Data protection

How to disclose information safely

Removing personal data from information requests and datasets

ico.

Information Commissioner's Office

Contents

- The General Data Protection Regulation (GDPR) came into effect on 25 May 2018. The Data Protection Act 1998 will be replaced in the UK with the Data Protection Act 2018.
- Our approach to considering the disclosure of personal data under the Freedom of Information Act 2000 (FOIA) and the Environmental Information Regulations 2004 (EIR) remains largely the same and our existing guidance is still of use. We will amend it in due course. However, there are a few key points to consider.
- The definition of personal data and sensitive personal data have changed, as have the data protection principles and the rights of subject access. Please see our Guide to the General Data Protection Regulation for more detailed information.
- If the information constitutes the personal data of third parties, public authorities should consider whether disclosure would breach the data protection principles. (In the case of special category or criminal offence data, public authorities must also satisfy one of the conditions listed in Article 9 of the GDPR). Principle (a) under Article 5 is the most applicable.
- When considering whether disclosure of information is a breach of principle (a), a public authority should first consider whether disclosure is lawful and then whether it is fair. The lawful basis that is most likely to be relevant is legitimate interests under Article 6.1(f).
- The Data Protection Act 2018 amends FOIA and the EIR so that the legitimate interests lawful basis is applicable to public authorities when they are considering disclosure.
- Competent authorities for the purposes of the law enforcement provisions (law enforcement bodies) should consider the application of principle (a) of the GDPR for disclosures under FOIA and the EIR.

Introduction

1. The Data Protection Act 1998 (the DPA) is based around eight principles of good information handling. These give people

specific rights in relation to their personal information and place certain obligations on those organisations that are responsible for processing it.

2. An overview of the main provisions of the DPA can be found in The Guide to Data Protection.

3. This is part of a series of guidance, which goes into more detail than the Guide, to help organisations to fully understand their obligations and promote good practice.

4. This guidance explains some of the most common types of inappropriate disclosures that the ICO has seen in recent years but also includes other types that data controllers and public authorities should be aware of.

5. This guidance is for organisations disclosing information which has been derived from personal data and requires further processing to ensure that individuals cannot be identified from that information. This includes organisations releasing data as part of a subject access request. It is also relevant to public authorities responding to a freedom of information or environmental information requests or proactively publishing data as part of a publication scheme or otherwise making data available.

Overview

- There are a number of different ways that personal data can be stored in a file and disclosed by mistake.
- Exporting data to simple file formats such as CSV (Comma Separated Value) and inspecting the file can help highlight many potential unauthorised disclosures.
- Mistakes can also be made in the redaction process and other file types such as email or images can contain a wealth of additional meta-data which needs to be considered.

What the law says

6. There are several instances when an organisation will need to remove personal data from information prior to release as follows:

- When responding to Subject access requests under the DPA;
- When proactively making information available under the Freedom of Information Act (FOIA) or the Environmental Information Regulations (the EIR);
- When responding to information requests under FOIA or the EIR and disclosing third party personal data would breach one of the data protection principles:
- When redacting information that is outside the scope of an FOIA or EIR request is the most efficient way of releasing relevant information that should be disclosed;
- When making personal data available for re-use under the Reuse of Public Sector Information Regulations (RPSI) would breach the data protection principles.

Subject access requests (DPA)

7. The Guide to Data Protection explains that responding to a subject access request may involve providing information that relates both to the individual making the request and to another individual.

8. Section 7(4) of the DPA states:

> 4. Where a data controller cannot comply with the request without disclosing information relating to another individual who can be identified from that information, he is not obliged to comply with the request unless—
> (a) the other individual has consented to the disclosure of the information to the person making the request, or
> (b) it is reasonable in all the circumstances to comply with the request without the consent of the other individual.

9. The Subject access code of practice explains that if it is not possible to gain the consent of the third-party then it may still be possible to provide some information, having edited or 'redacted' information that would identify the third-party.

10. The ICO's guidance on Access to information held in complaint files also explains that information which constitutes personal data of a third party should not be released under subject access unless it would be reasonable in all the circumstances (DPA s.7(4)).

11. Redacting personal data from the information requested means that some information can be released without breaching the data protection principles.

12. Redaction can also be used to remove information which is out of scope of the subject access request because it is not the applicant's personal data.

Information requests under FOIA and the EIR

13. The Guides to Freedom of Information and the Environmental Information Regulations explain that public authorities are obliged to publish certain information about their activities and members of the public are also entitled to request information from public authorities.

