

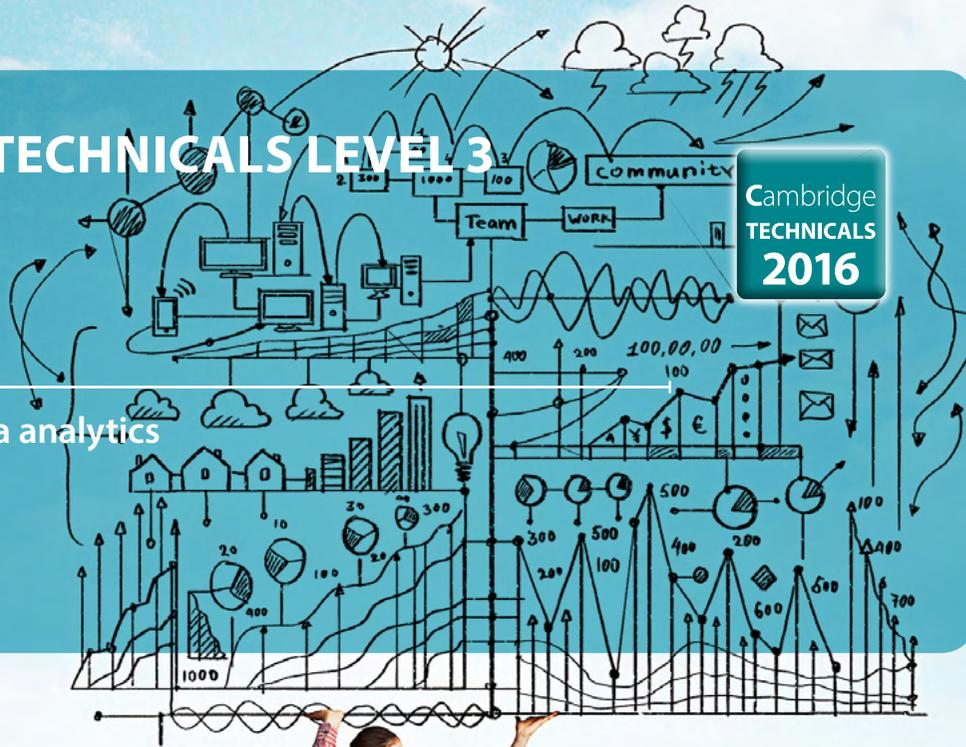
Cambridge TECHNICALS LEVEL 3

Cambridge
TECHNICALS
2016

IT

Unit 22 – Big data analytics DELIVERY GUIDE

Version 2



CONTENTS

Introduction	3
Related Activities	4
Key Terms	6
Misconceptions	7
Suggested Activities:	
Learning Outcome (LO1)	8
Learning Outcome (LO2)	11
Learning Outcome (LO3)	12

INTRODUCTION

This Delivery Guide has been developed to provide practitioners with a variety of creative and practical ideas to support the delivery of this qualification. The Guide is a collection of lesson ideas with associated activities, which you may find helpful as you plan your lessons.

OCR has collaborated with current practitioners to ensure that the ideas put forward in this Delivery Guide are practical, realistic and dynamic. The Guide is structured by learning outcome so you can see how each activity helps you cover the requirements of this unit.

We appreciate that practitioners are knowledgeable in relation to what works for them and their learners. Therefore, the resources we have produced should not restrict or impact on practitioners' creativity to deliver excellent learning opportunities.

Whether you are an experienced practitioner or new to the sector, we hope you find something in this guide which will help you to deliver excellent learning opportunities.

If you have any feedback on this Delivery Guide or suggestions for other resources you would like OCR to develop, please email resources.feedback@ocr.org.uk.

OPPORTUNITIES FOR ENGLISH AND MATHS SKILLS DEVELOPMENT AND WORK EXPERIENCE

We believe that being able to make good progress in English and maths is essential to learners in both of these contexts and on a range of learning programmes. To help you enable your learners to progress in these subjects, we have signposted opportunities for English and maths skills practice within this resource. We have also identified any potential work experience opportunities within the activities. These suggestions are for guidance only. They are not designed to replace your own subject knowledge and expertise in deciding what is most appropriate for your learners.



English



Maths



Work

Please note

The timings for the suggested activities in this Delivery Guide **DO NOT** relate to the Guided Learning Hours (GLHs) for each unit.

Assessment guidance can be found within the Unit document available from www.ocr.org.uk.

The latest version of this Delivery Guide can be downloaded from the OCR website.

