from accessing sites that sold Nazi memorabilia. This ban applied not only to the French branch of Yahoo!, but also the United States site. This tested legal jurisdiction as well as technical capabilities because selling Nazi memorabilia is not illegal in the United States.

Online book giant Amazon also found itself in difficulty after the German government insisted that German-language versions of Adolf Hitler's book *Mein Kampf* were not to be sold to German customers over the internet. Germans got around this ban by purchasing the English version instead. Anti-Nazi campaigners were also appalled to learn that Amazon's automatic listing of similar books had also operated in connection with *Mein Kampf*, giving customers a suggested list of further Nazi-related reading.

One popular method of registering disapproval about receiving spammed email is to **flame** the sender. It might be satisfying to clog the sender's email address with many angry replies, but this also inconveniences anyone else using that system. In effect, you might be causing just as much disruption as the spammer. Therefore, you may need to make an ethical decision about replying or putting up with the spam.



The screenshot shows the Google Search Settings page. At the top is the Google logo. Below it is the 'Search Settings' header. The page is divided into several sections:

- Search results:** Includes links for 'Languages' and 'Help'.
- Safe Search Filters:** A description of SafeSearch and a checkbox to 'Turn on SafeSearch' (which is currently unchecked). A 'Lock SafeSearch' link is also present.
- Google Instant predictions:** A question 'When should we show you results as you type?' with three radio button options: 'Only when my computer is fast enough.' (selected), 'Always show Instant results.', and 'Never show Instant results.' A note states 'Instant is currently on for web search. Manually change it below.'
- Results per page:** A slider ranging from 10 (Faster) to 100 (Slower). The slider is set to 10, and a label indicates 'Google Instant shows 10 results.'
- Spoken answers:** A question 'When you search by voice, should we speak the answers aloud?' with two radio button options: 'Speak answers for voice search' (selected) and 'Just show text'.
- Where results open:** A checkbox to 'Open each selected result in a new browser window.' (unchecked).
- Search history:** A description of how search history is used for recommendations and a link to 'turn off or edit your search history at any time.'

At the bottom right, there are 'Save' and 'Cancel' buttons.

FIGURE 5.9 In search engines such as Google, there are settings that allow you to filter out unwanted content when you perform searches, such as adult images and websites

Resolving legal and ethical tensions

Working in an organisation with ethical, legal or social tensions can be tense and uncomfortable. Often tensions arise through the lack of clarity in policy documents that exist within the organisation to govern the use of the information system.

For example, a situation might arise in which an employee needs to perform a task on the network that their manager needs to approve. The manager is sick and the employee had decided to take it upon themselves to access the manager's account in order to get the task done. Although this action is not illegal, it is unethical and it could result in disciplinary actions for the employee.

In addition to legislation such as the *Copyright Act 1968* and the *Privacy Act 1988*, organisations might have policy documents that mandate who has access to areas of the information system and the penalties for misuse of information within an organisation.

**THINK ABOUT
INFORMATICS
5.12**

Read through your schools acceptable-use policy and annotate it to show guidelines that cover ethical and legal uses of the school network.

As part of this Unit, you are not required to solve the dilemma, but rather the process used to solve it.

The suggested six steps for handling ethical dilemmas act as a framework that can be used within an organisation to solve legal, ethical or social tension.

Six steps to solving ethical dilemmas

- 1 **Identify the problem.** What decision has to be made and what facts are required?
- 2 **Identify the stakeholders.** Who are they? What interests do they have? Who is the key player?
- 3 **Identify possible consequences.** What options are available? What are the likely consequences?
- 4 **Identify ethical standards.** Are there any applicable laws? Are there any codes of conduct or standards that could be applied? Is there a precedent?
- 5 **Evaluate options.** Identify strengths and weaknesses. Identify the option that causes least harm. Can the decision be reversed?
- 6 **Make a decision.** Select the preferred option. Justify the option.

Once the organisation has solved its problem, it may wish to update any policies concerned to ensure that the process for resolution is clearer.

ESSENTIAL TERMS

- biased** a judgement that is based on prejudice or faulty reasoning
- code of conduct** a set of conventional principles and expectations considered binding on any member of a particular group
- data** raw, unorganised facts, figures and symbols fed to a computer during the input process; data can also mean ideas or concepts before they have been refined
- dilemma** occurs when there is a choice between two equally undesirable options
- ethics** principles of right and wrong that are accepted by an individual or a social group; ethical behaviour often guides policy makers within organisations
- flame** to criticise or heap indignities on an individual in a public email message, often resulting from an inappropriate or uninformed remark made by the individual who is flamed; spammers are often flamed for ignoring netiquette

IMPORTANT FACTS

- Information is obtained when **data** is manipulated by the computer's processor into a meaningful and useful form (also becoming its output).
- The large amounts of data that is being gathered and analysed has become very valuable to organisations. Large data sets or **big data** is assisting with competition, productivity growth and innovation.
- Data can be **compromised** through activities such as deliberate attacks of theft (disgruntled workers or criminals that will make money selling the data), loss of devices (such as accidentally leaving laptops at airports or in a taxi), neglect (not erasing data when recycling computer hardware) and not following appropriate data handling procedures and policies.
- As part of the Privacy Act, the **Australian Privacy Principles (APPs)** were devised to set out the standards, rights and obligations for collecting, handling, holding, accessing, using, disclosing and correcting personal information. There are 13 APPs.
- The **Privacy and Data Protection Act 2014 (PDPA)** was introduced by the Victorian Government. It replaced the *Information Privacy Act 2000* and the *Commissioner for Law Enforcement Security Act 2005*.
- The other key Victorian law relating to privacy is the **Health Records Act 2001**. This Act governs the collection and handling of confidential medical records.



TEST YOUR KNOWLEDGE

GOALS AND OBJECTIVES OF ORGANISATIONS AND INFORMATION SYSTEMS

- 1 Explain the difference between an organisational goal and a mission statement.
- 2 What is a vision statement?
- 3 How are values important to the content of mission and vision statements?
- 4 Where does the purpose of mission and vision overlap?
- 5 Why are mission and vision important for organisational goals and objectives?
- 6 Apart from making a profit, think of an organisational goal for each of the following types of organisations:
 - a a veterinary clinic
 - b a large shopping centre
 - c an airline
 - d a brewery
 - e a public library.

INFORMATION MANAGEMENT STRATEGIES

- 7 Explain why an organisation must comply with legal requirements.
- 8 Briefly summarise the role and scope of the three key laws affecting privacy of information.
- 9 Why have these laws been introduced?
- 10 If you believe that the privacy of your information has been breached by the Australian Taxation Office, to whom can you complain?
- 11 What are the penalties for breaches of the *Privacy and Data Protection Act 2014*?
- 12 For each of the following breaches of privacy, suggest which privacy law would cover it.
 - a You find that your employer has published your tax file number on the internet.
 - b Medical records are found at the tip.
 - c A bank refuses to give you a loan because the manager claims your credit record is poor, when it should actually be very good.
 - d A consultant working for the Victorian government passes on your VCE results to a friend without your permission.
 - e A website you visit asks for personal information from you, but does not display its privacy policy.

ETHICS AND INFORMATION SYSTEMS

- 13 Explain how ethical requirements differ from legal requirements.
- 14 What are some ethical responsibilities that employers and employees have to one another?
- 15 What is the purpose of a code of conduct?
- 16 What type of ethical restrictions can be applied to accessing the internet at work?

APPLY YOUR KNOWLEDGE

- 1 Locate a company and identify:
 - a the company's mission statement
 - b its organisational goals
 - c at least one information system within the company and explain how it meets one of the organisational goals.
- 2 A large Australian financial institution sends a policy document to all clients on how the company handles personal information. The document includes the following statements.

Your personal information

Personal information is information or opinion that allows others to identify you. It includes your name, age, gender, contact details, as well as your health and financial information. We will act to protect your personal information in accordance with the National Privacy Principles or an industry privacy code. We collect personal information to provide you with the products and services you request as well as information on other products and services offered by or through us.

How we use your personal information

Personal information may be used and disclosed within the company to administer our products and services, as well as for prudential and risk management purposes and, unless you tell us otherwise, to provide you with related marketing information. We also use the information we hold to help detect and prevent illegal activity. We cooperate with police and other enforcement bodies as required or allowed by law.

- a Explain why the document is sent to clients.
- b Provide reasons for including the statements within the document. Also included in the document is the following statement.

We disclose relevant personal information to external organisations that help us provide services. These organisations are bound by confidentiality arrangements. They may include overseas organisations.

- c How should the organisation ensure that a client's privacy is protected if they need to send personal details overseas?
- 3 The Children's Singing and Dancing Academy is an organisation that runs singing, dancing and acting classes specifically aimed at school-age children. The company offers classes after school on most weeknights and on weekends in various locations around Melbourne. Children from many suburbs participate in this extracurricular activity. The company is a not-for-profit organisation set up specifically to broaden children's interest in the performing arts. The Academy has a webpage and advertises its classes and locations through this medium. Pictures taken of students during their end-of-semester performances are used for advertising the company. The company relies on technology usage by creating brochures, updating their webpage and storing clients' data. Although the data stored is primarily about the children, data on the parents/guardians of the children is also stored.
 - a What particular data on children is stored at the Children's Singing and Dancing Academy?
 - b What particular data on the parents/guardian is stored at the Children's Singing and Dancing Academy?
 - c Why does the company need to store data on their clients?
 - d If you were to write an organisational goal for the Children's Singing and Dancing Academy, what would it be?
 - e What does the Children's Singing and Dancing Academy need to do to ensure they are compliant with the *Privacy Act 1988*?
 - f What measures are needed for the Children's Singing and Dancing Academy to protect the integrity of data and information?

SECURITY MEASURES

Key knowledge

After completing this chapter, you will be able to demonstrate knowledge of:

Interactions and impact

- reasons for preparing disaster recovery plans, and their scope, including evacuation, backing up, restoration and test plans
- possible consequences for organisations that fail to follow or violate security measures
- criteria for evaluating the effectiveness of information management strategies

Digital systems

- role of people, processes and digital systems in the management of data and information
- types and causes of accidental, deliberate and events-based threats to the integrity and security of data and information
- physical and software security controls for preventing unauthorised access to data and information, and for minimising the loss of data accessed by authorised and unauthorised users
- the advantages and disadvantages of using networks and cloud computing for storing and disposing of data and information.

For the student

This chapter relates to what you need to know for Unit 4, Outcome 2. It concentrates on threats to data, protecting data, disaster recovery plans, digital systems, networks and cloud computing.

For the teacher

This chapter, together with Chapter 5, is based on Unit 4, Area of Study 2, and provides the key knowledge required to complete Unit 4, Outcome 2. This chapter provides an overview of people and processes in information systems, and networks.

It explores in detail threats to data, data protection measures, disaster recovery planning, digital systems, consequences of security failures and cloud computing.

Preparing for the School-assessed Coursework (SAC)

Together with Chapter 5, Chapter 6 prepares you for the Unit 4 SAC. This chapter continues the study of information management and its importance to organisations, but focuses chiefly on security measures, and network and cloud computing.

We will talk about the different types of threats to data and information, which include **accidental threats**, **technical threats**, **deliberate threats** and **event-based threats**.

Next, **disaster recovery plans** will be addressed, particularly in terms of why organisations need to plan for disasters, how they plan for disasters and the scope of their plans, including testing and data recovery strategies.

Sometimes, organisations fail to follow security measures or even violate them outright, so we will also discuss the consequences of such actions, including penalties, loss of reputation and loss of income.

Next, the chapter will discuss how data and information management is conducted across the components of information systems, which comprise people, processes, digital systems and data, and we will delve into the topics of people, processes and digital systems, concluding with a look at networking. Following this, we will look at the advantages and disadvantages of using networks and cloud computing for storing and disposing of data and information.

The chapter concludes with the criteria that can be used to evaluate the performance of information management strategies. These criteria fall under two major categories: effectiveness and efficiency.

Threats to data and information

Data and information are the life-blood of any organisation and its most valuable assets. Data and information can be damaged, destroyed, stolen, copied or deleted in multiple ways, most of which are avoidable. While there are many threats to an organisation's information systems, the right **physical and software security controls** and procedures can both prevent and minimise the effects of data loss.

Threats to data and information come under the categories of accidental, technical, deliberate and event-based. The following sections show each category of threat using tables.

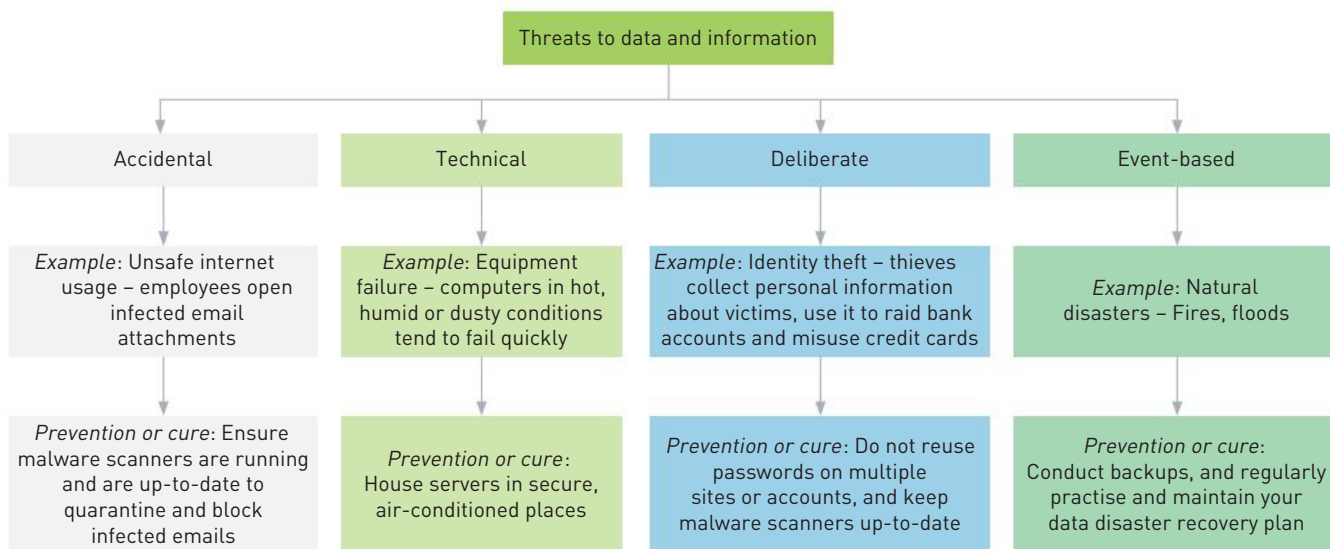


FIGURE 6.2 Threats to data and information



FIGURE 6.1 Chapter map

Accidental threats

Data security is also discussed in Chapter 3.