14. Section 40(1) of the Freedom of Information Act (FOIA) states:

> 40. (1) Any information to which a request for information relates is exempt information if it constitutes personal data of which the applicant is the data subject.
>
> (2) Any information to which a request for information relates is also exempt information if—
> (a) it constitutes personal data which do not fall within subsection (1), and
> (b) either the first or the second condition below is satisfied.
>
> (3) The first condition is –
>
> (a) in a case where the information falls within any of the paragraphs (a) to (d) of the definition of "data" in section 1(1) of the Data Protection Act 1998, that the disclosure of the information to a member of the

public otherwise than under this Act would contravene –

(i) any of the data protection principles, or

(ii) section 10 of that Act (right to prevent processing likely to cause damage or distress), and

(b) in any other case, that the disclosure of the information to a member of the public otherwise than under this Act would contravene any of the data protection principles if the exemptions in section 33A(1) of the Data Protection Act 1998 (which relate to manual data held by public authorities) were disregarded

15. Information that is the personal data of the applicant is exempt from FOIA under section 40(1) and any request should be processed under section 7 of the DPA. Regulation 5(3) of the EIR is the equivalent provision and has the same effect.

16. Public authorities should not proactively publish third party personal data or release it in response to a request if doing so would breach one of the data protection principles. In such circumstances, section 40(2) (and regulation 13 of the EIR) exempt third party personal data from disclosure in response to a request. In most instances the first data protection principle is most relevant and public authorities must carefully balance the legitimate public interest in disclosure against any prejudice to the data subject's rights, freedoms or legitimate interests when deciding whether personal data can be made available.

Requests for re-use of public sector information

17. The Guide to RPSI explains that information produced as part of a public task can be re-used by an individual or other organisation for a purpose other than the initial public task it was produced for.

18. Section 5(7) of RPSI states:

(7) These Regulations do not apply to—
(a) a document where access is excluded or restricted under information access legislation including on the grounds of protection of personal data, protection of national security, defence or public security, statistical confidentiality or commercial confidentiality (including business, professional or company secrets); or
(b) any part of a document which—
(i) is accessible under information access legislation; and
(ii) contains personal data the re-use of which would be incompatible with the law concerning the protection of individuals with regards to the processing of personal data.

19. An earlier decision to release the information under the appropriate legislation (eg FOIA, EIR, the Freedom of Information (Scotland) Act or the Environmental Information (Scotland) Regulations) may have resulted in a number of redactions in order to comply with the data protection principles.

20. When considering a request to re-use information that contains personal data or redacted information public sector bodies should not automatically assume that the redactions made to the previously disclosed information are sufficient in the context of responding to an application for re-use.

21. Re-use of information may involve significant further processing, combining with other datasets or being made readily accessible to the wider public. As such public sector bodies should closely inspect whether information sought for re-use includes personal data and whether permitting such re-use would comply with the data protection principles.

Background

22. Tools such as graphs, charts and tables can arrange thousands of data points into a more understandable form to educate, inform or convey information hidden within the data.

23. Such summarisation methods can also be an important anonymisation technique and provide a privacy protecting mechanism to allow organisations to publish or make further

use of data derived from individuals without a risk of identification. Such techniques are commonly used to prepare information in response to a freedom of information request or publishing information on a website.

24. Software packages in use today have many hidden dangers. A chart or summary table might not appear to contain any personal data on the surface, but it could in fact have a copy of the individual data points embedded within and allow this data to be made accessible with nothing more than a couple of clicks. Complex file types can also contain meta-data which may not be appropriate for disclosure, such as photographs with embedded GPS coordinates or the routing information of an email.

Example data

25. A range of practical examples are given throughout this guidance to illustrate how personal data can be disclosed in error. The examples used in this guidance are fictional so any apparent personal data included does not relate to living individuals.

26. A full description of the process used to create the dataset can be found in the Appendix and a copy of the dataset used is available in ExampleDataset.csv.

Example software

27. A range of examples are also given using commonly available and frequently used software tools such as Microsoft Office but also open source alternatives such as LibreOffice and OpenOffice.

28. The use of a particular software tool or standard is to provide a practical example to describe a particular issue or solution which may or may not be unique to that tool.

29. In July 2014 the Cabinet Office published a set of open standards for use in government technology. Whilst the use of open standards may be preferable, in a number of cases, converting data into one of these preferred formats (eg PDF/A) is not practical. Furthermore the issues discussed here may

occur when sharing data within an organisation or between two private sector organisations.

Freedom of Information Act requirements for datasets, form and format

30. There is other ICO guidance available on the requirements of the FOIA and the EIR for Means of communication (section 11), Form and format of information (regulation 6) and Datasets (sections 11A, 19 & 45). For datasets this includes a requirement to provide information in an electronic form which is capable of reuse, if reasonably practicable. Public authorities should also consider FOIA section 45 code of practice on datasets. A reusable form means that the dataset is in a machine readable form and based on open standards. The most used format often used to meet this requirement is CSV.