UNIT AIM

Data is all around us and the amount of data being gathered is growing. Big Data deals with extremely large data sets that can be analysed computationally to reveal patterns, trends and associations. This is especially the case in relation to human behaviour and interactions. In this unit you will learn what Big Data is, how it can be gathered, analysed and used by businesses. You will also look at how a company could use big data for planning for the future.

Due to the increased use of Big Data this is an optional unit in the Emerging Digital Technology Practitioner, Application Developer and Data Analyst specialist pathways.

Unit 22 Big data analytics

LO1	Understand the scope of Big Data
LO2	Be able to process Big Data for business purposes
LO3	Be able to provide information resulting from processing Big Data

To find out more about this qualification please go to: <http://www.ocr.org.uk/qualifications/cambridge-technicals-it-level-3-certificate-extended-certificate-introductory-diploma-foundation-diploma-diploma-05838-05842-2016-suite>

The activities within this teaching and learning resource must not be used for summative assessment purposes. As part of our teaching we expect support to be given to your learners; such support is not permissible for summative assessment and is likely to be considered malpractice.

Cambridge
TECHNICALS
2016

2016 Suite

- New suite for first teaching September 2016
- Externally assessed content
- Eligible for Key Stage 5 performance points from 2018
- Designed to meet the DfE technical guidance

RELATED ACTIVITIES

The Suggested Activities in this Delivery Guide listed below have also been related to other Cambridge Technicals in IT units/Learning Outcomes (LOs). This could help with delivery planning and enable learners to cover multiple parts of units.

This unit (Unit 22)	Title of suggested activity	Other units/LOs	
LO1	What is Big Data?	Unit 17 Internet of everything	LO1 Understand what is meant by the Internet of Everything (IoE)
LO1	Use of Big Data – customer preferences	Unit 17 Internet of everything	LO1 Understand what is meant by the Internet of Everything (IoE)
LO1	Use of Big Data – infrastructure challenges	Unit 1 Fundamentals of IT	LO1 Understanding computer hardware LO3 Understanding business IT systems
		Unit 2 Global information	LO1 Understand where information is held globally and how it is transmitted
		Unit 17 Internet of everything	LO1 Understand what is meant by the Internet of Everything (IoE)
		Unit 18 Computer systems hardware	LO1 Understand the components of a computer system
LO1	Use of Big Data – infrastructure solutions	Unit 1 Fundamentals of IT	LO1 Understanding computer hardware LO3 Understanding business IT systems
		Unit 2 Global information	LO1 Understand where information is held globally and how it is transmitted
		Unit 17 Internet of everything	LO1 Understand what is meant by the Internet of Everything (IoE)
		Unit 18 Computer systems hardware	LO1 Understand the components of a computer system
LO2	Big Data sources	Unit 13 Social media and digital marketing	LO2 Understand the use of social media in business
LO2	Big Data risks	Unit 1 Fundamentals of IT	LO5 Understand ethical and operational issues and threats to computer systems
		Unit 2 Global information	LO4 Understand the legal and regulatory framework governing the storage and use of global information LO6 Understand the principles of information security
		Unit 4 Computer networks	LO1 Understand what is meant by cyber security LO2 Understand the issues surrounding cyber security LO3 Understand measures used to protect against cyber security incidents LO4 Understand how to manage cyber security incidents
LO2	Preparing Big Data for analysis	Unit 7 Data analysis and design	LO1 Understand the purpose and stages of data analysis and design
LO2	Processing Big Data – preparation, import and manipulation	Unit 7 Data analysis and design	LO1 Understand the purpose and stages of data analysis and design LO3 Be able to develop data design solutions to meet business requirements
		Unit 10 Business computing	LO3 Be able to use tools to edit and analyse data
LO2	Processing Big Data – extraction	Unit 22 Big data analytics	LO1 Understand the scope and challenges of Big Data
LO2	Evaluating results	Unit 10 Business computing	LO3 Be able to use tools to edit and analyse data
		Unit 13 Social media and digital marketing	LO2 Understand the use of social media in business
LO3	Presenting results presentation methods – video Presenting results presentation methods – presentation Presenting results presentation methods – formal reports Presenting results presentation methods – graphical diagrammatic	Unit 7 Data analysis and design	LO4 Be able to present data analysis and design solutions to stakeholders
		Unit 10 Business computing	LO4 Be able to present data analysis outcomes