The integrity and security of data can be accidentally threatened in multiple ways, as described in Table 6.1.

TABLE 6.1 Accidental threats to the security and integrity of data

Cause	Explanation	Prevention or cure
User error	The biggest cause of data loss is the person using it, and can result from inexperience, lack of skill, carelessness, failure to save regularly, or poor documentation. Reports regularly suggest that user error accounts for 60% to 75% of all data loss. Error can also occur as a result of lack of training or supervision, or an employee being given access to inappropriate levels of data. Staff must be trained properly and monitored sufficiently so they do not: • disclose private information, such as passwords • delete vital records or files • damage databases • damage hardware; for example, by accidentally spilling coffee on a file server • invite unauthorised people into secure areas • choose weak passwords.	Train staff thoroughly in the use of software, computers and company procedures, such as file naming conventions, and how to choose strong passwords and recognise social engineering. Use access hierarchy to give employees only the level of data access that they need to do their jobs.
Poorly managed backups	Being unable to recover from data loss can be caused by poor backup procedures, such as backing up infrequently, inadequate storage media, or a lack of testing of the backup regimen.	Ideally, organisations should backup all their data every day and store their backups offsite. They should also test their backup procedures to ensure that they are working properly and backups can be restored. They should also use quality storage media. <i>You</i> should also backup frequently and follow the same procedures, albeit on a smaller scale.
Poor management 	Poor office management can lead to many different types of data disasters, such as the following. • Staff who are dismissed may delete files or steal data before they leave. • A disgruntled IT manager may leave without telling anyone the network administrator password, so the company is locked out of its own system. • Employees have unauthorised duplicate keys cut that can later be used to enter the building without authorisation. • Staff are untrained in cases of emergency. • The network's resources are wasted on illegal unauthorised purposes, such as downloading torrents. • Security updates are not installed. • Staff are permitted to use unauthorised or unapproved equipment, such as unsecured wireless access point, unsecured or infected personal laptops, or infected software.	• Revoke the network access of employees before they are fired. • Keep a secret master password (unknown even to the IT manager) that gives complete access to the digital systems. • Use copy-restricted keys or electronic security passes that cannot be copied. • Carry out drills to practice all aspects of the disaster recovery plan. • Use network monitoring software to detect intrusions and to log all network activity, such as opening or emailing sensitive documents. • Enforce policies forbidding the use of unauthorised equipment.
Lost data storage devices 	Massive amounts of sensitive, secret or valuable data can be lost if: • USB keys are left plugged into public computers • portable hard disks or laptops are left in taxis, on public transport or in public places • obsolete computers are sold without first being reformatted.	• Encrypt all portable storage devices and laptops. • Reformat or destroy the hard disks of obsolete computers.

TABLE 6.1 (continued)

Cause	Explanation	Prevention or cure
Unsafe internet usage	- Porn, gambling, hacking or file sharing sites can open up an entire network to malware via Trojans or drive-by downloads. - Employees may open infected email attachments, reply to spam, respond to phishing attempts, click links in suspicious emails, or mistakenly send valuable data to the wrong email address.	- Use the network's proxy server to block access to all internet sites apart from those needed by staff to do their work. - Ensure malware scanners are running and are up-to-date.

Technical threats

Technical threats originate from the hardware and software used in a digital system. Generally, these are a result of failures, bugs, infections or similar problems. A list of technical threats appear in Table 6.2.

TABLE 6.2 Common technical threats

Cause	Explanation	Prevention or cure
Equipment failure 	- Failure of servers, hard disks or core communications equipment, such as switches and routers, can cause loss of access to network data, data loss and loss of productivity - Computers in humid, hot or dusty conditions tend to fail quickly.	- Do not buy cheap or underpowered equipment. - Maintain and update hardware properly and handle equipment carefully. - House servers in secure, air-conditioned environments 
Electric problems 	Blackouts, brownouts, spikes and power surges can damage sensitive electronic equipment.	Servers need battery-backed uninterruptible power supplies (UPS).
Software bugs	Bugs (programming errors) in an operating system, device driver or application can crash and lose data, or make computers vulnerable to attack.	Regularly update software and operating systems.
Improperly configured equipment	Dangerous security holes or poor performance can be caused by equipment that has been configured incorrectly; for example, if: - wireless encryption has not been enabled - the manufacturer's default security passwords have not been changed - the GPU co-processing option has not been turned on - outdated or incorrect device drivers are being used.	Read the manual. Visit manufacturer forums to learn of problems and cures from other product users.



Shutterstock.com/Kotomiti Okuma

FIGURE 6.3 Most public libraries, schools and internet cafes have large collections of lost flash drives. Never trust a vital document only to a Flash drive. Save it also to disk, upload it to the cloud or email it to yourself so you have a backup.

TABLE 6.2 (continued)

Cause	Explanation	Prevention or cure
Malware infection	Once malware enters a system, it can spread and expose an entire network to exploitation and data loss. Malware includes worms, Trojans, viruses, spyware, adware, root kits and ransomware .	Use vulnerability management software to perform automated scans of systems to discover security weaknesses such as open ports, unpatched software and SQL injection vulnerability. 



Corbis/Kim Kulish

FIGURE 6.4

Uninterruptible power supplies can keep a PC or server running after a blackout. A smart UPS can tell the server when it can no longer supply power and that it is time for the server to gracefully shut itself down.

Deliberate threats

Networks are attractive targets for organised crime syndicates, as well as for political, religious or commercial adversaries. Networks contain massive amounts of valuable data. Stealing data can be tempting because it can often be easily accessed remotely with little effort or risk of being detected or punished. A list of deliberate threats is described in Table 6.3.

TABLE 6.3 Common deliberate threats to networks

Cause	Explanation	Prevention or cure
Malicious staff	It is estimated that 18% of deliberate attacks on systems come from staff within an organisation. Of those, half come from digital systems staff. Staff with privileged access may delete data out of spite or steal valuable data to use as a bargaining tool when seeking a job with a rival company.	Hire external consultants to conduct security audits and examine networks for: • secret 'back door' entrances so employees cannot enter the system after leaving the company • booby traps that damage the system if employees are aggrieved • unauthorised network privileges, such as staff giving themselves administrator access.
Crackers	Crackers (malicious hackers) may penetrate systems to plant botnet software or keyloggers so they can control systems remotely to steal data, discover passwords or banking information, conduct DDoS (distributed denial of service) attacks or turn computers into zombies that deliver spam.	Constantly run malware scanners. Run network monitoring software to identify attempts to break in. 

TABLE 6.3 (continued)

Cause	Explanation	Prevention or cure
Identity theft	Once enough personal information is known about victims, their bank accounts can be raided or their credit cards misused. Victims may make matters worse: • by reusing the same password on different sites • by not protecting their computers • with careless browsing practices.	• Do not reuse passwords on different sites. • Keep malware scanners up to date. • Avoid suspicious sites, such as gambling, pornography or pirate software sites. • Browse in 'incognito' mode. • Carefully check software installations for devious extra software additions, such as toolbars. • Use multi-factor authentication whenever it is offered.
Equipment thieves 	Thieves may walk off with expensive hardware, along with the irreplaceable data stored on it.	• Improve physical security and strengthen security procedures; for example, keep doors locked and do not allow non-staff near company computers. • Record serial numbers of equipment so police can identify stolen equipment. • Encrypt hard disks.
Vandals	People may physically damage equipment or electronically deface a website for fun, revenge, fame, money or a challenge.	• Improve physical and electronic security. • Keep your website files backed up.
Espionage	• Political espionage aims to discover state secrets of a rival nation. • Corporate espionage aims to learn about a competitor's new product, business plans, or how much they plan to bid on a government tender.	• Encrypt all documents so they are unreadable even if they are stolen. • Run continuous network monitoring software to detect break-ins.
Fraudsters	Scammers, phishers, and social engineers manipulate gullible staff to gain unauthorised information or payment by trickery.	Train staff to recognise deceptive practices.

Event-based threats

Natural or manmade events can threaten or devastate an organisation's data and infrastructure.

TABLE 6.4 Possible natural or manmade threats

Cause	Explanation	Prevention or cure
Natural disasters	Disasters such as fires, floods, storms, lightning and earthquakes can cause equipment failures and even destruction. While some may be reasonably anticipated based on the time of year and location, such as bushfires in a Victorian summer or an earthquake in Tokyo, others may be completely unpredictable and unexpected, such as the deadly Newcastle earthquake of 1989.	• Conduct backups, ensuring that all backups are kept offsite. • Maintain and practise your data disaster recovery plan. • Consider storing data in buildings designed to withstand earthquakes or that are elevated above floodplains.
Manmade disasters	Acts of war, riots, terrorism or civil unrest can cause devastating disruptions to information systems.	Monitor the news. With enough early warning, data managers may have enough time to protect data.

THINK ABOUT INFORMATICS 6.1

- 1 Investigate and explain the meaning and importance of multifactor authentication in preventing identity theft.
- 2 Research the Stuxnet worm. What does it do? Who created it? Why did it become famous?
- 3 As autonomous technology develops, such as driverless cars, self-flying planes and computer-controlled infrastructure (including power stations, dams, oil, wells and air traffic control), what potential is there for catastrophic deliberate damage to be done?

THINK ABOUT INFORMATICS 6.2

'Prevention is better than cure.' Does this statement ring true when it comes to data loss? Discuss.



NewsPix/News Ltd

FIGURE 6.5 Some tragic disasters, such as the 1989 Newcastle earthquake, cannot be predicted or even anticipated

Disaster recovery plans

A disaster recovery plan (DRP) is a comprehensive scheme that explains how to prepare for, survive and recover from a disaster that affects information systems infrastructure.

It is prepared in advance, tested regularly and executed immediately after a disaster to minimise disruption and restore normal business operation as soon as possible.

A well-designed DRP has several benefits:

- it provides a sense of security and confidence to workers
- it prevents further problems that might be caused by panic
- it reduces the amount of uncertainty and decision-making in stressful moments
- it neatly explains procedures to new employees.

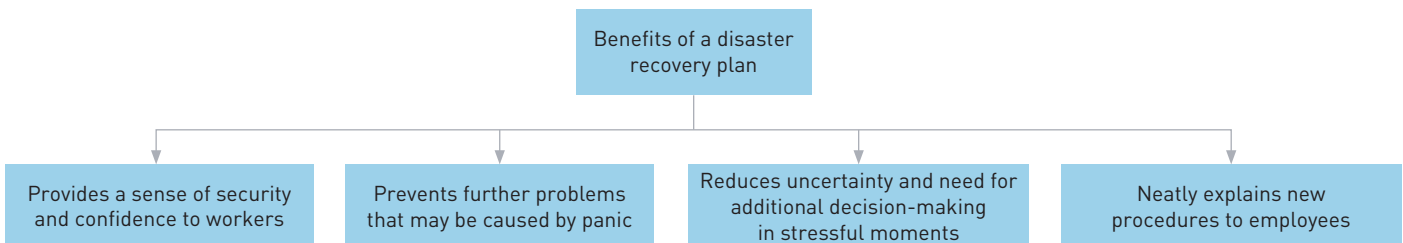


FIGURE 6.6 Benefits of a disaster recovery plan

The four main components of a DRP cover:

- 1 evacuation
- 2 backing up data
- 3 data restoration
- 4 testing of disaster plans.

Regardless of how well you protect your data and information against threats, there is still a chance that it can be damaged, lost, stolen, copied or destroyed. Only a fool would consider their data completely and permanently invulnerable. New threats are invented or discovered every day, and even the cleverest system manager cannot hope to prevent disasters caused by all known and unknown threats.

Disaster planning is about accepting the fact that one day, in spite of all your careful planning, your data and information will be stolen or destroyed. Whether it happens deliberately or accidentally no longer matters: the DRP is your last chance to survive and recover. Without a means of recovering lost data, an organisation struck by data tragedy faces a high probability of quick corporate death.

Make sure you create your DRP when you create your organisation. If you join an organisation and it does not have a DRP, make it a priority. A DRP needs to exist before disaster strikes. Inventing the comprehensive and efficient action plan while the office burns down is a bit like learning to swim after you have fallen overboard: you will have left things much too late.

Creating an effective, complete DRP requires considerable planning, and this planning must pass through distinct stages.

Scope

Risk assessment

Risk assessment analyses the probability of different risks to your data and information, and the potential consequences of each of those risks. For example, rural organisations would be very concerned with the threat of bushfires. City organisations would be more concerned with burglary. It is more sensible to pay the most attention to the most likely and most damaging risks to your data.

- An organisation with many unskilled workers may need to focus on continual backups and staff training.
- A school may focus on vandalism and the misuse of accounts by students.
- Military establishments are very concerned with protecting secrets against espionage and physical attack.
- A company headquartered in the mountains may need to anticipate avalanches more than a company located next to the beach.

Each organisation's needs will be different.

Invest in outside consultants to inspect your organisation, test your security and examine your equipment and procedures to identify potential risks or weaknesses. Where possible, take preventative measures to reduce the chance of identified risks actually occurring. Buy a fireproof safe and install fire alarms; in flood-prone areas, move servers onto high racks. Train staff not to give out passwords to smooth-talking strangers on the phone and improve malware defences. Invest in a continuous data backup scheme and install network monitoring software to detect intrusions. Every dollar spent on hazard reduction saves four times that in recovery costs.



FIGURE 6.7 Preventative measures to reduce occurrences of identified risks

A larger organisation would assign a disaster planning manager.