Hidden data

Hiding in plain sight

31. The simplest case of data being disclosed in error can occur when data is not immediately visible on the screen but elsewhere within the file. This can be due to a range of design choices or the rendering of certain formatting styles. For example, when setting up a template a user might have chosen to 'hide' certain data by setting the font colour to be the same as the background (eg white on white or black on black). An example of this type of formatting is HiddenDataExample.xlsx.

32. Whilst hiding data in this manner prevents personal data being disclosed on a printed version of the file, it will still remain within the source file. This personal data is at risk of unintended disclosure if the electronic version is distributed. Highlighting the text or changing the font colour will expose it (see Figure 1).

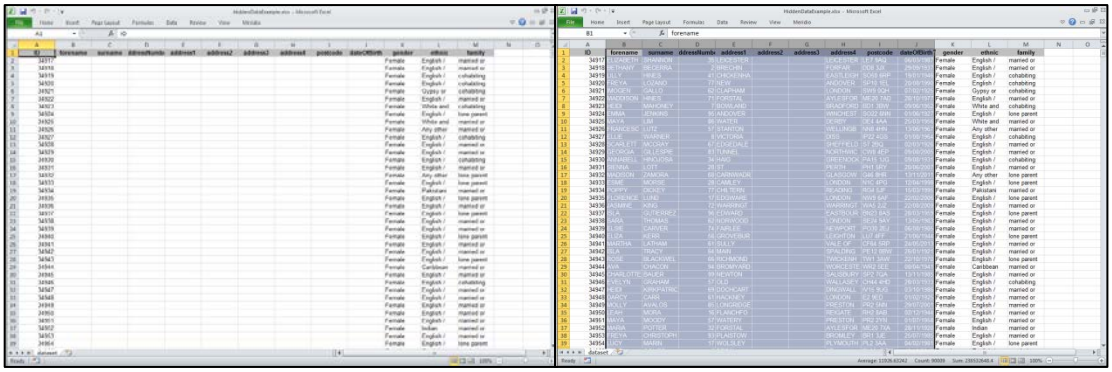


Figure 1: The left hand side shows the identifying elements hidden from immediate view as the font colour is set to white. The right hand side shows the text easily revealed simply by selecting the relevant cells

33. Another example of where data might be hidden from obvious view is when it is placed in the fringes of a file where is it not expected to be found.

34. As an example, Microsoft Excel 2007 and upwards support up to 16,384 columns and 1,048,576 rows of data. A user might place data outside of the normal visible area with the aim to hide it from being displayed on a standard sized monitor. For example PeripheryDataExample.xlsx where the personal data has been moved to columns AA to AI.

Solution: Hiding personal data in this manner is not good practice, is an ineffective way of removing or masking personal data for the purposes of redaction and therefore should be avoided.

A more appropriate practice would be to control access to the file containing the personal data. An alternate view of the data can be created for those individuals who do not need access to the identifying elements.

If you are trying to control printed versions of the file or the data must exist within the file a more appropriate practice would be to set a print area.

Ensuring that information systems have comprehensive and descriptive documentation which indicate the location and content of all data can also reduce the risk that data is 'lost' within a particular file.

Hidden rows and columns

35. A common method of 'hiding' data within a spreadsheet is through the use of hidden rows or columns.

36. Figure 2 shows part of a Microsoft Excel 2010 spreadsheet that contains a short log of property rentals and associated payment information. The data can also be viewed in HiddenColumnsExample.xlsx.

	A	C	D	E	F
1	Property	Tenancy start date	Payment day	Monthly rent	Comments
2	1	05/03/2013	15	£ 495.00	
3	2	09/11/2009	15	£ 575.00	
4	3	15/06/2014	1	£ 495.00	
5	4	01/07/2012	1	£ 525.00	
6	5	01/09/2010	5	£ 400.00	
7					
8					

Figure 2: A log of property rentals and associated payment information

37. Column B is not visible in the column headings but that does not mean that there is no Column B within the file or that it does not contain any personal data. Column B has been set to be hidden, meaning hidden from view. Selecting 'Unhide' from the appropriate submenu is a trivial series of clicks to return the data to full view (shown in Figure 3).