This unit (Unit 22)	Title of suggested activity	Other units/LOs	
LO3	Target audience consideration	Unit 7 Data analysis and design	LO4 Be able to present data analysis and design solutions to stakeholders
		Unit 10 Business computing	LO4 Be able to present data analysis outcomes
LO3	Big Data recommendations	Unit 7 Data analysis and design	LO4 Be able to present data analysis and design solutions to stakeholders
		Unit 10 Business computing	LO4 Be able to present data analysis outcomes

KEY TERMS

Explanations of the key terms used within this unit, in the context of this unit

Key term	Explanation
Big Data velocity	The pace at which data flows in from sources like business processes, machines, networks and human interaction with things like social media sites, mobile devices etc.
Big Data veracity	The biases, noise and abnormality in data. Is the data that is being stored and mined meaningful to the problem being analysed?
Clickstream	This is a series of mouse clicks made by a user while accessing the Internet. It is monitored so that a person's interests can be ascertained.
Competitive advantage	Using data to enable a company to outperform its rivals.
Correlation	How much one thing is related to another; for example when the temperature increases then the sales of ice cream go up.
Data cleansing	Amending or removing data in a data set that is not correct, duplicated and/or incomplete.
GIGO	GIGO (garbage in, garbage out). The quality of output is determined by the quality of the input so if the data being analysed is incorrect then the output will also be flawed.
Hadoop cluster	This relates to a computational cluster designed specifically for storing and analysing huge amounts of unstructured data in a distributed computing environment.
Lead time	Lead time is, for example, the time between the placement of an order and the delivery of a new car from a manufacturer. This lead time may vary depending on many factors e.g. if the car is imported from another country, there may be specific customisations a client may want on their car etc.
Market trends	A market moving in a certain way over time.
Scalability	Scalability is the possibility of a network handling a growing amount of data and its potential to be enlarged in order to accommodate that growth.
Semi-structured data	This is mainly made up from a mark-up language (e.g. XML) i.e. a language for exchanging data on the web. These languages contain user-defined data tags which make them machine-readable.
Sociodemographics	Sociodemographics are made up of different variables; these include for example gender, age, level of education, employment status, marital status, total number of persons living in the house.
Socioeconomics	Socioeconomics analyses how the local, regional economy, or the global economy influences society.
Structured data	For example, tabular data found in spreadsheets and relational databases.
Unstructured data	Unstructured data is formed from the likes of text, images, audio, and video.
Virtualisation	Virtualisation is creating a virtual (rather than actual) version of something, including virtual computer hardware platforms, operating systems, storage, and computer networks.

MISCONCEPTIONS

Some common misconceptions and guidance on how they could be overcome

What is the misconception?	How can this be overcome?	Resources which could help
That all data is valuable	<p>In order to make data valuable, it must be sorted, processed and distributed.</p> <p>Having a hypothesis before starting to draw down data sets allows you to guide your data analysis, looking at relevant data and in a relevant time frame. For example, it would be fairly pointless to look at the sales of ice cream as compared to air temperatures in the UK winter months.</p>	<p>Ensure that the hypothesis is well defined before starting to look at data sets and meet with the client to ensure that the correct parameters are set out.</p> <p>See page 14 of: https://www.henley.ac.uk/html/hwss/files/Henley-Centre-for-HR-Excellence-Big-Data-Research-paper.pdf</p>
Big Data is often better than little data	<p>The use of little data can sometimes be more effective (little data is simply a smaller data set). It is often a case of quantity versus quality. Even though Big Data has a huge quantity of information, the quality of that information may not always be what we are looking for, and it often needs to be sorted and validated before processing. Sometimes with little data, the information is more controlled and clean thus making it more useful.</p>	<p>Ensuring that any data that is provided by the client is sufficient to be able to analyse trends and to extract relevant information. However, this data ideally must be big enough to allow Big Data analytics to be demonstrated.</p> <p>http://www.dataversity.net/how-content-intelligence-makes-big-data-valuable/ http://www.datamation.com/applications/big-data-9-steps-to-extract-insight-from-unstructured-data.html http://www.information-management.com/news/news/five-approaches-to-extract-value-from-big-data-10024889-1.html</p>