THINK ABOUT INFORMATICS 6.3

- 1 The owner of a florist, Bill, advises his brother, Ben, that he should consider their DRP confidential and not to give it out to anyone else. Why?
- 2 Their DRP states, 'In the event of fire someone should call the fire brigade.' Bill thinks that Ben could have written this better. Reword this statement.
- 3 What events should automatically trigger the review of a DRP, and why?

Relying on a single solution in an emergency may turn out to be short sighted. A good DRP will make **contingency plans** if preferred procedures happen to be unavailable. For example, if staff cannot reach the emergency meeting point, they should proceed to a secondary location. If the general manager cannot be located, contact another person.

Allocating emergency responsibilities

Responsibilities during a disaster should be allocated to specific individuals, with secondary people assigned if the primary person is absent at the time of the disaster. A good DRP will state very clearly *who* is responsible for doing *what*, *where* and *when*.

Allocated tasks include:

- ongoing preventative measures, such as backing up; annually recharging and testing fire extinguishers; changing smoke alarm batteries; and updating the DRP, emergency kit, and emergency signage
- assigning roles during emergencies, such as checking that the toilets have been evacuated, shutting down servers, retrieving DRP documentation and performing head counts
- allocating jobs during recovery, such as setting up the secondary workplace, contacting the insurance company, informing clients and partners of the disaster, acquiring replacement equipment, restoring backups and testing that recovery is complete.

Inventory

Audit all of your digital systems (hardware and software) including make, model and version numbers so that you can obtain replacements quickly. Update your audits whenever equipment changes. It would be inconvenient and troublesome if you could not restore your backups because you did not know what backup software you had been using or even the operating system the computers had used. Print the inventory and add it to your DRP emergency kit.

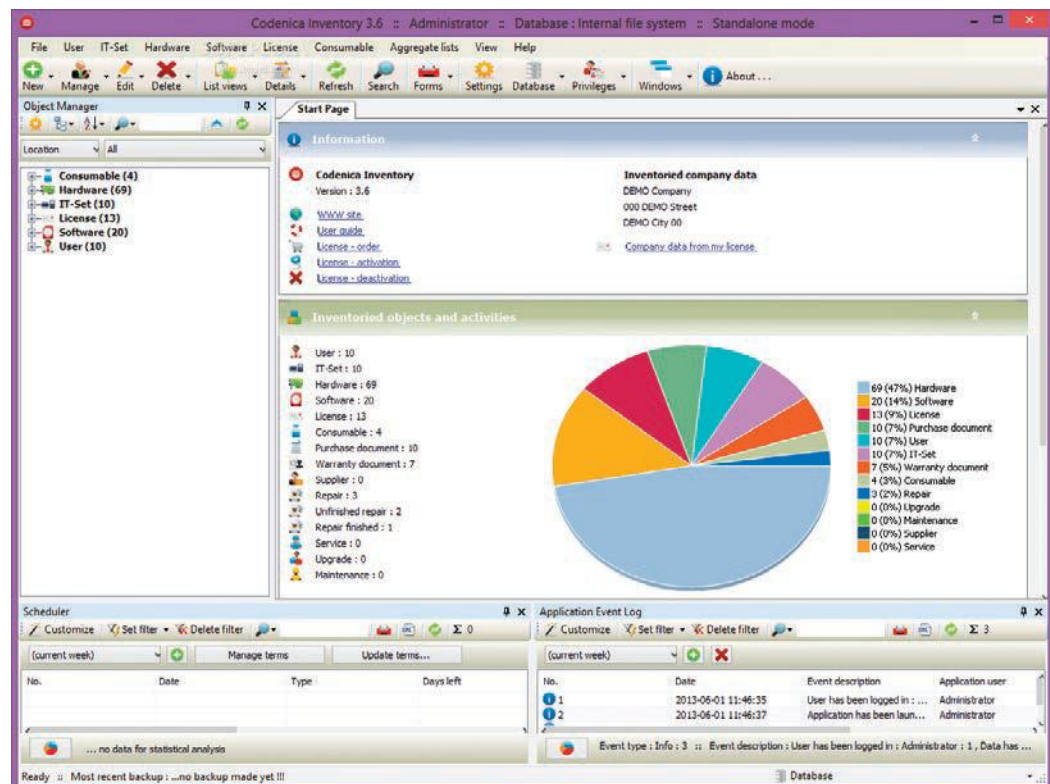


FIGURE 6.8 Inventory software makes it easier to audit your digital systems equipment

Key information

Document key information in printed form so it can be found easily during an emergency evacuation. Even the calmest, coolest head can become panicked or go blank in a disaster, so it is vital that the information not only be available electronically or via one person. An emergency kit that contains that key information should be created and kept in an easily accessible location, near an exit that is not likely to be blocked off in case of fire. Do not bury it in the bottom of a filing cabinet in the corner of the office that is furthest from the exit. Do not save it only electronically – you may have no way of accessing it when you need it the most. Remember:

- print it
- keep it near the exit
- keep a backup copy offsite, just in case.

A good DRP will assume the worst and prepare for it. It should include all the data and information you need to recover, and are unlikely to be able to memorise, such as insurance policy numbers, contact phone numbers and login details. Figure 6.9 details key information that should be included in a DRP. Remember to review the information in the emergency kit regularly to check its accuracy.

Businesses often have reciprocal agreements with other branches of their organisation, or other organisations, to provide emergency workplaces or facilities to each other.

Australians living in bushfire-prone areas know the value of a survival plan and a pre-packed kit containing emergency survival items that can be grabbed in an instant during an escape. When a fire front approaches at the speed of a car, there is no time to wonder about what to pack.

It is unwise to have only one person who knows critical passwords. If the person is injured or killed in the disaster, an organisation might find itself locked out of its own systems. Keep copies of master passwords safely stored offsite.

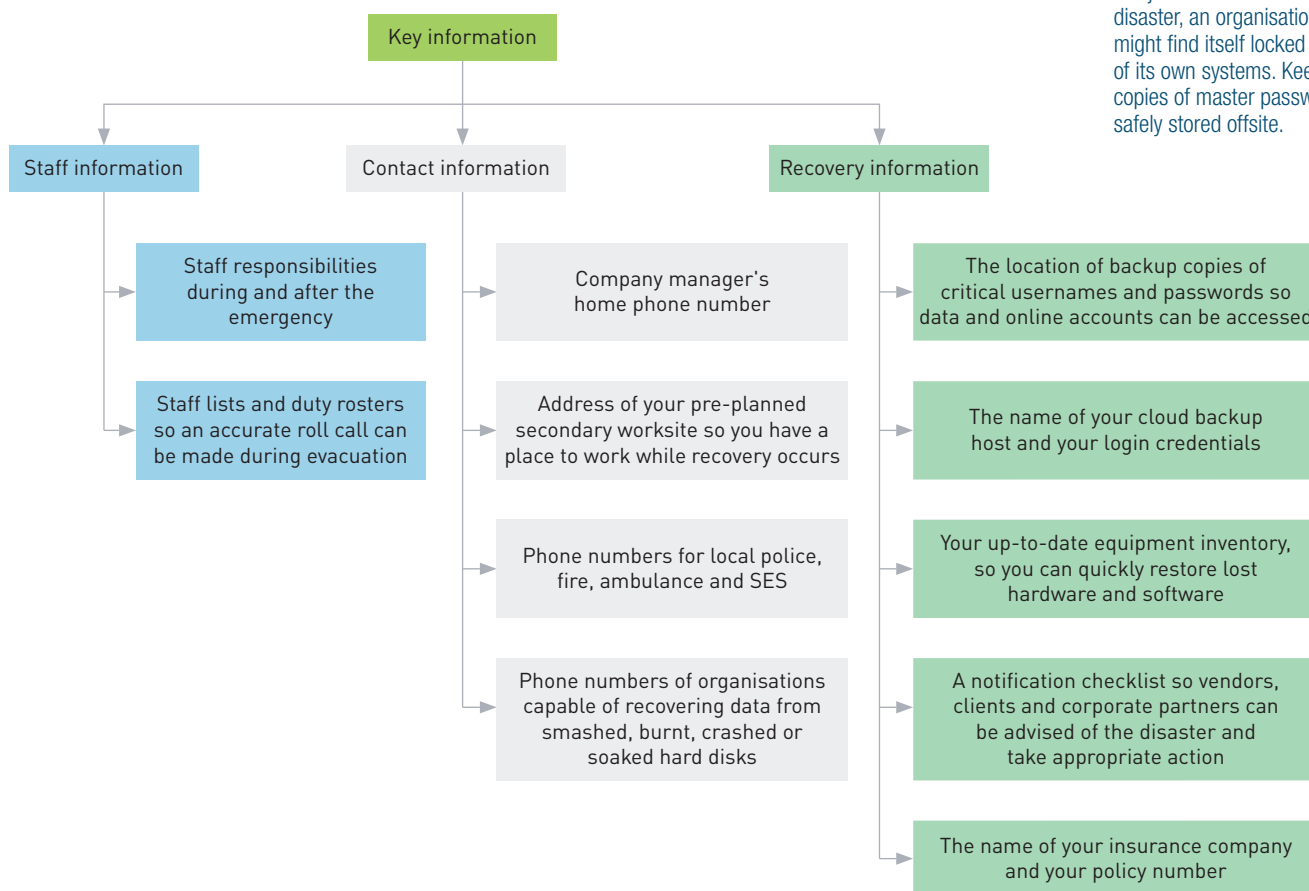


FIGURE 6.9 Key information to include in a DRP

Backup scheme

The core of a DRP must describe an **effective data backup scheme**: if an organisation’s data cannot be recovered, little else matters. The plan should specify *how* and *when* backups are done, and by *whom*.

The most common backup scheme involves daily incremental backups and weekly full backups, as described in Table 6.5. Other types of backups include differential backups and partial backups.

TABLE 6.5 Common types of backups

Frequency/Type	Description
Daily incremental	• A backup of new or changed data since the last full or incremental backup that is quick and consumes less storage space than a full backup • By itself, it will not restore a computer to its previous working state
Weekly full	• A backup of all files in the entire system, including the operating system • Basic building block of restoring a computer to its previous working state • After data loss, the most recent full backup is restored, followed by all incremental backups that have been made since then
Differential	• Similar to an incremental backup, but saves changes since the last <i>full</i> backup rather than since the last <i>incremental</i> backup
Partial	• Only backs up a portion of the file system, such as a folder or individual files within a file system, such as a single hard disk or multiple-disk array • An incremental backup can therefore be <i>full</i> incremental (saving new or changed files throughout the entire file system) or a <i>partial</i> incremental (saving new or changed files within a limited part of the file system)

Backups are usually scheduled to run at night because files that are open and in use often cannot be copied. Nightly backups also prevent networks being slowed down by the backup’s heavy disk activity.

Even minor data loss can be a crisis for modern organisations. An increasingly common solution is continuous data protection (CDP), which saves all versions of all data immediately when it is created or changed, and saves it offsite using the **cloud** or a remote network server. Online CDP providers you may have used personally during the course of your studies include Dropbox and Google Drive.

TABLE 6.6 Comparing incremental and differential backup behaviours: document versions are shown in parentheses. To restore the file system, you would firstly load Monday’s full backup. Next, you would either add all of the small incremental backups or add the single most recent larger differential backup.

Monday			
File system	Full backup	Incremental	Differential
Operating system	Operating system	N/A	N/A
Doc1 (v1)	Doc1 (v1)		
Doc2 (v1)	Doc1 (v1)		
Tuesday			
File system	Full backup	Incremental	Differential
Operating system	N/A		
Doc1 (v2)		Doc1(v2)	Doc1(v2)
Doc2 (v1)			
Doc3 (v1)		Doc3 (v1)	Doc3 (v1)

TABLE 6.6 (continued)

Wednesday			
File system	Full backup	Incremental	Differential
Operating system	N/A		
Doc1 [v2]			Doc1 [v2]
Doc2 [v2]		Doc2 [v2]	Doc2 [v2]
Doc3 [v1]			Doc3 [v1]
Doc4 [v1]		Doc4 [v1]	Doc4 [v1]

Backup media

Several backup media (methods of storage) are available, each with its own strengths and weaknesses. The most popular backup media are described in Table 6.7.

TABLE 6.7 Pros and cons of backup media

Medium	Strengths	Weaknesses
Tape  FIGURE 6.10 A quarter-inch cassette (QIC) backup tape	Common and popular in the business world	- Wears out with use and degrades over time - Limited capacity - Slow to save/restore data - Tapes and drives can be expensive - Can be erased by a strong magnetic field - Declining in favour of network or cloud backups
External hard disk drive	- Fast, high-capacity, cheap, portable, easily available and easy to use - Can be automated - Data never leaves your control	- Will fail over time - Can be damaged by mishandling - Can be erased by a strong magnetic field
Online, cloud 	- Backups occur automatically, immediately and continuously - Relatively cheap - Offsite storage - Large (sometimes unlimited) capacities are available - Every version of a document over time can be recovered	- Slow to upload large files - Valuable data is entrusted to an outsider - Will not work without an internet connection
CD/DVD	- Relatively cheap to implement - Read-only data cannot be changed later; for example, to cover up embezzlement or fraud - Very portable	- Easily damaged and degrades over time - Limited capacity: 740MB for CD, 4.7GB for DVD, 25GB for Blu-ray - Must be written manually - Read-only disks are not reusable
USB Flash drive 	- Small, very portable and cheap - Easily available and convenient	- With a limited rewrite capacity, they can fail without warning - USB 2.0 devices can be slow - Easily lost

A WORM drive, such as a non-rewriteable CD or DVD, is one that is 'Write Once Read Many' times.

Not even an on-site fireproof safe is acceptable for storing data backups. No safe is completely fireproof or theft-proof.

Warning: Data that has been compressed and/or encrypted requires decompression and decryption before it can be restored. Without the right software, the backup media will be useless. In the DRP, be sure to document the software (and decryption key, if necessary) that is required to restore data from the backup media.