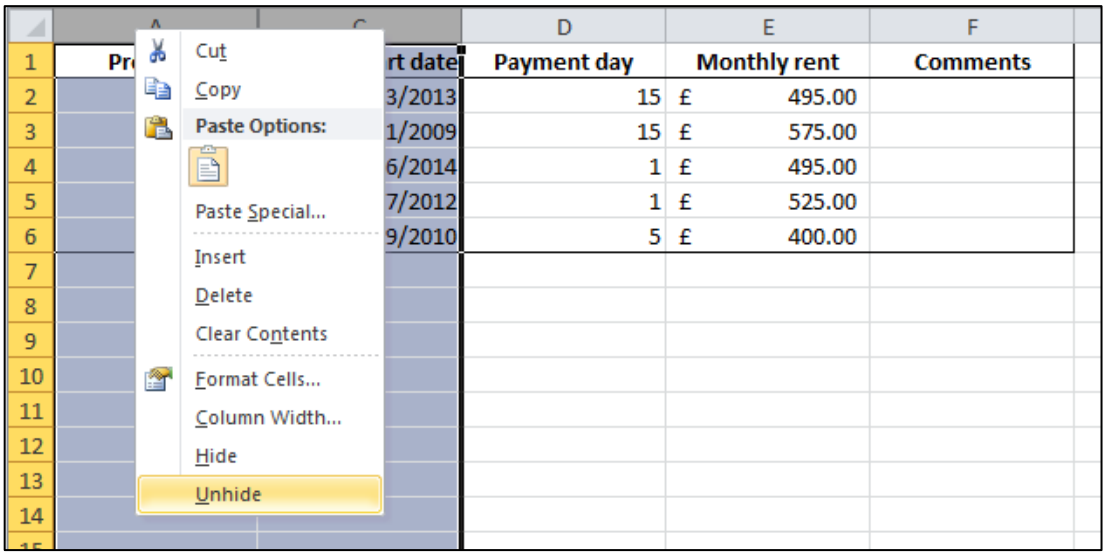


Figure 3: Menu selection required to unhide Column B

38. Once the column has been unhidden personal data contained within the column is accessible to anyone who has access to the file (Figure 4). Therefore the hiding of data in this manner is not an effective or appropriate means to protect against unauthorised access. It is also obvious, from the column headings themselves when hidden columns exist.

	A	B	C	D	E	F
1	Property	Tenant	Tenancy start date	Payment day	Monthly rent	Comments
2	1	Mr A	05/03/2013	15	£ 495.00	
3	2	Mrs B	09/11/2009	15	£ 575.00	
4	3	Mr C	15/06/2014	1	£ 495.00	
5	4	Ms D	01/07/2012	1	£ 525.00	
6	5	Dr E	01/09/2010	5	£ 400.00	
7						

Figure 4: Personal data contained within the previously hidden Column B is now visible

39. Whilst hidden rows and columns might seem a trivial example, it is important to remember that the decision to release a file may have been taken by an individual who has no working knowledge of the file or its original purpose. The original author of the file may even have left the organisation many years previous and it may not be immediately apparent that the data fields were hidden from view.

40. As previously mentioned, hidden rows and columns can be identified from the fact that the (column and row) headings do not flow in a consecutive order. In some versions of Microsoft Excel hidden data can also been identified using the Document Inspector function (Figure 5) which is accessed via the Check for Issues button (Figure 6). The precise method for accessing the Document Inspector depends on the specific version.

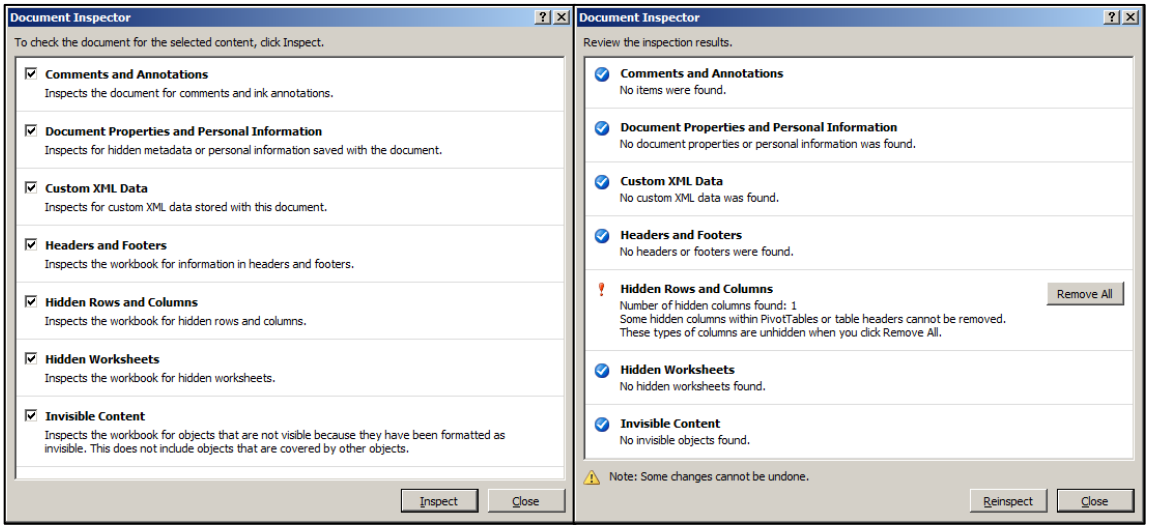


Figure 5: The Document Inspector window (left) and results (right) which can identify hidden rows and columns in Microsoft Excel 2010