SUGGESTED ACTIVITIES

LO No:	1		
LO Title:	Understand the scope of Big Data		
Title of suggested activity	Suggested activities	Suggested timings	Also related to
What is Big Data?	<p>Tutors could explain briefly what the Big Four are and then divide the group into four, giving each group one of the Big Four to research.</p> <p>Tutors could ask learners to look at the following which covers many of the areas below: http://ac.els-cdn.com/S0268401214001066/1-s2.0-S0268401214001066-main.pdf?_tid=5eb33152-c833-11e5-8f42-0000aacb361&acdnat=1454256046_1bd4d10133c08d6070c6f385bb807969</p> <p>Learners may put together a diagrammatic representation of the area they have been given starting with mind mapping the different areas that are associated with their topic. This is beneficial as a diagrammatic representation is often the way that the Big Four are portrayed and it will also allow the learners to think about the idea that often diagrams can convey more meaning than a lot of text. Learners could then design a poster that can be displayed to the remainder of the group (tutors should check the validity of the content before the rest of the group see it). http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html https://hrboss.com/articles/big-data-infographic</p>	1 hour	Unit 17 LO1
Use of Big Data – trends	<p>Tutors could give the group a small data set that shows patterns and market trends. If there is a possibility of linking with a local industry it may be possible to make this more relevant.</p> <p>If a local shop or larger company could make some data available it would be possible to look at trends (some patterns may be hidden depending on the data set given). Data may also be available from their centre (though it would need to be anonymised due to data protection) – it may be possible to look at sociodemographics and examination result outcomes, trends in travel to the centre. For example, a local clothes shop may show trends in winter and summer clothes. It may be possible for learners to think about the impact of this and the planning that needs to go into forecasting these trends and the impact it may have. This may mean looking into the ordering process (especially if there is a long lead time and the goods are imported), the space needed to store goods until they are needed and the cost to store large quantities of clothing. This can apply to many different industries e.g. car manufacturers before a new number plate comes out. Learners may also look at Google Analytics that allows companies to measure their advertising return on investment (ROI) as well allowing for the tracking and reporting of website traffic. http://www.ibmbigdatahub.com/blog/big-data-trends-top-eight-analytics-lessons-business</p> <p>Learners could look at how Cancer Research UK is using Big Data to personalise cancer treatment: http://www.computerweekly.com/feature/Genomics-England-exploits-big-data-analytics-to-personalise-cancer-treatment</p>	30 minutes	



<p>Use of Big Data – customer preferences</p> 	<p>Tutors could get a range of store cards and give these out to the group and who then think about why store cards are used. Learners could consider what the advantages are for the company and the holder. Learners would need to consider the information that is being supplied by the holder e.g. sociodemographics, travel patterns to store, shopping trends by socioeconomic groups, reaction to special offers and advertising.</p> <p>The group could research the advantages and disadvantages using sites such as: http://www.theguardian.com/news/datablog/2013/oct/31/are-loyalty-cards-really-worth-it http://www.select-statistics.co.uk/article/blog-post/how-do-supermarkets-use-your-data</p> <p>Learners could think about the data stored and how much of it there is on a daily basis (this may be real time data or batch data stored either on a network of servers or at a central hub); e.g. is real time data being collected or is batch data being held, from processing transactions at a supermarket, to be sent to head office at regular intervals. They could present their findings as a report to the group.</p>	30 minutes	Unit 17 LO1
<p>Use of Big Data – impact on organisations</p> 	<p>Linking in with the activity above, tutors could introduce how Big Data is used to gain competitive advantage by businesses that use it to target certain groups with offers at certain times of the year.</p> <p>The group could research the local area and supermarkets that have set up in the area and think about why they have been opened in those locations compared to others e.g. is this because of the road network infrastructure that makes it possible for deliveries to be made and customers to park and shop more easily, does the socioeconomic makeup of that area make it viable (for example, are ‘higher end’ supermarkets placed in one area while more budget supermarkets are being opened in a different area), are the supermarkets in out-of-town retail parks or in town centre shopping areas? http://iveybusinessjournal.com/publication/why-big-data-is-the-new-competitive-advantage/</p> <p>Learners could also research how data is being communicated between public sector organisations, for example Government Connect. The following website provides a starting point for this: http://searchsecurity.techtarget.co.uk/definition/UK-Government-Connect-Secure-Extranet-GCSX</p>	1 hour	
<p>Use of Big Data – infrastructure challenges</p>	<p>Tutors could introduce why Big Data is now possible in order to give some context to why this is now relevant to organisations, possibly due to storage capacities. Tutors could introduce Hadoop clusters at this point and allow learners to research what they are.</p> <p>Learners could look at old storage capacities for the likes of floppy disks and compare this to innovations like cloud storage and capacity changes. It may be relevant to look at the computing power that was used to send the first person to the moon compared to the processing power that is available today. http://www.enterprisenetworkingplanet.com/datacenter/network-infrastructure-concerns-for-big-data.html http://www.computerweekly.com/news/2240186002/Inadequate-datacentre-infrastructure-is-a-barrier-to-big-data-analytics http://www.bluedata.com/blog/2015/05/big-data-case-study-hadoop-infrastructure/</p>	1 hour	Unit 1 LO1, LO3 Unit 2 LO1 Unit 17 LO1 Unit 18 LO1