Getty Images/Joby Sessions/N-Photo Magazine

FIGURE 6.11
Network-attached
storage (NAS) device

According to the Symantec Disaster Recovery Research 2007 report, 48% of backup recovery tests fail.

Backup media should be stored offsite. It is senseless to store data backups in the same location as the originals. If a disaster destroys the master copy, it will also destroy the backup copy. This is why backing up to a LAN is not wise.

In case the backup manager is absent and someone else has to take over, the backup procedure should be documented clearly so it can be replicated by another person easily. Leaving it to individuals to figure out their own backup procedures is a dangerous practice. If the sole backup expert suddenly leaves, an organisation might find itself wondering how to make and restore backups.

Backups are often compressed to achieve maximum storage in a given space. They are also often encrypted so they are useless to a thief.

Automated backup software can also carry out **data deduplication** so unnecessary copies of files are not wastefully saved multiple times.

It is far more convenient for a team to backup and protect one storage location than it is to store files across the hard disks of many separate computers. This is a good argument for creating a local area network (LAN) with central document storage on the server or a network-attached storage (NAS) device (Figure 6.11). Employees would be expected to save all of their work to this central networked location rather than to their local hard disks. Not only are an organisation's documents easier to backup, but they are able to be easily accessed and shared by anyone on the LAN.

Testing the backup scheme

If you do not regularly test your backups, you are not really protecting your data and information. You must test your backup scheme so that you can be sure your data can be restored. Assuming that your backups will be successful without confirming with a test may be catastrophic when the day comes that you actually need to use them. You may find that your backup scheme is copying the wrong network folders, the backup media have deteriorated or that they are not large enough.

Some companies have never tested a data restore from tape. Backups should be tested to ensure they are still saving the right data and can be retrieved in case of emergency.

Backup schemes can go out of date, unnoticed. New servers may be installed, but not added to the backup schedule. New software or operating system changes may suddenly cause the backup software to misbehave without being detected. Storage requirements may grow to the point that the current backup media are no longer big enough, or the midnight backup is still not finished by start of business the next day because the backup is now so large and slow. Test your backup scheme at least once a year to detect such problems.

TABLE 6.8 Methods for testing a backup scheme

Method	Explanation
Create a few typical (dummy) documents and delete them a few days later.	Can they be recovered from backup? Try recovering them two weeks, a month or a year later. Are they still recoverable? The loss of some data may not be noticed for quite some time if it is not used regularly.
Recover a backup of a database, video or spreadsheet that is several years' old.	Can the document still be opened after applications have been gradually upgraded or replaced over time? While most major applications try to retain their ability to open documents created in previous versions, there often comes a time when backward compatibility cannot be maintained and old documents can only be opened by old applications.
Perform integrity tests of tapes or hard disks.	This will detect media faults that could lead to corrupt data.
Completely restore from backup to another location or to a different hardware.	Remember that in a real disaster, the original hardware may need to be replaced with newer or different equipment. Will the restored system still boot after the operating system and applications have been restored? If the destination hardware is different to the current hardware, some software (e.g. SQL Server, Exchange Server, Windows) may not react well.
Conduct failover and failback tests.	Some advanced systems have built-in redundancy, where all data is mirrored (copied automatically to secondary equipment). If the primary system fails, the secondary equipment activates. In such a system, you need to conduct a failover test to verify that control smoothly passes to the secondary system, and then conduct a failback test to check that control is passed back to the primary system when the disaster has been fixed.
Use a hot, warm or cold site.	Some organisations use hot sites, warm sites or cold sites where a third-party organisation provides a workplace in case of disaster. • A hot site replicates a system and keeps a continuous, real-time copy of all data. You could walk in to the hot site and commence work immediately as if no disaster had ever occurred. • A warm site provides a clone of a system, but you must provide your own data backups to install onto the clone system when the disaster occurs. • A cold site can be configured to act as a replacement system while disaster recovery takes place, but it has neither cloned hardware nor live data backups. If you use a secondary site, you must also test that it can provide a functional workplace.

Before starting your backup test remember to:

- plan exactly which operations will be tested, as well as where and how they will be tested
- notify staff, in case their work will be affected
- record the result of the test
- review and improve the backup scheme in the light of results of the test.

Testing the disaster plan

Organisations must regularly practise their emergency procedures and their disaster plan. This ensures that the procedures remain appropriate. It also helps staff who might otherwise be prone to anxiety become familiar and comfortable with procedures and thus less likely to add to the chaos if a real emergency occurs. Calm and confident staff will respond efficiently and effectively, but this requires drilling and repetition of procedures.

Emergency procedures need to be documented in a form that is very clear and simple, so mistakes are not made in the heat of the moment.

Do not try a full test restore on your primary digital system. Downing the system and performing the restoration will stop all network activity for hours and, if it fails, your network will be broken and data will be lost, which is exactly what the test was trying to avoid!

The 'warmer' your secondary site is, the more it will cost.

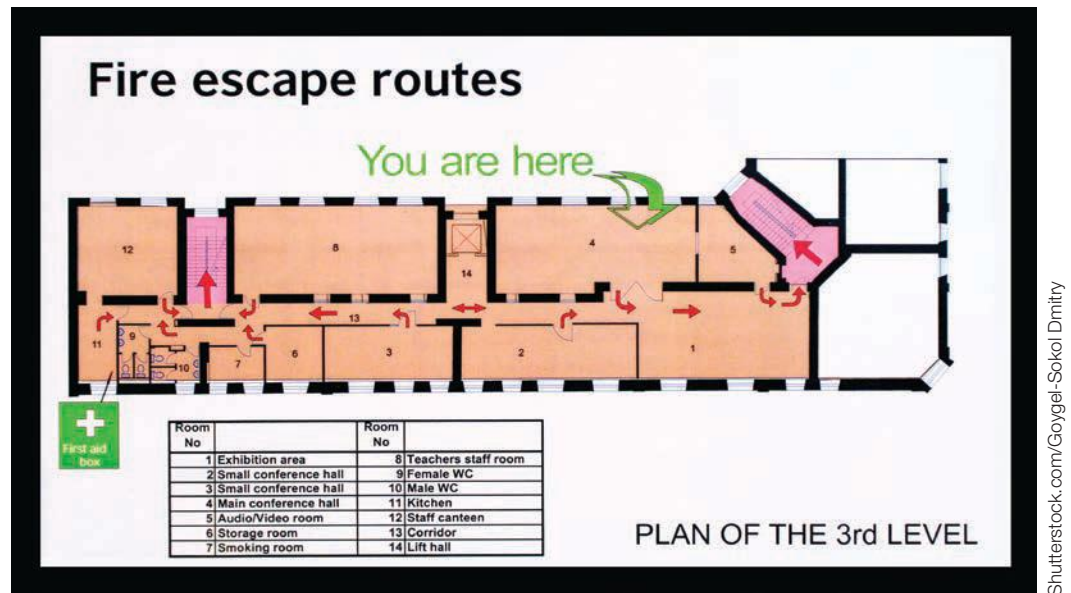


FIGURE 6.12 An emergency evacuation route map



FIGURE 6.13 Simple emergency procedure documentation

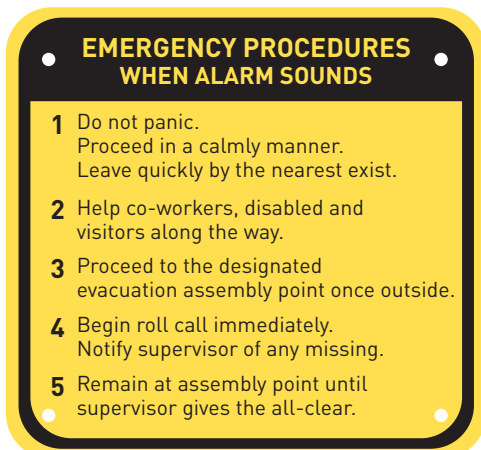


FIGURE 6.14 Text-based emergency procedures

Staff should be assigned to specific roles so everyone knows what they need to do and no vital steps are missed. Use rehearsal drills to practise evacuation. It is important to make sure you know how many people need to be accounted for and who is offsite or absent. Evacuating to the designated assembly point regularly also reminds staff to go to the correct place. If staff scatter during an emergency, it will be very difficult to account for the missing.

Practise using firefighting equipment. It is important that all staff members know where the equipment is, as well as how it works.

Perform emergency data backups. Even the best backups may be hours out of date, so new data may still be lost after restoring from backups. If you have any advance warning of disaster, you may still have time to copy the latest data. Always make sure you have storage media at hand that can be used for an emergency data backup. Can you quickly find where the new data is stored on the network?

Shut systems down safely. If staff panic and shut down systems without caution, there is a risk of losing unsaved data and damaging sensitive hardware. Exercise particular care with network servers.

Practise notifying emergency services. Do staff know what information to provide, such as the nature of the emergency and the address? Consider including a mobile phone in the emergency kit.

Emergency drills should be both pre-announced and surprise events, and should be conducted regularly enough that all staff are familiar with the procedures. Feedback should be collected after drills and errors or omissions in the plan should be identified and fixed.

Recovery strategies

After a disaster comes recovery. The workplace is restored. Replacement equipment is acquired and installed. Data backups are restored. Complete recovery is verified according to pre-planned criteria, such as that shown in Figure 6.15.

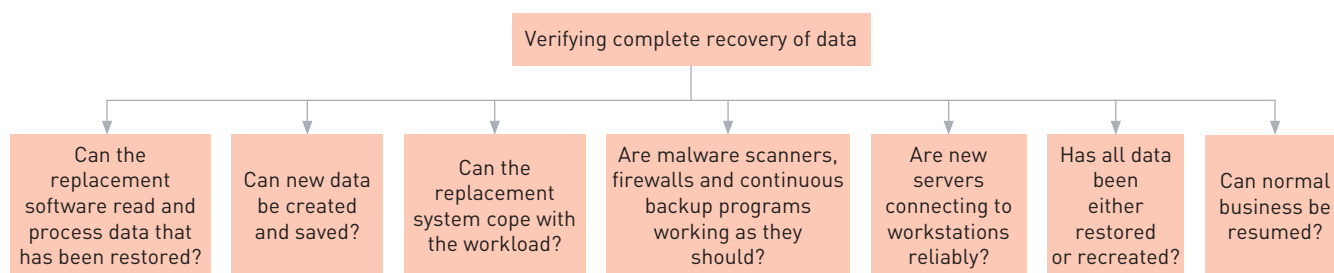


FIGURE 6.15 Criteria for verifying the complete recovery of data

Consequences of security failure

Organisations that fail to secure the data and information they hold, and suffer losses or breaches as a result of this, may be subject to penalties or prosecution. For example, if private personal information is lost, damaged or exposed, organisations may be prosecuted under the Privacy Act. If tax records are lost, organisations may be penalised or prosecuted by the Australian Taxation Office. However, the consequences of failing to protect stored and communicated data go beyond the legal and punitive.

Organisations may struggle to conduct normal business because, in some cases, data loss means they are unable to pay wages to staff or pay their suppliers. Further costs, labour and disruptions will be caused by the need to recreate lost or damaged data, and repair or replace damaged, destroyed or stolen equipment. If normal business is disrupted, the organisation will also lose income.

Data security failures may result in organisations losing their trade secrets to competitors. Depending on the level of publicity involved in the breach, the organisation may also lose public reputation as trustworthy and thus reduced customer loyalty. Even its stock market value may decline.

In 2011, Sony's Playstation Network incurred a major security failure that became one of the most famous security breaches in history, with a variety of consequences. The infographic shown as Figure 6.17 outlines the security failure, how it unfolded and how it could have been avoided.

Some of the potential consequences of security failures for organisations were discussed in Chapter 3.

The highly destructive Stuxnet worm was released in 2005, but not discovered as a threat until 2010. The effects of the damage over the course of those five years are unknown.

After a security breach in 2009, Heartland Payment Systems (credit card processors) in the USA lost 50% of their market value. They accrued \$139.4 million in breach-related expenses, including legal fees, forensic costs, fines and other settlement costs. It took them over a year to recover from the stock market plunge.



FIGURE 6.16 Consequences of data security failure to organisations



NelsonNet additional resource: Figure 6.17 Infographic: Playstation Network hack, p. 290