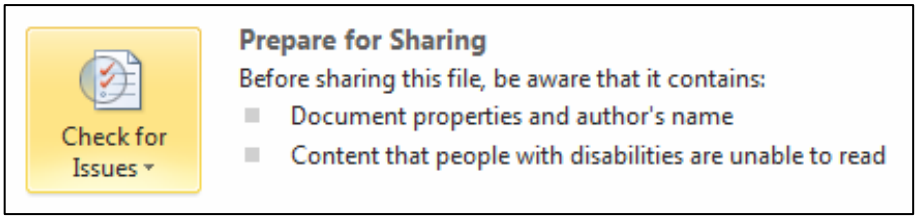


Figure 6: The Check for Issues button in Microsoft Excel 2010

41. An entire worksheet can also be hidden from view. It is less obvious that a worksheet has been hidden as they can be renamed so it is more difficult to notice that a sequential number or letter is missing. However, relying on obscurity as a security measure is poor practice and not to be considered as an appropriate measure to prevent unauthorised access (as hidden sheets can be trivially unhidden).

42. Some word processing software such as LibreOffice and Apache OpenOffice Writer include a feature to conditionally hide paragraphs of text.

43. Some software packages allow the author to password-protect specified fields, pages, columns, rows, worksheets or the entire file. Whilst this may afford some protection against accidental or unauthorised modification of the data it would only be considered appropriate protection against unauthorised or unlawful access if the personal data was protected with an appropriate encryption algorithm and the key or password has also remained a secret. Password protection which makes a file or data elements read-only may not fit this requirement.

Solution: Export to CSV (comma separated value) format.

A solution to the problem of hidden data fields is to export the data into a simple text format such as CSV. This may also be something to consider when disclosing a dataset in a machine-readable open format.

CSV is a format where only the visible text is exported. Columns in spreadsheets are separated with a comma. The CSV file format does not support complex features such as hidden data fields, formulae, type formatting (eg bold and italic) or comment boxes. If a cell contained a formula then just the formula result would be exported.

The exported data can be manually validated by opening the CSV file and inspecting the data. Figure 7 shows the data from the worksheet in Figure 2 exported to CSV and displayed in Notepad. The CSV file itself can also be viewed in HiddenColumnsExample.csv

Simple file formats such as CSV also offer compatibility with a greater range of software packages.