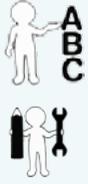
Use of Big Data – infrastructure solutions	<p>Linking in with the previous activity, learners could look at how infrastructure such as cloud, Hadoop clusters, virtualisation, and solid state has changed (in terms of capacity and cost) and how this has made the storage and analysis of Big Data possible.</p> <p>Tutors could divide the group into six and give each group one of the areas to research in the Teaching Content under 1.4 Infrastructure challenges posed by Big Data. Learners could then present their findings to the rest of the group.</p>	1 hour	Unit 1 LO1, LO3 Unit 2 LO1 Unit 17 LO1 Unit 18 LO1
---	---	--------	---

SUGGESTED ACTIVITIES

LO No:	2		
LO Title:	Be able to process Big Data for business purposes		
Title of suggested activity	Suggested activities	Suggested timings	Also related to
Big Data sources 	<p>Tutors could divide the class into five groups and let each group research one of the Big Data sources outlined in Teaching Content 2.1. It may be useful if there is a contact in industry who could allow the group to visit their premises to show the learners how Big Data is collected from sources (especially those that are web-based). N.B. This could be used in the tutor/centre plan for Meaningful Employer Involvement.</p> <p>https://datafloq.com/read/understanding-sources-big-data-infographic/338</p>	1 hour	Unit 13 LO2
Big Data risks	<p>Learners could research the risks of Big Data; this may link into other qualifications, particularly any relating to database work in the areas of data cleansing and data validity. This will also follow on from any work that has been carried out on legislation and ethics.</p> <p>Tutors could get the learners to enhance and apply any work that has been carried out in earlier years on this subject, maybe at GCSE, or in other qualifications or units.</p> <p>http://data-informed.com/the-5-biggest-risks-of-big-data/</p>	1 hour	Unit 1 LO5 Unit 2 LO4, LO6 Unit 4 LO1, LO2, LO3, LO4
Preparing Big Data for analysis 	<p>Tutors could at this point give learners access to a small data set that includes some invalid data, some data that is wrongly input, and allow learners to start to look at the data reflecting on the areas in Teaching Content 2.3; this will allow learners to gain experience with a smaller data set before starting work on the Big Data main task. Learners could look at data that is captured by wearable tech e.g. fitness data or weather station data and then its potential use.</p> <p>http://bigdata-madesimple.com/26-popular-techniques-for-analysing-big-data/ http://documents.software.dell.com/statistics/textbook/data-mining-techniques</p>	3 hours	Unit 7 LO1
Processing Big Data – preparation, import and manipulation 	<p>Depending on the method that learners will be using to analyse the data, the data that has been used in the activity above and is prepared ready for analysis could be imported into the software that will be used to analyse it and valid questions asked of the data that link to Teaching Content 1.2 Use of Big Data and 1.3 Impact on organisations.</p> <p>http://www.developer.com/db/understanding-big-data-processing-and-analytics.html</p>	3 hours	Unit 7 LO1, LO3 Unit 10 LO3
Processing Big Data – extraction 	<p>Linking to the previous activity, once the data has been manipulated the learners could look at the best way to extract any areas they have found that link to the Teaching Content 1.2 Use of Big Data and then how best to display this – tutors could compare and contrast the use of looking at data compared to tables, or compared to graphs and other graphics that could be used.</p>	4 hours	Unit 22 LO1
Evaluating results 	<p>Tutors could ask learners to show the results of their data to the group and then ask the group to evaluate the results that have been found – this may be interesting if the same data set has been used for all learners. Different trends and patterns and correlations may be found.</p> <p>http://betterevaluation.org/blog/big_data_in_evaluation</p>	2 hours	Unit 10 LO3 Unit 13 LO2