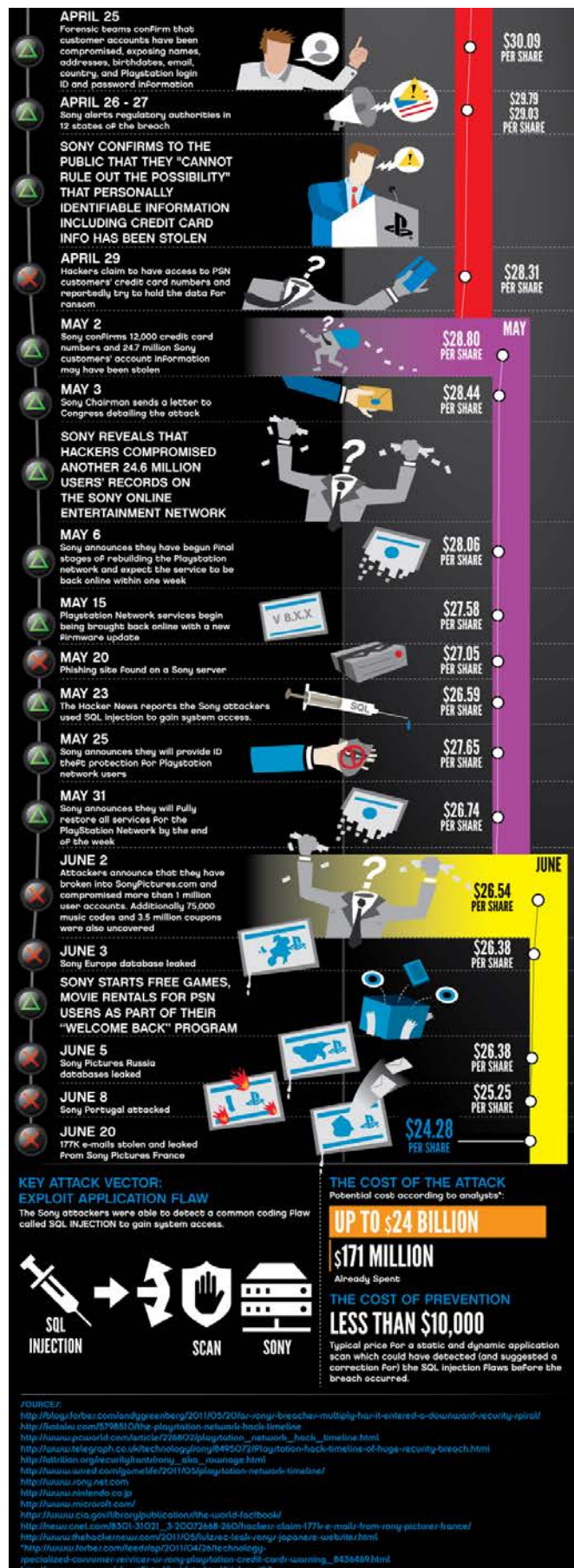
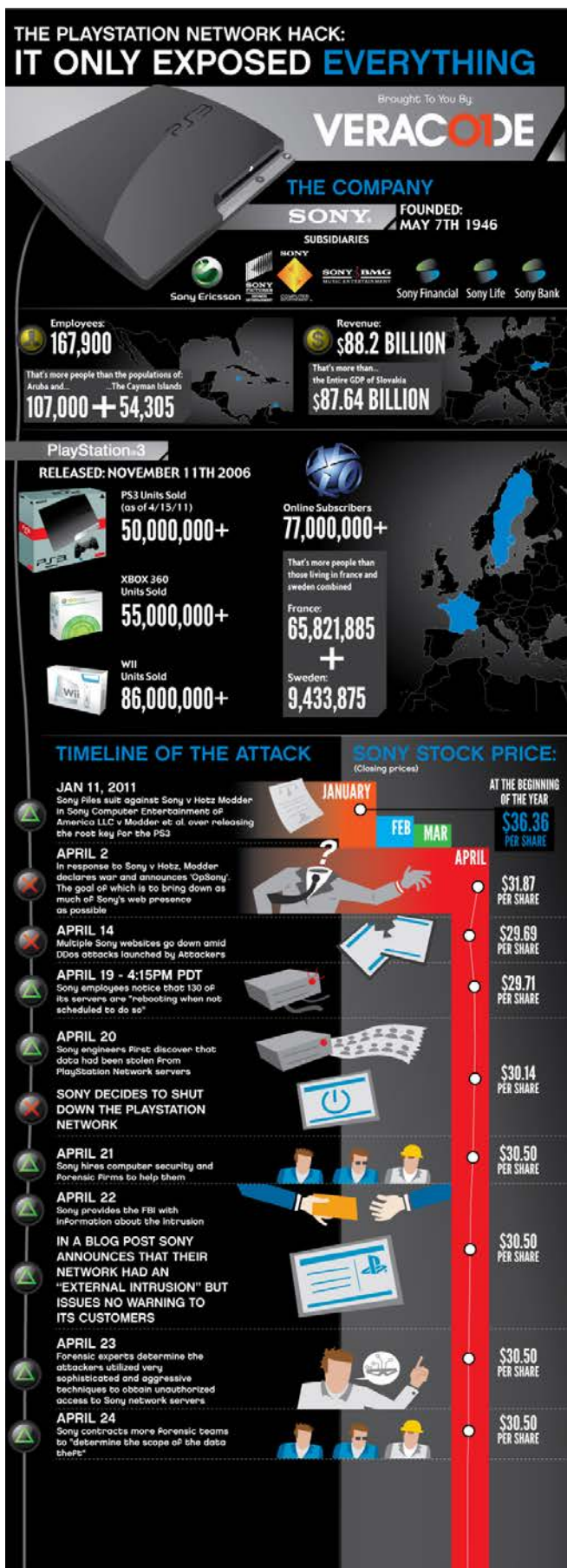


FIGURE 6.17 The Playstation Network hack: download this infographic in full size from the NelsonNet student website

Components of data and information management

Data and information management is carried out on information systems, which are made up of **people, processes, digital systems** and **data** itself. The following sections discuss the people, processes and digital systems.

People

The people in information systems are both users and support personnel and they fulfil many different roles. The people in information systems are described in Figure 6.18.

THINK ABOUT INFORMATICS 6.4

Research the roles of people in information systems. Find information on:

- one support role
- one development role
- one management role.

Write a one-paragraph (4–5 sentences) summary of each role.

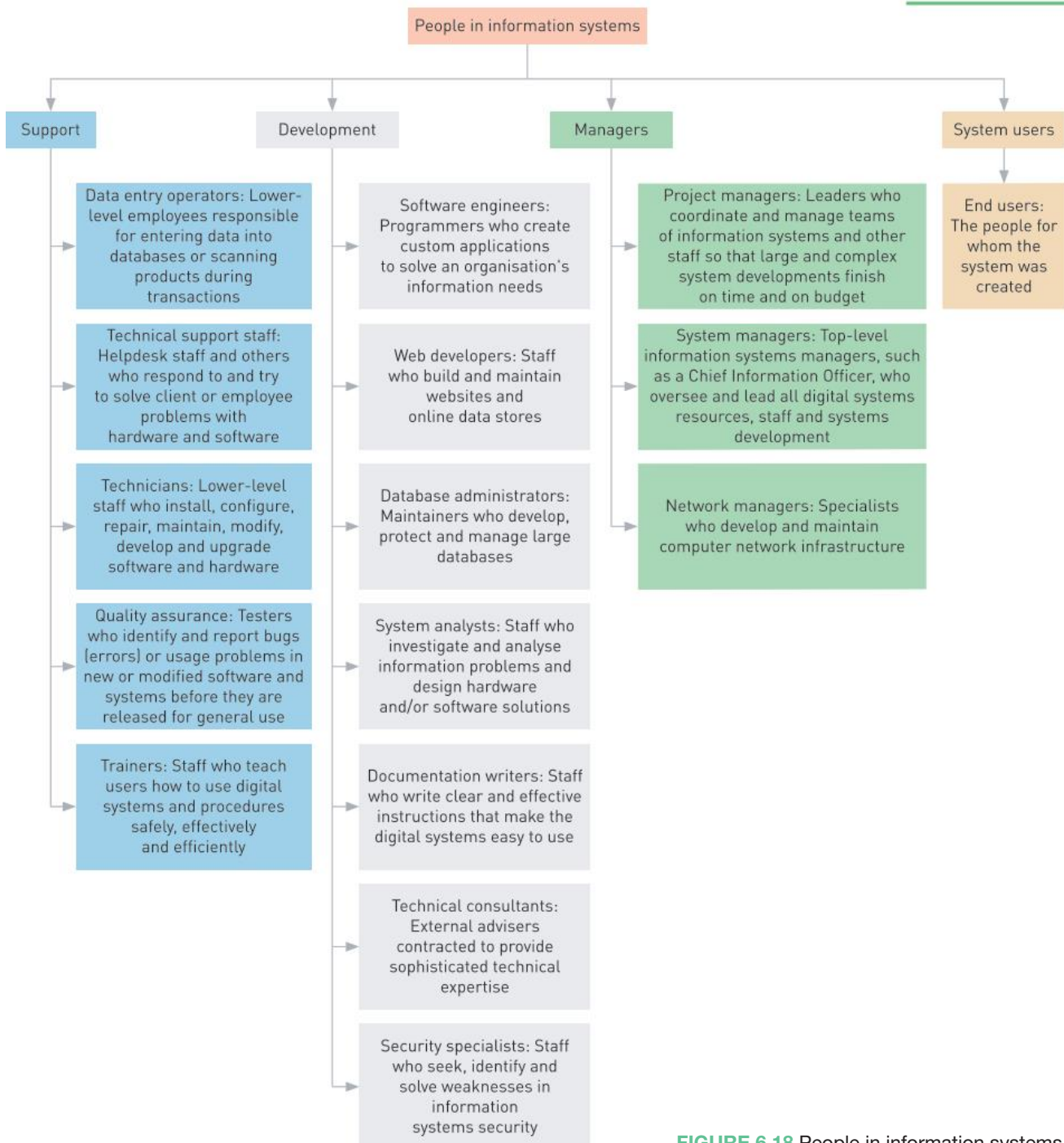


FIGURE 6.18 People in information systems

Processes

The typical processes in data and information management include acquisition, input, validation, processing, storage, retrieval, output, communication and disposal.

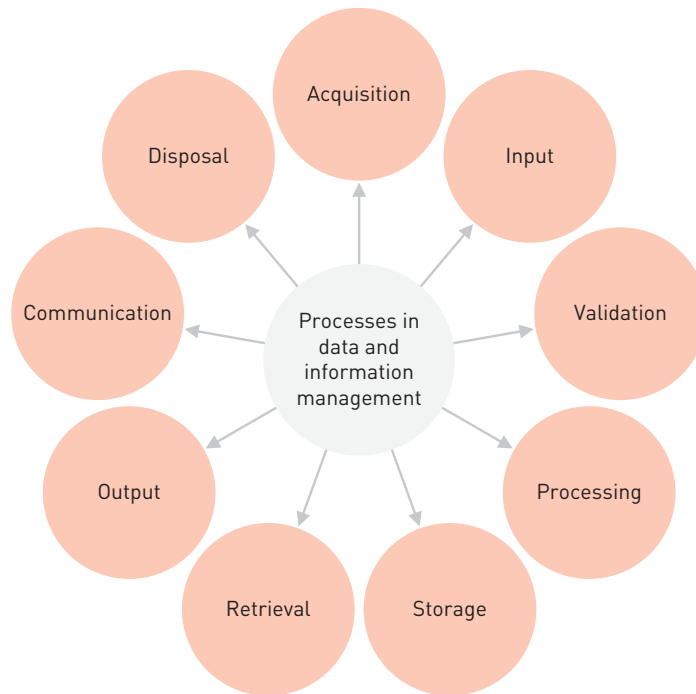


FIGURE 6.19 Typical processes in data and information management

The following section discusses each of these processes in more detail.

Acquisition

Acquisition gathers raw data from the world outside the information system. Data may also be acquired electronically in situations where the system monitors sensor inputs such as keystrokes, automated CCTV cameras, incoming missiles on radar, water levels in dams, vehicular traffic sensors or rain gauges.

Input

Input requires the user to enter acquired data into an information system, using a keyboard, mouse, barcode, QR code, voice recognition, custom control panel (such as a Myki card reader) or another input device.

Validation

Validation checks that input data exists and falls within reasonable data types and ranges. Manual validation, such as proofreading, is used to check that data makes logical sense. However, it is important to note that validation cannot and does not have the capacity to check the *accuracy* of data.

Processing

People cannot interpret or understand raw input data, so **processing** is used to manipulate or convert the data into digestible, useful information. This information will therefore have meaning and value to the people who need to use it. Processing tasks are many and varied, but a large part of processing involves converting data or information into a different form.

One way to detect data entry errors is by using double entry checks. Data are entered twice and the two entries are checked for consistency.

Refer to Chapter 2 for discussion of data sources (surveys, interviews, observation, databases, websites), primary, secondary, qualitative and quantitative data.

Converting is achieved using many different methods. For example, you might convert the **file format** of an audio file from FLAC to MP3. Converting may also take the form of changing **units**, such as going from US dollars to Australian dollars, or from Fahrenheit to Celsius.

Other examples of processing tasks include summarising, such as calculating totals, averages and standard deviations; translating both human and programming languages; pattern matching, such as face recognition, spell checking and **fuzzy searches**; and predicting, such as forecasting the weather or even predictive text on smartphones. The types of processing tasks are varied and expansive. These are merely a few. Note that the output of processing should always be tested.

Storage

Storage involves both seemingly simple and more complex processes. Data and information is stored in fields, records, databases, XML files, sequential or random files on disk, or in the cloud. Copying data and information to another place and keeping the original is called backing up. As discussed earlier in the chapter, backups are a crucial aspect of information management.

Sometimes stored data is unneeded, so it can be archived. Archiving involves moving the unneeded data to offline storage, such as DVD, and deleting the original. Other examples of storage processes include compression, which reduces storage space and communication times, and encryption, which can be used to protect data from being viewed, copied or modified by unauthorised people. Popular compression formats include ZIP, RAR, lossless run-length encoding Lempel-Ziv and LZW. Popular encryption types include PGP, SSL, TLS and Macrovision. This is not an exhaustive list of storage processes. Many other storage processes exist.

Retrieval

Data and information may be retrieved from local, networked or online databases. Data may also be retrieved from disk drives, websites and the cloud. To retrieve data, you will often need to search by one or more fields in a database using queries.

Output

Output is the process for when information is formatted and presented for maximum ease of interpretation and use. Output tends to employ colour and visual elements such as photos, graphs, animations and other data visualisations. Multimodal solutions, gauges and GUI controls such as radio buttons and sliders are also common. Output may be displayed on monitors, printers, data projectors, speakers, LED lights, dials, LCD displays, virtual reality helmets and custom-built control panels (such as the flight simulators used to train jet pilots) and other devices.

Communication

Communication involves the use of both wired and wireless networks. Wired and wireless networking is discussed briefly in this chapter on pages 296–8, and in more depth in Chapter 3.

Disposal

Disposal refers to the safe removal of unwanted or obsolete data and information. Disposal techniques include deletion, file wiping, reformatting of disks and the physical destruction of obsolete media. Disposal methods need to ensure that personal and sensitive information is still protected from exposure, so they need to be appropriately secure.

THINK ABOUT INFORMATICS 6.5

Research and list two examples of each of the following additional processing tasks.

- Categorising
- Standardising
- Statistical analysis
- Sorting
- Reformatting
- Repairing
- Comparing
- Reorganising
- Aggregating

The *Copyright Act 1968* encourages the use of encryption and forbids the importation, possession or use of any technology, such as Blu-ray rippers, designed to remove or bypass the encryption of digital intellectual property.

THINK ABOUT INFORMATICS 6.6

Research and list two examples of each of the following additional types of storage processes.

- Redundancy and mirroring
- Caching
- Error detection
- Cleansing
- Appending
- Verification

See also Chapter 3 of *Computing VCE Units 1 & 2*, which is titled 'Networks'.