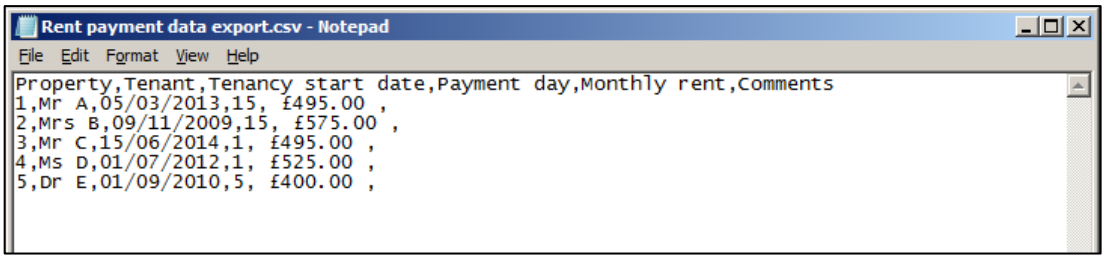


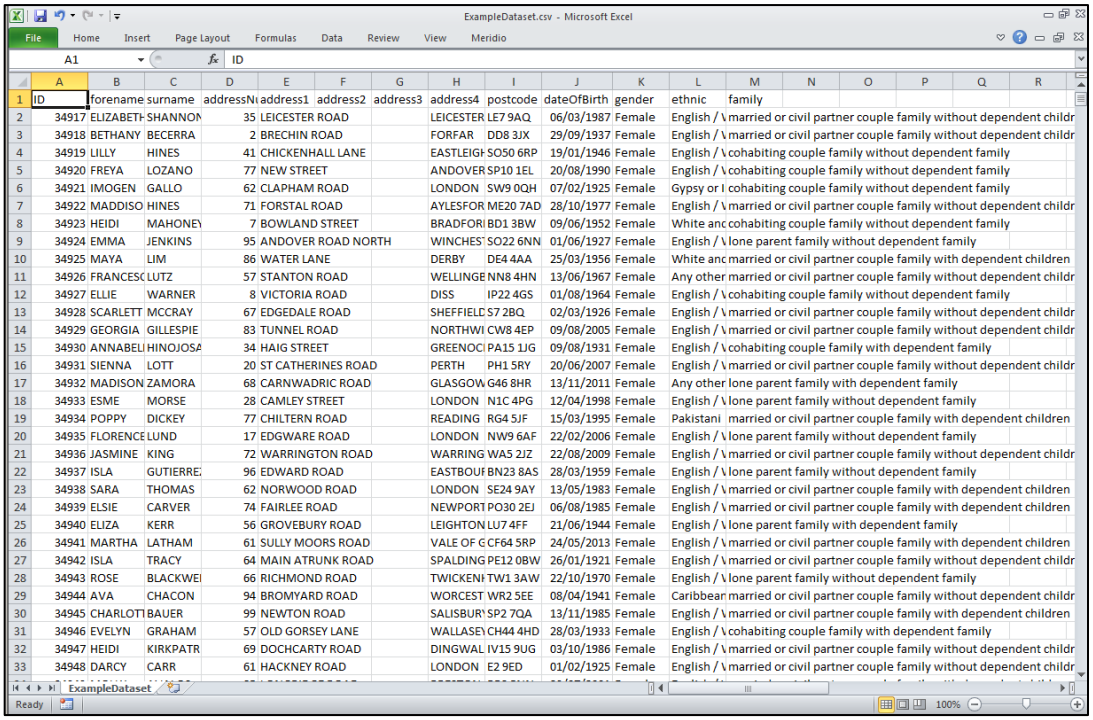
Figure 7: An example of the data shown in Figure 1 exported to CSV. Note the data in column B is clearly visible even though it was hidden when the file was exported.

Solution: Use the Document Inspector to identify hidden data in Microsoft Office files.

Pivot tables

44. A pivot table can be used to summarise a large set of data. This can create an automatic summary of the underlying data.

45. Figure 8 shows an extract of the dataset created for this guidance displayed in Microsoft Excel 2010. A full description of the method used to create this dataset can be found in the Appendix and viewed in ExampleDataset.csv.



ID	forename	surname	addressNo	address1	address2	address3	address4	postcode	dateOfBirth	gender	ethnic	family
34917	ELIZABETH	SHANNON	35	LEICESTER ROAD			LEICESTER	LE7 9AQ	06/03/1987	Female	English / \	married or civil partner couple family without dependent childr
34918	BETHANY	BECERRA	2	BRECHIN ROAD			FORFAR	DD8 3JX	29/09/1937	Female	English / \	married or civil partner couple family without dependent childr
34919	LILLY	HINES	41	CHICKENHALL LANE			EASTLEIGH	SO50 6RP	19/01/1946	Female	English / \	cohabiting couple family without dependent family
34920	FREYA	LOZANO	77	NEW STREET			ANDOVER	SP10 1EL	20/08/1990	Female	English / \	cohabiting couple family without dependent family
34921	IMOGEN	GALLO	62	CLAPHAM ROAD			LONDON	SW9 0QH	07/02/1925	Female	Gypsy or I	cohabiting couple family without dependent family
34922	MADDISO	HINES	71	FORSTAL ROAD			AYLESFOR	ME20 7AD	28/10/1977	Female	English / \	married or civil partner couple family without dependent childr
34923	HEIDI	MAHONEY	7	BOWLAND STREET			BRADFORI	BD1 3BW	09/06/1952	Female	White anc	cohabiting couple family without dependent family
34924	EMMA	JENKINS	95	ANDOVER ROAD NORTH			WINCHES	SO22 6NN	01/06/1927	Female	English / \	lone parent family without dependent family
34925	MAYA	LIM	86	WATER LANE			DERBY	DE4 4AA	25/03/1956	Female	White anc	married or civil partner couple family with dependent children
34926	FRANCESC	LUTZ	57	STANTON ROAD			WELLINGE	NN8 4HN	13/06/1967	Female	Any other	married or civil partner couple family without dependent childr
34927	ELLIE	WARNER	8	VICTORIA ROAD			DISS	IP22 4GS	01/08/1964	Female	English / \	cohabiting couple family without dependent family
34928	SCARLETT	MCCRAY	67	EDGEDALE ROAD			SHEFFIELD	S7 2BQ	02/03/1926	Female	English / \	married or civil partner couple family without dependent childr
34929	GEORGIA	GILLESPIE	83	TUNNEL ROAD			NORTHWI	CW8 4EP	09/08/2005	Female	English / \	married or civil partner couple family without dependent childr