SUGGESTED ACTIVITIES

LO No:	3		
LO Title:	Be able to provide information resulting from processing Big Data		
Title of suggested activity	Suggested activities	Suggested timings	Also related to
Presenting results presentation methods – video	<p>Tutors could lead a session on how to compose a formal video presentation that will be shown to a client to illustrate results. These hints and tips on producing business videos may help: http://www.videouniversity.com/articles/ten-tips-for-producing-highly-effective-business-videos/</p> <p>Learners could then use their own small data set to carry this out.</p>	3 hours	Unit 7 LO4 Unit 10 LO4
Presenting results presentation methods – presentation	 <p>Linking with the activity above, undertake a tutor-led demonstration of how to compose a formal business presentation including speaker notes. Again, the learners may wish to research what makes an effective business presentation using sites such as: http://www.thinkoutsidetheslide.com/five-tips-to-make-powerpoint-business-presentations-more-effective/</p> <p>Learners could then use their own small data set to carry this out.</p>	3 hours	Unit 7 LO4 Unit 10 LO4
Presenting results presentation methods – formal report	 <p>Linking with the previous two activities, the tutor could show learners how to produce a formal business report and, if there is an industrial link with a local company, it could ask the learners to provide an example of a formal report, indicating the formats and house styles that are used. Learners may have conversations with their client so that their reports comply with the company's house style for reports.</p> <p>Learners could then use their own small data set to carry this out.</p>	3 hours	Unit 7 LO4 Unit 10 LO4
Presenting results presentation methods – graphical diagrammatic	 <p>Learners could be shown how best to present data, linking into Teaching Content 2.4 Processing Big Data and how this is produced. Tutors could demonstrate the use of software to create these graphical representations. Learners could research on the Internet how best to present data in graphical format, for example: http://faculty.atu.edu/mfinan/2043/section30.pdf</p> <p>Learners could then use their own small data set to carry this out.</p>	3 hours	Unit 7 LO4 Unit 10 LO4
Target audience consideration	<p>Tutors could divide learners into groups and ask them to prepare their presentation for different audiences as outlined in Teaching Content 3.1 Presenting Results target – audience considerations. Learners can combine the activities from previous Learning Outcomes in order to form a presentation to the given audience. Guidance could be given by the tutor on how best to achieve this and some examples could reinforce how businesses present to different audiences. http://www.marketingdonut.co.uk/marketing/customer-care/understanding-your-customers/understand-your-customers-better-with-big-data</p>	2 hours	Unit 7 LO4 Unit 10 LO4

Big Data recommendations

Tutors could try to link up with a local company to look at how it makes recommendations to its team or board members and if possible either visit the head or regional office or invite a guest speaker. N.B. This could be used in the tutor/centre plan for Meaningful Employer Involvement. It may be possible to gather some reports online or through local libraries or council offices.

1 hour

Unit 7 LO4
Unit 10 LO4



We'd like to know your view on the resources we produce. By clicking on the 'Like' or 'Dislike' button you can help us to ensure that our resources work for you. When the email template pops up please add additional comments if you wish and then just click 'Send'. Thank you.

Whether you already offer OCR qualifications, are new to OCR, or are considering switching from your current provider/awarding organisation, you can request more information by completing the Expression of Interest form which can be found here: www.ocr.org.uk/expression-of-interest

OCR Resources: the small print

OCR's resources are provided to support the delivery of OCR qualifications, but in no way constitute an endorsed teaching method that is required by OCR. Whilst every effort is made to ensure the accuracy of the content, OCR cannot be held responsible for any errors or omissions within these resources. We update our resources on a regular basis, so please check the OCR website to ensure you have the most up to date version.

This resource may be freely copied and distributed, as long as the OCR logo and this small print remain intact and OCR is acknowledged as the originator of this work.

OCR acknowledges the use of the following content:
Cover image: Sergey Nivens/Shutterstock.com
Square down and Square up: alexwhite/Shutterstock.com

Please get in touch if you want to discuss the accessibility of resources we offer to support delivery of our qualifications: resources.feedback@ocr.org.uk

Looking for a resource?

There is now a quick and easy search tool to help find **free** resources for your qualification:

www.ocr.org.uk/i-want-to/find-resources/

ocr.org.uk/it

OCR Customer Contact Centre

General qualifications

Telephone 01223 553998

Facsimile 01223 552627

Email general.qualifications@ocr.org.uk

OCR is part of Cambridge Assessment, a department of the University of Cambridge. *For staff training purposes and as part of our quality assurance programme your call may be recorded or monitored.*

© **OCR 2018** Oxford Cambridge and RSA Examinations is a Company Limited by Guarantee. Registered in England. Registered office 1 Hills Road, Cambridge CB1 2EU. Registered company number 3484466. OCR is an exempt charity.



Cambridge
Assessment