Information storage and disposal policy: Department of Health, Government of Western Australia (PDF).

Organisations and government departments may have policies that govern how they handle both the storage and disposal of the data and information that they hold. The PDF of the information storage and disposal policy for the Western Australian Health Department in the weblink opposite is a particularly useful example of how organisations address storage and disposal.



FIGURE 6.20 A hard disk before physical shredding



FIGURE 6.21 A hard disk after physical shredding – the ultimate form of disposal

Digital systems

Digital systems in information systems are made up of three key components: hardware, software and networks. Digital systems are integral to how data is stored and disposed of. Both independently and collectively they contribute to controlling how the integrity and security of data and information can be managed.

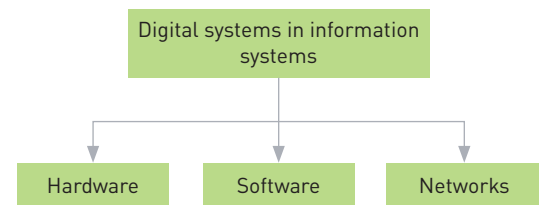


FIGURE 6.22 Components of digital systems in information management

Hardware

Hardware is the physical equipment used in a digital system. It is composed of items such as CPUs, keyboards, modems, disk drives, monitors, printers and mice. Hardware, discussed in more depth in Chapter 3, is made up of input hardware (Table 3.8), output hardware (Table 3.9), storage hardware (Table 3.10) and communication hardware (Table 3.11).

Software

Software is made up of programmed instructions that control the behaviour of CPUs and computer systems. Written in source code, the instructions are converted by compiler software into executable code, which are commands that the CPU can understand, such as `notepad.exe`. Software is usually divided into three categories: system, applications and utilities.

System software

System software is programming that runs the information system and allows other software to run. The most prominent part of system software is the operating system, such as Windows, Mac OS, Linux, Android and Chromium. It also includes device drivers, which translate operating system commands into instructions for a particular piece of hardware (such as a printer), function libraries (such as DLL files), and firmware to control devices such as modems, TVs, hard disk drives and MP3 players. GUI (graphical user interface) operating systems make underlying complex system software easy to use with windows, icons, menus and pointers (WIMP).

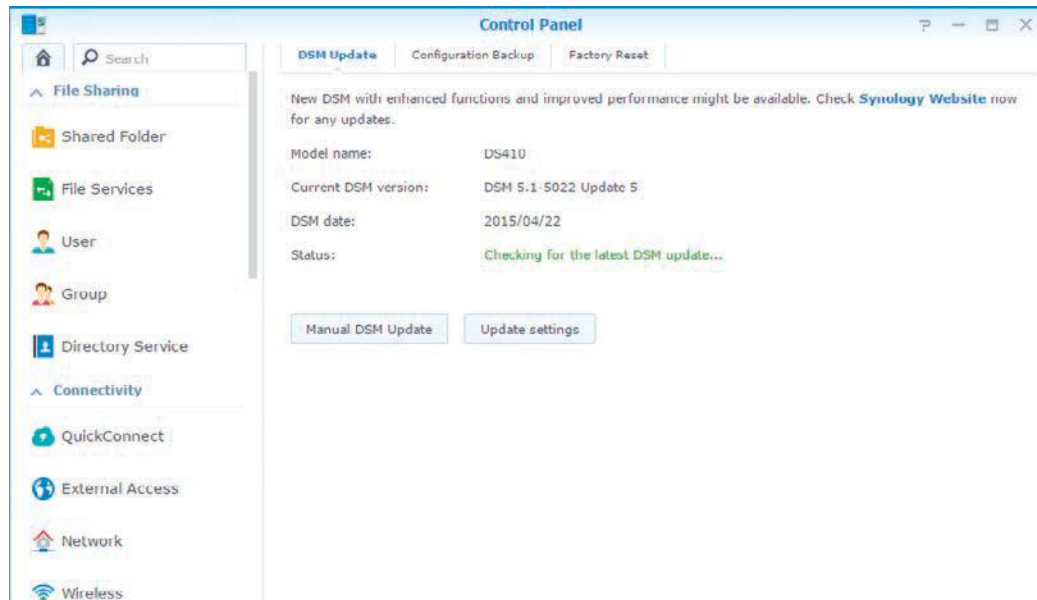


FIGURE 6.23 Many electronic devices, such as the NAS shown above, now can have their operating system's firmware updated via download

Applications

Applications are programs used to process data into useful information to get work done or complete larger tasks. Examples of applications include word processors such as Microsoft Word, spreadsheets such as Microsoft Excel, drawing programs such as Adobe Illustrator, open-source programs such as GanttProject, custom-written programs for specialised purposes and web browsers such as Google Chrome.

Utilities

Utilities are small programs with a very limited scope that usually perform a single, useful task or function. In doing so, they generally extend the usefulness and functionality of a computer system. Examples of utilities include programs that partition hard disks, burn DVDs, scan for viruses, capture screenshots, convert video formats and transfer files using FTP. Specific utilities in the Windows operating system include the Calculator and Character Map.

Some people also recognise a fourth category: development software, which includes programming languages, compilers, debuggers and code editors. Others categorise utilities under system software.

Networks

A network is made up of computers and devices connected by communications channels that facilitate communication and the sharing of resources between users. Resources include data, information, internet connections, printers and storage. Networks can be internal to an organisation or cover the whole world by connecting to the internet.

The simplest form of network, found in homes and small offices, is the local area network (LAN), consisting of two or more **nodes** in a limited area, such as a few rooms or a single building. While most LANs have only a few computers, there can be any number of computers in a LAN. The defining characteristic of a LAN is that its nodes are physically close to each other, and not spread over a broad geographical area.

The technologies used in networks are varied. Refer to Chapter 3 for more information.

There are distinct advantages and disadvantages to using networks, as well as cloud computing, to store and dispose of data and information.

Networks and cloud computing

Advantages of networks

Storing data and information on a network is useful for a number of reasons. Resources stored on a network, perhaps using a NAS, are centralised, which makes sharing quicker, easier and document storage more secure. Databases and documents can be accessed simultaneously by multiple users.

Storing documents on individual hard drives would create a file versioning problem, where every user would have their own copy of a file that would be edited separately. Storing a central master copy on a network ensures that users are always using the one definitive version of a file. In an organisation, staff also need to access and work on each other's files directly. Storing them on a network ensures that this is easy to do.

Creating disaster recovery plans and managing backup schemes is easier if storage is network-based (or cloud-based). The IT department of an organisation will be able to focus on one or two specific drives to restore and backup rather than the drives of every individual user in the organisation. This is much easier to manage.

It is critical that data stored on networks be disposed of securely. An advantage of disposing of data and information on a network is that network media, such as disk drives or NAS, can be reused by being easily overwritten. Data disposal can also be managed by a larger organisational policy when it is stored on a network.

Disadvantages of networks

The drawback of storing data and information on networks is the same risk that is always inherent in centralisation: once people rely on a single piece of core equipment, such as a server, any equipment failures will have widespread implications. Users may not be able to log in to their computers, print, load any documents stored on the network, check email, or access a master database. When a network 'goes down', business effectively stops for many modern organisations because they no longer have alternative manual procedures to let them continue normal operations.

Many network outages can be prevented by thinking ahead.

- Buy quality equipment. Cheaper alternatives often use poor quality components that fail quickly as a result of stress or age.
- Protect core equipment with an uninterruptible power supply (UPS).

- Keep core equipment away from people in a cool, dust-free environment, preferably air conditioned.
- Have backup replacements for key equipment, such as a switch, modem, power supply and network cable.
- Keep the operating systems and anti-virus software of **workstations** and servers up-to-date. Apply security patches immediately to prevent malware infection.
- Train users in the proper use of the network to avoid accidental or negligent damage or data loss.
- Conduct routine maintenance of equipment.

A disadvantage of disposing of data on a network is that it can be quite difficult to truly dispose of data securely. In addition, truly disposing of some network data can require a fair amount of labour – such as demagnetising or destroying the magnetic media on which it is stored.

The cloud



The cloud is any internet-based software and data storage facility that is used as an alternative to, or additional to, **local** storage. It can also include online services that carry out processing on a user's behalf.

Cloud-based email is the most common example of cloud processing and storage. Originally, people used the **POP** protocol for email. This required installing email software, such as Mozilla Thunderbird, on a local computer, downloading emails from the mail server and storing them on a particular computer.

People with multiple computers, such as a home desktop, a laptop, and a work computer, had to download the same emails onto each device and then perform the same mail management, such as sorting or deleting on each individual device. It was a frustrating waste of bandwidth, storage, time and patience.

In 1986, the Internet Message Access Protocol (**IMAP**) email protocol was invented. This is essentially a cloud application. Apart from a web browser, no software has to be installed locally because the programming lives on the internet. All emails are also stored on the internet, and can be accessed from any computer without having to download them all again. Mail and software preferences can be managed centrally. People could begin to check their email using any computer at any location.

Most people now use IMAP cloud-based email services, such as Gmail, Hotmail, Outlook and Yahoo! mail.

The simplicity and ease of cloud-based email has since attracted many people and businesses. Many use similar services for carrying out their processing in the cloud rather than on their local computer for tasks other than email. Tasks include word processing, spreadsheets, forms for data collection, slideshows, drawing, calendars, games, databases, project management, image editing and special-purpose applications.

Advantages

Storing data and information in the cloud allows for great mobility. The user can access their material at any time from any location around the world. In addition, storing data in the cloud rather than on a network or servers reduces the need for organisations to rely on hardware for storage, which reduces setup costs. In cases of disasters, data and information stored in the cloud is more easily recoverable than anything that has been stored on local equipment. This also saves money on maintaining and upgrading this storage hardware.

For many people, the main benefit of cloud computing is that the cloud host is using their massive processor power and storage capacity to run applications so you only need a very cheap, low-powered device with few resources to view the results.

Similar to network storage, cloud storage also allows simultaneous collaborative data access. It also reduces the risk of versioning. However, multiple copies are kept in different locations to guarantee backups and data redundancy.

There is less need for local data protection because sensitive data is not stored locally. There is less chance it can be stolen or damaged. There are massive amounts of storage space available, and even small, free accounts tend to have sizeable amounts of space for single users. The costs for professional use are regular and predictable, unlike the ad hoc spending needed to pay for local digital systems resources.

There are also cloud versions available now of popular major applications, including Microsoft Office 365 and Adobe Creative Cloud. Managers enjoy not having to buy, maintain, and upgrade expensive software suites, servers and high performance computers because all the processing is done by online servers. All a staff member needs is an inexpensive laptop or **Chromebook** with a web browser to display the processing completed by a remote supercomputer.



Getty Images/Glen Chapman/AFP

FIGURE 6.24 A cheap, light, cloud-based Chromebook

Disadvantages

The main disadvantage of storing data and information in the cloud is that without an internet connection, you cannot access it. Data access is also slow and expensive when compared with local storage, especially when uploading massive media files using an **ADSL** internet connection.

You are also entrusting your data to a third-party corporation who may not secure it properly, or may even misuse or abuse it. You will have ongoing costs for storage, processing and access to software applications. If the host fails or cancels the account, you will lose access to your data. If the cloud host is hacked or fails in its own security, it could open up valuable data to loss, theft or damage. Cloud hosts are frequent hacker targets.

Once committed to a cloud service's software, you might find yourself 'locked in' to their software. As result, you may find it too difficult to move to a different host if it means transferring huge amounts of data and converting documents to a different service's software.

Evaluating information management strategies

To evaluate whether information management strategies are effective, you can apply a number of criteria. These criteria should correlate with the functional and non-functional requirements for a solution – the reasons for its development and design, as well as its ease of use. These requirements are effectively measures of success that will be monitored over time.

Monitoring may use logs and records; for example, to count numbers of errors or measure speed. It may also involve using interviews or questionnaires for opinion-based issues, such as ease of use, happiness or feeling of comfort using the information system.

Evaluation does not check that a solution can cope with suddenly removing a power plug, or that it calculates the total of invoices correctly. Such issues are determined during testing before a solution is released. To reiterate: Evaluation checks that a solution is achieving the goals or objectives for which it was originally designed.

Evaluation criteria fall under effectiveness or efficiency.

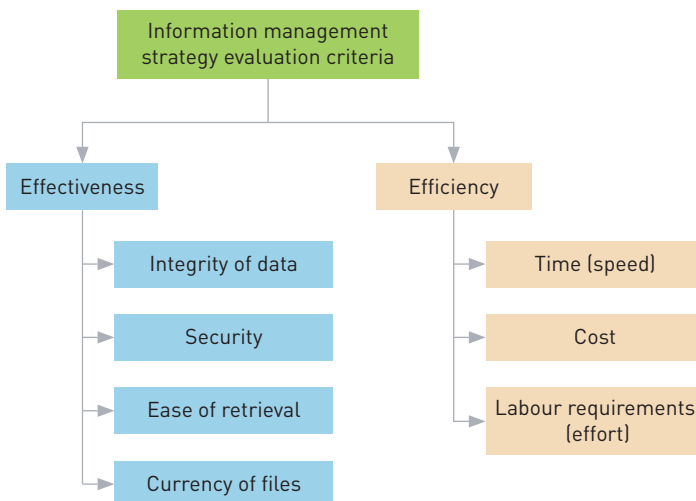


FIGURE 6.25 Information management strategy evaluation criteria

Effectiveness

Effectiveness criteria measure the quality of the strategy – how well it performs. The measures of effectiveness in an information management strategy are integrity of data, security, ease of retrieval and currency of files. Each measure is expanded upon in the section that follows.

Integrity of data

Generated information shows that queries return data that is both relevant and accurate. There should be few or no complaints from customers or users about inaccurate data or information about their accounts. Network logs should report few or no occasions where data storage errors have occurred. Unreasonable data should be reliably and appropriately queried or rejected.

Security

There should be no reports lodged by users about deleted data, or their data being accessed, used or revealed by unauthorised parties. Data disposal procedures should not have accidentally exposed sensitive data and information. There should be no incidences of successful security breaches – from hackers, crackers or otherwise.