34930	ANNABEL	HINOJOSA	34	HAIG STREET			GREENOC	PA15 1JG	09/08/1931	Female	English / \	cohabiting couple family with dependent family
34931	SIENNA	LOTT	20	ST CATHERINES ROAD			PERTH	PH1 5RY	20/06/2007	Female	English / \	married or civil partner couple family without dependent childr
34932	MADISON	ZAMORA	68	CARNWADRIC ROAD			GLASGOW	G46 8HR	13/11/2011	Female	Any other	lone parent family with dependent family
34933	ESME	MORSE	28	CAMLEY STREET			LONDON	N1C 4PG	12/04/1998	Female	English / \	lone parent family without dependent family
34934	POPPY	DICKEY	77	CHILTERN ROAD			READING	RG4 5JF	15/03/1995	Female	Pakistani	married or civil partner couple family with dependent children
34935	FLORENCE	LUND	17	EDGWARE ROAD			LONDON	NW9 6AF	22/02/2006	Female	English / \	lone parent family without dependent family
34936	JASMINE	KING	72	WARRINGTON ROAD			WARRING	WA5 2JZ	22/08/2009	Female	English / \	married or civil partner couple family without dependent childr
34937	ISLA	GUTIERRE	96	EDWARD ROAD			EASTBOUI	BN23 8AS	28/03/1959	Female	English / \	lone parent family without dependent family
34938	SARA	THOMAS	62	NORWOOD ROAD			LONDON	SE24 9AY	13/05/1983	Female	English / \	married or civil partner couple family with dependent children
34939	ELSIE	CARVER	74	FAIRLEE ROAD			NEWPORT	PO30 2EJ	06/08/1985	Female	English / \	married or civil partner couple family with dependent children
34940	ELIZA	KERR	56	GROVEBURY ROAD			LEIGHTON	LU7 4FF	21/06/1944	Female	English / \	lone parent family with dependent family
34941	MARTHA	LATHAM	61	SULLY MOORS ROAD			VALE OF G	CF64 5RP	24/05/2013	Female	English / \	married or civil partner couple family with dependent children
34942	ISLA	TRACY	64	MAIN ATRUNK ROAD			SPALDING	PE12 0BW	26/01/1921	Female	English / \	married or civil partner couple family without dependent childr
34943	ROSE	BLACKWEI	66	RICHMOND ROAD			TWICKENI	TW1 3AW	22/10/1970	Female	English / \	lone parent family without dependent family
34944	AVA	CHACON	94	BROMYARD ROAD			WORCEST	WR2 5EE	08/04/1941	Female	Caribbean	married or civil partner couple family without dependent childr
34945	CHARLOTT	BAUER	99	NEWTON ROAD			SALISBUR	SP2 7QA	13/11/1985	Female	English / \	married or civil partner couple family with dependent children
34946	EVELYN	GRAHAM	57	OLD GORSEY LANE			WALLASE	CH44 4HD	28/03/1933	Female	English / \	cohabiting couple family with dependent family
34947	HEIDI	KIRKPATR	69	DOCHCARTY ROAD			DINGWAL	IV15 9UG	03/10/1986	Female	English / \	married or civil partner couple family without dependent childr
34948	DARCY	CARR	61	HACKNEY ROAD			LONDON	E2 9ED	01/02/1925	Female	English / \	married or civil partner couple family without dependent childr