Ease of retrieval

Depending on the type of information system, staff surveys or customer feedback will reveal that they are happy with the ease of retrieving relevant data and information from the database or strategy. It should not be unnecessarily complicated or difficult to locate the relevant data and information using queries, searches, sorting, filtering and so on. This also relates into efficiency criteria – time (speed); that is, users should not report that it is unnecessarily time-consuming to retrieve the data and information because of the complexity of retrieval processes.

Currency of files

Logs should report few or no complaints from staff or customers about data and information being out of date. This also relates to integrity of data – timeliness. An example of a customer complaint for currency of files might include a university website that has class timetables listed that are more than a year old, but none for the current or upcoming semesters.

Efficiency criteria

Efficiency criteria relate to measurable, numeric or factual issues.

Time (speed)

This criterion evaluates a strategy based on its speed. Logs should report few or no complaints from staff or customers about the strategy being slow to respond, tending to lag or slow to use overall.

Cost

This criterion evaluates a strategy based on ongoing costs. If the ongoing costs of operating the strategy outweigh the benefits of using the strategy, the strategy is not efficient. Evaluation criteria can address start-up costs, but it is more likely that ongoing costs will be of more interest here. For example, if the strategy included a sophisticated type of laser printer and software within it, ongoing costs would include supplies such as toner, drum cartridges and paper, as well as power.

Labour requirements (effort)

This criterion evaluates a strategy based on the number of labour hours required to produce output. If the labour hours required to produce output are exceedingly high, they have an effect also on cost and have a negative effect on the efficiency of the strategy.

Next steps

Work your way through the chapter summary that follows in preparation for completing your end-of-year exams. Follow your teacher's instructions with regard to the SAC.

ESSENTIAL TERMS

accidental threats user incompetence or error, equipment failure, electrical problems, software failure or misconfiguration, loss of data devices, unsafe email or web usage, poor backups, and mismanagement

ADSL an acronym of Asymmetric Digital Subscriber Line; 'asymmetric' means that upload and download speeds are not equal. Typically, download speeds may be 10 times faster than uploads since most home and business users download much more than they upload

botnet malware software that allows an unsuspecting victim's computer (the zombie) to be controlled remotely by a cracker to attack other sites, or to use the victim to send spam

Chromebook a laptop with minimal processing power and little storage capacity that runs Google's Chromium OS operating system; they rely on the internet to connect to cloud services to perform processing and storage

cloud internet-based data storage and/or processing

DDoS attacks when a botmaster instructs thousands of infected zombie computers to simultaneously flood a target system with massive numbers of requests for data and the target, unable to cope with the mass of incoming traffic, collapses under the onslaught; DDoS attacks are used by criminals to blackmail organisations, or by groups wishing to harm an enemy

deliberate threats malware, hackers and fraudsters, identity theft, equipment theft, vandalism, espionage, disgruntled employees and service provider misbehaviour

disaster recovery plans (DRP) prepared in advance to explain how to prepare for, survive and recover from catastrophic data loss

drive-by downloads infections caused by visiting infected or malicious sites that upload malware when a visitor clicks a deceptive link on a webpage

event-based threats natural disasters – such as fire or flood – and manmade disasters such as war, civil unrest and terrorism

fuzzy search a search that returns results that are approximate rather than exact matches; for example, searching for 'analyze data' also finds results for 'analyse data'

IMAP abbreviation of internet message access protocol, an email protocol

keyloggers secretly installed utilities that record keystrokes and mouse clicks on infected computers and send data like passwords and credit card information to criminals

local installed 'where the user is' as opposed to being available online and/or remotely

node any device that actively sends, receives, or forwards data on a network, including computers, servers, printers, routers, switches, WAPs (wireless access points) and NAS (network-attached storage) devices

POP an abbreviation of post office protocol, an email protocol; version 1 was developed in 1984

ransomware data on a vulnerable system that has been encrypted by an attacker who only supplies the unlocking key if a ransom is paid

testing checking whether output information is accurate

workstation a networked computer; a standalone computer is not networked

IMPORTANT FACTS

- 1 **Threats to data and information** are accidental, deliberate, technical and events-based.
- 2 A DRP starts with a **risk assessment** to discover the types and probabilities of threats your data will face.
- 3 A DRP assigns **disaster responsibilities** to individuals.
- 4 A DRP includes **inventories** of hardware and software that will need to be replaced.
- 5 A DRP contains **vital information** that will be needed after a disaster.
- 6 The main feature of a DRP is a **good backup** and **data recovery scheme**.

- 7 **Backups** should be regular, tested, documented and stored offsite.
- 8 **Continuous data protection** backs up new or changed data immediately to the cloud.
- 9 A DRP explains **emergency procedures** to take during emergencies, such as an evacuation plan, system shutdown and how to use firefighting equipment.



CHAPTER SUMMARY

- 10 A **DRP** explains how to restore hardware, software and data, and how to evaluate the success of the recovery.
- 11 There are many serious consequences of a failure of **data security**, including loss of reputation, inability to carry on normal business and loss of trade secrets to competitors.
- 12 There are many roles for **people** in systems, such as programmers, security specialists and technicians.
- 13 **Data management processes** are as follows: acquisition, input, validation, processing, storage, retrieval, communication and disposal.
- 14 **Digital system hardware** includes equipment to do input, processing, storage, output and communication.
- 15 **Software** controls **hardware** and includes system software, applications and utilities.
- 16 **Cloud advantages** include mobility, cost savings and data protection.
- 17 **Cloud disadvantages** include getting locked in to a provider, entrusting data to outsiders and complete failure if the internet connection is lost.
- 18 There are many **benefits** to creating a network, including cost-saving, central control and communication.
- 19 **Network protection** includes investing in quality equipment, protecting and maintaining hardware and software, and training users.

TEST YOUR KNOWLEDGE

- 1 Justify the need for a disaster recovery plan to a reluctant small-business owner.
- 2 List four main consequences if a business ignores or violates security measures, and justify your choices.
- 3 What are the main features of an effective backup scheme?
- 4 What criteria would you use to evaluate your family's information management needs, and what methods could you use to evaluate those criteria?
- 5 Imagine you are the owner of a small business and are advertising for an IT support employee. Write a job description for the following roles and describe the responsibilities for each role.
 - a Technician c Security specialist
 - b Trainer d Project manager
- 6 Explain, using examples, how the processes of validation, testing and evaluation differ.
- 7 List two hardware devices used during each of these stages: input, processing, storage, output and communication. For each pair of devices, contrast their differences, strengths and weaknesses.
- 8 Complete this table by listing the top five data defences. If in doubt, use your local area as a reference.



Review quiz

	Physical security	Software security
1		
2		
3		
4		
5		

- 9 Describe five important procedures that should be used to protect data and information. Do not repeat items used in the previous question.
- 10 Complete this table of the top five threats to data. If in doubt, use your local area as a reference.

	Accidental	Deliberate	Event-based
1			
2			
3			
4			
5			

- 11 Write a publicity webpage for a cloud service provider that promotes the advantages of cloud computing.
- 12 Explain why a business with even just two employees should set up a local area network.

APPLY YOUR KNOWLEDGE



This task should not be used as a complete outcome for Informatics Unit 4, Outcome 2 because it does not include knowledge of relevant legislation.

Your cousin Ben is a small business owner in Frankston with five employees. His company sells flower pots across Australia. His one-room office consists of five stand-alone desktop computers and a laptop. One computer has an ADSL internet connection and a printer.

- 1 Justify why a business of Ben's size should have a disaster recovery plan.
- 2 Create a draft disaster recovery plan, including:
 - a risk assessment outlining the threats his data faces in Frankston
 - a data backup scheme
 - evacuation procedures
 - data restoration
 - a method of testing the plan.
- 3 Explain the possible consequences of Ben not taking the threats to his data seriously.
- 4 If Ben did adopt different information management practices, how could he tell whether they were better than his old practices?
- 5 What IT staff might he need to use to support his business?
- 6 Waste paper, often containing financial information and customer details, are crunched up and thrown into a rubbish bin. Deleted files are dragged into the PC's desktop Recycle Bin. Suggest a better disposal procedure for Ben and explain why it is sensible.
- 7 Ben is concerned that a neighbouring business has recently been broken into and computers were stolen. What suggestions might you offer to improve physical security at his office?
- 8 Surprisingly, Ben has an internet connection, but no network. How would you convince Ben to set up a LAN?
- 9 Ben needs to replace his computers because they are old and slow, and because he is running out of hard disk space. He currently uses an online backup service, Oddle Poddle Online, which is *still* backing up his hard disk two months after he started the process. How would you convince him to go beyond cloud storage and consider using cloud computing instead of investing in new hardware and updated application software?

INDEX

- accessibility
 - data collection design principle 19, 205
 - message communication 186, 192–3
- accidental threats to data and information
 - 275, 276–7
- accuracy (data integrity) 114–21
 - challenges 115–18
 - clarity 114–15, 117, 120
 - completeness 114, 116–17, 118–20
 - consistency 114, 115, 118, 120–1
 - correctness 114, 115–16, 118
 - measures to improve 118–21
- accurate information 247–8
- acquisition and information systems 292
- ADSL 298
- age appropriateness and global message
 - 190–2
- algorithms 220
- alignment (data collection design principle)
 - 15, 206
- American Psychological Association (APA)
 - 107, 108–10
- analogue data 157
- analytic coding 90
- annotated diagrams 221
- antivirus software 173–4
- appearance (data collection design principle)
 - 15–17, 206–9
- application programming interface (API) 106
- archiving data 151
- artwork 169
- ASP 19
- ASP.Net 19
- attacks to data and information 278
- attractiveness and message communication 186,
 - 198–9
- attribute listening 214
- Australian Bureau of Statistics (ABS) 93
- Australian National Data Service 93
- Australian Privacy Principles (APPs) 131–2,
 - 258–60, 262
- authenticity (data integrity) 112–13
- authority 91
- average (statistical concept) 152, 153–4
- back-end tools 20
- balance (data collection design principle)
 - 16–17, 208–9
- bar graphs 158
- bech test 49
- biased information 248
- biometric identification 171–2
- Boolean data type 22, 23
- botnet 278
- brainstorming 209–11
- browsers 195–6
- cables, network 169
- case studies
 - Australia Card 258
 - Gantt chart for creating a database 139–43
 - Pavlov's dogs 78
 - Pizzas on the Go 27–8, 29, 35–6, 38, 39, 40,
 - 42, 45–6, 48, 57, 59
 - Ready Set Go gym 87–90
- causality (statistical concept) 152, 156–7
- chain of custody 91
- character (data type) 22, 23
- Charter of Human Rights and Responsibilities*
 - Act 2006* (Vic) 133–4
- Chromebook 298
- citation 108
- clarity
 - and data accuracy 114–15, 117, 120
 - and message communication 186, 193
- closed questions 94–5
- cloud computing 148–9, 297–8
 - advantages 297–8
 - disadvantages 298
- codes of conduct 265–6
- coding
 - analytic 90
 - descriptive 90
 - and open questions 95
 - qualitative data 86–7
 - theoretical 90
- column graphs 158
- communication and information systems 293
- communication hardware 165, 168–9
- communicating your message 186

complete information 247
 completeness
 and data accuracy 114, 116–17, 118–20
 and message communication 186, 198
 computer-use policies 265–6
 conditional formatting, database 164–5
 consistency and data accuracy 114, 115, 118, 120–1
 constraints (solution design) 80–1
 content (functionality) of data accuracy 114
 contrast (data collection design principle) 16, 207–8
 conventions for MMOS 202–4
 cookies 19, 20, 267
Copyright Act 1968 133, 269
 correctness and data accuracy 114, 115–16, 118
 correlation (statistical concept) 152, 156–7
 CSIRO's Data Access Portal 93
 culture and global message 188–9
 curating data 146

data
 accidental threats to 275, 276–7
 analogue 157
 backing up 57
 definition 73
 deliberate threats to 275, 278–9
 digital 157
 electrical protection 57–8
 encryption (storage and transmitted) 59, 173
 event-based threats to 275, 279–80
 manipulation 199–202
 and online transactions 5–9
 organising and storing 146–52
 primary 82–4
 protecting valuable 57–9, 170–5
 protection of rights 20–1
 qualitative and quantitative 85–91
 secondary 82–4
 security 170–5
 systems security software 58
 technical threats to 275, 277–8
 test 41–3
 testing the solution 49–57
 threats to 275–80
 types, characteristics of 22–3, 121–3
 user flow diagrams 11–13

 usernames and passwords 58
 visualisations 157–9, 170
 data, components of 291–6
 digital systems 294–6
 people 291
 processes 292–4
 data acquisition 94–107
 asking questions 94–7
 conducting interviews 97–8
 observing subjects 98–9
 by organisations 9–10
 querying resources 105–7
 surveys and questionnaires 99–105
 data acquisition software 10, 19–20
 data analytics 128–75
 data security 170–5
 digital systems 165–70
 file naming strategies 144–6
 identifying patterns and relationships 152–65
 legal requirements 129–34
 patterns and relationships 152–65
 project management 134–44
 solution requirements 79–80
 solution specifications 78–82
 data and information
 concept xiv, 247
 importance of 247–70
 see also information; information systems
 data collection
 design principles 15–20
 online 3–5
 protection of rights 20–1
 techniques for efficient and effective 13–15
 data collection design principles 15–20, 204–9
 accessibility 19, 205
 alignment 15, 206
 appearance 15–17, 204, 206
 balance 16–17, 208–9
 contrast 16, 207–8
 flexibility and ease of use 18–19
 functionality 18–20, 204, 205
 MMOS 204–9
 repetition 16, 206–7
 space 16, 208
 useability 18, 205
 data dictionary 36–8
 data integrity 30, 111–21
 accuracy 114–21

- authenticity 112–13
 - relevance 113–14
 - timeliness 111–12
- data patterns and relationships 152–65
 - basic statistics 153–7
 - conditional formatting 164–5
 - data visualisations 157–9
 - filtering 163
 - pivot tables 159–62
 - queries and searches 159
 - sorting 162
- data sources 92–4
 - government 92
 - government agencies 93
 - libraries 92
 - referencing 107–11
- data storage and organisation 146–52
 - archiving 151
 - cloud and file sharing 148–9
 - email 151
 - encryption 59
 - hierarchies 146–8
 - metadata 150–1
 - spreadsheets 149–50
 - synchronising files 152
- data structure diagram (DSD) 38–9
- data structures 121–3
- database, querying a 105–6
- database design tools 36–41
 - data dictionary 36–8
 - data structure diagram (DSD) 38–9
 - naming conventions 36
 - query design 39–40
 - testing the solution 49–57
- database elements 41–2
- database record 23
- database terminology 23–4
- databases 3, 23–6
 - and data storage 148
 - data types 22–3
 - relational 25–36, 44–8
- data.gov.au 92
- date (data type) 22, 23
- DDoS (distributed denial of service) 278
- de Bono's Six Hats 214–15
- deliberate threats to data and information
 - 275, 278–9
- dependencies (project management) 135
- dependent variable 75
- descriptive coding 90
- design ideas 209–20
 - attribute listening 214
 - creative techniques 209–13
 - de Bono's Six Hats 214–15
 - evaluating 218–20
 - tips for creative thinking 215–18
- design principles *see* data collection design principles
- design tools 220–3
 - input-process-output (IPO) charts 220
 - layout diagrams 223
 - mock ups or annotated diagrams 221
 - organisational and hierarchy charts 223
 - site map 221
 - storyboard 222
- digital data 157
- digital signatures 113
- digital systems 165–70
 - concept xiv, 2
 - definition 165
 - hardware 165–9, 294
 - networks 296
 - software 169–70, 294–6
- disaster recovery plans (DRP) 275, 280–9
 - recovery strategies 288–9
 - security failure 289–90
 - scope 281–7
 - testing disaster plan 287–8
- disposal and information systems 293–4
- distribution questions 104
- drive-by downloads 277
- dynamic webpage 19
- ease of use (data collection design principle)
 - 18–19
- economic constraints 81
- educating a global audience 186–93
- effectiveness 3
- efficiency 3
- email (data storage) 151
- employee monitoring 266–7
- encryption (storage and transmitted data) 59, 173
- entity–relationship diagram (ERD) 26, 28–9
- ethical dilemmas 270
- ethics
 - definition 265

- and information systems 265–70
- and the internet 267–9
- evaluation and problem-solving methodology 237–9
- evaluation criteria 237
- evaluation methods 238
- event-based threats to data and information 275, 279–80
- field (databases) 23
- field notes 99
- file backups 175
- file management (MMOS) 227–8
 - file system basics 227–8
 - planning disk structure 228
 - terms 228–9
 - utilities 229–30
 - version control 229
- file naming conventions (data analytics) 145–6
- file naming formats (data analytics) 144
- file naming strategies (data analytics) 144–6
- file sharing 148–9
- file storage 151
- files, synchronising 152
- filtering, database 163
- firewalls 173
- first normal form (1NF) 30, 31
- flat field (databases) 23, 24
- flexibility (data collection design principle) 18–19
- floating point (numeric data type) 22
- form (appearance) of data accuracy 114
- formats for MMOS 202–4
- formulas and summary statistics, testing 54–5
- frequency distribution tables 80
- functionality (data collection design principle) 18–20, 205
- fuzzy logic 121
- fuzzy searches 293
- Gantt charts 134, 137, 138–44, 169, 223–5
- gender inclusion and global message 187–8
- global economy 4
- government agencies 93
- government data sources 92
- GovPond 92
- graphic organisers 212–13
- graphs 80, 158–9

- Hansard 92
- hardware 165–9, 196, 294
- hashing 59
- Health Records Act 2001* 252, 256, 263–4
- hidden agenda in questioning 96
- hierarchies (data storage) 146–8
- hierarchy charts 223
- HTTPS 20
- human rights 133–4
- hypothesis 73–8
 - characteristics of a reasonable 75–6
 - formulating your 76–7
 - definition 73–5
 - proof, support and refutation 77
- images, editing 169
- independent variable 75
- information
 - accidental threats to 275, 276–7
 - accurate 247–8
 - biased 248
 - characteristics 247–9
 - clarity of message 249
 - complete 247
 - concept xiv, 247
 - deliberate threats to 275, 278–9
 - event-based threats to 275, 279–80
 - importance of 247–70
 - legal and ethical tensions 269–70
 - legal obligations 255–6
 - maximising opportunities 252–4
 - minimising risk 256
 - privacy policies 256, 257
 - storage and disposal, key legislation 256–65
 - technical threats to 275, 277–8
 - threats to 275–80
 - timely 247
 - unbiased 248
- information management strategies 252–6
 - effectiveness 299–300
 - efficiency criteria 300
 - evaluation 299–300
 - legal obligations 255–6
 - maximising opportunities 252–4
 - minimising risk 255
 - privacy policies 256

- information systems
 - codes of conduct 265–6
 - digital systems 294–6
 - disaster recovery plans (DRP) 275, 280–1
 - employee monitoring 266–7
 - ethics 265–70
 - ethics and the internet 267–9
 - legal and ethical tensions 269–70
 - organisational goals and objectives 249–52
 - and people 291
 - and processes 292–4
 - security failure 289–90
 - workplace responsibilities 265
- input and information systems 292
- input hardware 165–7
- input masks (test data) 42
- input–process–output (IPO) charts 36, 220
- integer (numeric data type) 22
- intellectual property legislation 133
- interactions and impact (concept) xiv, 2
- internet and ethics 267–9
- Internet Message Access Protocol (IMAP) 297
- interpolation 119
- interview technique 97–8
- interviews (data acquisition) 97–8
- jargon 193
- key concepts xiv, 2
- keyloggers 278
- language commonality and global message 190
- layout diagrams 11, 223
- leading questions 96
- legacy assessment 239
- legal constraints 81
- legal obligations (information management) 255–6
- legal requirements and data analytics 129–34
 - human rights 133–4
 - intellectual property legislation 133
 - legislation 129–33
 - privacy legislation 131–2
 - Spam Act 2003* 132–3
- legislation
 - and data analytics 129–33
 - data and information storage and disposal 256–64
 - information management 252
- libraries 92
- line graphs 159
- loaded questions 96
- logging out 172
- machine validation (online data acquisition) 9
- malware 58, 173, 278
- mandatory conventions 203
- manmade disasters 279
- manual validation (online data acquisition) 9
- many-to-many relationship (relational databases) 26
- mean (statistics) 153
- media and message communication 197
- median (statistics) 153
- Melbourne Data 92
- message clarity (information) 249
- message communication
 - accessibility 186, 192–3
 - accuracy 186, 195
 - attractiveness 186, 198–9
 - clarity 186, 193
 - communicating your message 186
 - completeness 186, 198
 - educating a global audience 186–92
 - effective 186–91
 - readability 186, 193–4
 - relevance 186, 195
 - useability 186, 195–7
- metadata 150–1
- milestones (project management) 135
- mindmapping 80, 211–12
- mission statement 249–50
- mock ups 221
- mode (statistics) 153
- modems 169
- multimodal authoring 231
- multimodal data 231
- multimodal online solution (MMOS) 185–99
 - accessibility 186, 192–3
 - accuracy 186, 195
 - attractiveness 186, 198–9
 - clarity 186, 193
 - communicating your message 186
 - completeness 186, 198
 - data manipulation software 231–2
 - definition 185
 - design ideas 209–20

- design principles 204–9
- documenting the progress of projects 223–47
- educating a global audience 186–92
 - an effective 186–91
 - formats and conventions 202–4
 - managing files 227–30
 - manipulating data 199–202
 - project plan assessment 239–40
 - readability 186, 193–4
 - relevance 186, 195
 - solution evaluation 237–9
 - testing 232–6
 - timeliness 186, 197–8
 - useability 186, 195–7
- multimodal resources 92
- naming conventions (database design tool) 36
- networks 296
 - advantages of 296
 - and communication 168–9
 - disadvantages of 296–7
- nodes 296
- normalisation (RDBMS structure) 26
- numeric fields (data types) 22, 23
- observation
 - approaches 99
 - disadvantages of 99
- observing subjects (data acquisition) 98–9
- one-to-many relationship (relational databases) 25
- one-to-one relationship (relational databases) 25
- ongoing services and organisations 4
- online data
 - acquisition 10
 - acquisition techniques used by organisations 9–10
 - collecting 3–5
 - user flow diagrams 11–13
 - and users 5–9
 - see also* data
- online form 3
- online resources 92
- open questions 94, 95–6
- optional conventions 203
- organisational charts 223
- organisational goals 249–51
- organisations
 - goals and objectives (information systems) 249–52
 - and online data 3–5
- outcomes viii–xi
- outlier (statistics) 153
- output and information systems 293
- output hardware 165–7
- paraphrasing 107
- passwords (data protection) 58, 171
- people and information systems 291
- PHP 19
- physical security 170–1
- physical security controls 275
- pie charts 158
- primary data 82–4
- primary key (databases) 23
- privacy legislation and data analytics 131–2
- privacy policies (data protection) 20–1
- pivot tables 159–62
- plug-ins 195
- POP 297
- ports 168
- preferred conventions 203
- printers 166–7
- Privacy Act 1988* 252, 256, 257–62, 269
- Privacy and Data Protection Act 2014* 252, 256, 262–3
- privacy policies (information management) 256, 257
- problem solving
 - approaches to xiv, 2
 - testing the solution (data management) 49–57
- problem-solving methodology xii–xiii
 - design stage xii–xiii
 - and evaluation 237–9
 - problem analysis xii
 - solution development xiii
 - solution evaluation xiii
 - stages xii
- processes and information systems 292–4
- processing and information systems 292–3
- project management 134–44
 - concepts 135
 - definition 134
 - processes 135–44
- project plan assessment 239–40
- projects, documenting the progress of 223–47

- protection of rights (data) 20–1
 - privacy policies 20–1
 - security protocols 20
- qualitative data 85–91
 - coding 86–7
- quantitative data 85–91
- queries (database tables) 24–5, 47–8, 159
- query by example (QBE) 105–6
- querying a database 105–6, 159
- querying other resources 106–7
- querying resources (data acquisition) 105–7
- query design (database design tool) 39–40
- query selection criteria, testing 52–3
- questionnaires 99–105
 - advantages/disadvantages 100
 - writing 101–4
- questions (data acquisition)
 - closed 94–5
 - distribution 104
 - faults in questions 96–7
 - hidden agenda in 96
 - leading 96
 - loaded 96
 - open 94, 95–6
- quotations in referencing 111
- ransomware 278
- readability and message communication 186, 193–4
- reference list 111
- referencing data sources 107–11
- relational database management systems (RDBMS)
 - creating a RDBMS structure 25–36
 - definition 25
 - developing 44–8
 - entity 26
 - entity–relationship diagram (ERD) 26, 28–9
 - normalisation 26, 30–6
 - setting up queries 47–8
 - table normalisation 30–6
 - tables and relationships 44–5
- relational databases 25–6
- relevance (data integrity) 113–14
- repetition (data collection design principle) 16, 206–7
- resources, querying (data acquisition) 105–7
- retrieval and information systems 293
- risk assessment analysis 281–2
- routers 169
- rubric 90–1
- scatter graphs 159
- searches, database 159
- second normal form (2NF) 30, 31–3
- secondary data 82–4
- secure sockets layer (SSL) 20
- security
 - data 170–5
 - failure, consequences of 289–90
 - physical 170–1
 - software 171–5
- security protocols (data protection) 20
- sequencing (project management) 137
- school-assessed coursework (SAC) 275
- shipping and returns policy 20, 21
- significance (statistical concept) 152, 154–5
- site map 221
- software 169–70
 - antivirus 173–4
 - applications 295
 - digital systems 294–5
 - security 58, 171–5
 - security controls 275
 - utilities 295
- solution requirements (data analytics) 79–80
 - data requirements 79–80
 - functional requirements 79
 - non-functional requirements 79
- solution specifications (data analytics) 78–82
 - constraints 80–1
 - scope 81–2
- sorting, database 162
- space (data collection design principle) 16, 208
- Spam Act 2003* 132–3
- spreadsheet charts 158–9
- spreadsheets 122, 149–50
- spyware 20
- standard deviation (statistical concept) 152, 154
- statistical concepts 152
- statistics, basic 153–7
- storage hardware 165, 167–8
- storyboard 222

- structured query language (SQL) 105, 106
- subjective assessment 98
- survey tools 170
- surveys 99–105
- switches 169
- synchronising files 152
- system goals (organisational information) 251
- systems security software 58

- table normalisation (entity–relationship diagram (ERD) 30–6
- table validation, testing 49–52
- task completeness 18
- task identification (project management) 135–6
- technical constraints 81
- technical threats to data and information 275, 277–8
- test data 41–2, 233
- test plan 41–2
- testing
 - formulas and summary statistics 54–5
 - MMOS 232–6
 - procedure (data management) 49–57
 - query selection criteria 52–3
 - sorting 55–6
 - the solution (data management) 49–57
 - types 233
- testing table 235–6
 - validation 49–52
- text data type 22, 23
- theoretical coding 90
- third normal form (3NF) 30, 34–5
- time allocation resources (project management) 137–8

- timeliness
 - data integrity 111–12
 - and message communication 186, 197–8
- timely information 247
- transport layer security (TLS) 20
- type face and message communication 194

- unbiased information 248
- useability
 - data collection design principle 18, 205
 - message communication 186, 195–7
- user acceptance testing 49
- user flow diagrams (UFD) 11–13
- usernames (data protection) 58
- users and online data transactions 5–9
 - convenience 5–8
 - reducing cost 8–9
 - variety of choice 8

- validation 4
 - and information systems 292
 - techniques 231–2
 - testing table validation 49–52
- validation rules (test data) 42
- validation techniques (test data) 42–3
- vested interest 116
- visualisations, data 157–9, 170

- white space and message communication 194
- wireless access points 169
- work breakdown structure (WBS) 135
- workstations 297

- XML files 105

ISBN: 978-0170364829