Figure 8: An extract of the example dataset displayed in Microsoft Excel 2010

46. A summary of this dataset can be created using the pivot table feature. Figure 9 shows the pivot table created to summarise the number of individuals by gender and family type. The pivot table can be viewed in file PivotTableExample.xlsx.

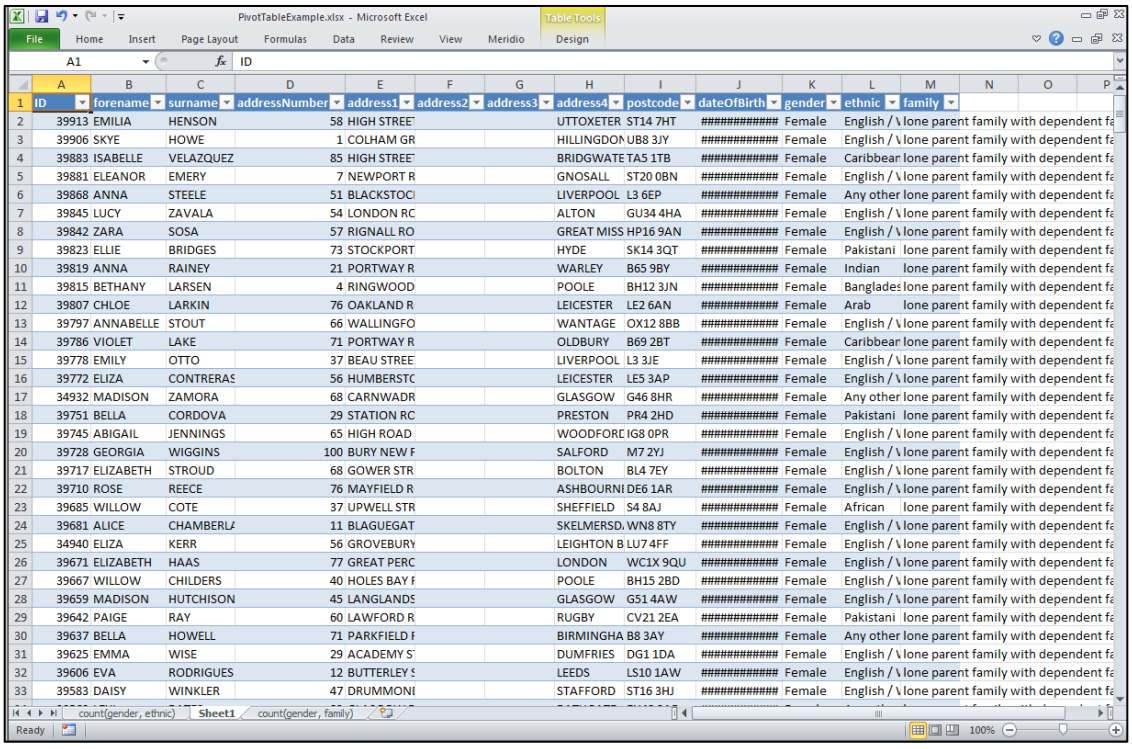
Row Labels	Count of ID
Female	**5000**
cohabiting couple family with dependent family	277
cohabiting couple family without dependent family	456
lone parent family with dependent family	518
lone parent family without dependent family	500
married or civil partner couple family with dependent children	1238
married or civil partner couple family without dependent children	2011
Male	**5000**
cohabiting couple family with dependent family	300
cohabiting couple family without dependent family	442
lone parent family with dependent family	507
lone parent family without dependent family	506
married or civil partner couple family with dependent children	1216
married or civil partner couple family without dependent children	2029
Grand Total	**10000**

Figure 9: A pivot table displaying a summary of individuals by Gender and Family type

47. As with hidden data fields, despite the fact that the underlying data is not immediately visible on the screen it can still be accessed. A double-click on the pivot table can signal to the

software to automatically extract the data used to calculate the clicked data and display this in a new worksheet.

48. Even if the worksheet containing the original data is deleted from the workbook or if the pivot table is copied into a new workbook, the underlying data may be copied across with it, making the data accessible to the user (eg file PivotTableExample.xlsx does not contain a worksheet containing the original dataset). Figure 10 shows the result of double-clicking on the count of 'Female / Lone parent family with dependent family'. The personal data of the 518 individuals which were summarised in that row are extracted by the software and displayed in a new worksheet.



ID	forename	surname	addressNumber	address1	address2	address3	address4	postcode	dateOfBirth	gender	ethnic	family
39913	EMILIA	HENSON	58	HIGH STREE			UTTOXETER	ST14 7HT	##########	Female	English / \	lone parent family with dependent fa
39906	SKYE	HOWE	1	COLHAM GR			HILLINGDON	UB8 3JY	##########	Female	English / \	lone parent family with dependent fa
39883	ISABELLE	VELAZQUEZ	85	HIGH STREE			BRIDGWATE	TA5 1TB	##########	Female	Caribbean	lone parent family with dependent fa
39881	ELEANOR	EMERY	7	NEWPORT R			GNOSALL	ST20 0BN	##########	Female	English / \	lone parent family with dependent fa
39868	ANNA	STEELE	51	BLACKSTOCI			LIVERPOOL	L3 6EP	##########	Female	Any other	lone parent family with dependent fa
39845	LUCY	ZAVALA	54	LONDON RC			ALTON	GU34 4HA	##########	Female	English / \	lone parent family with dependent fa
39842	ZARA	SOSA	57	RIGNALL RO			GREAT MISS	HP16 9AN	##########	Female	English / \	lone parent family with dependent fa
39823	ELLIE	BRIDGES	73	STOCKPORT			HYDE	SK14 3QT	##########	Female	Pakistani	lone parent family with dependent fa
39819	ANNA	RAINEY	21	PORTWAY R			WARLEY	B65 9BY	##########	Female	Indian	lone parent family with dependent fa
39815	BETHANY	LARSEN	4	RINGWOOD			POOLE	BH12 3JN	##########	Female	Banglades	lone parent family with dependent fa
39807	CHLOE	LARKIN	76	OAKLAND R			LEICESTER	LE2 6AN	##########	Female	Arab	lone parent family with dependent fa
39797	ANNABELLE	STOUT	66	WALLINGFO			WANTAGE	OX12 8BB	##########	Female	English / \	lone parent family with dependent fa
39786	VIOLET	LAKE	71	PORTWAY R			OLDBURY	B69 2BT	##########	Female	Caribbean	lone parent family with dependent fa
39778	EMILY	OTTO	37	BEAU STREE			LIVERPOOL	L3 3JE	##########	Female	English / \	lone parent family with dependent fa
39772	ELIZA	CONTRERAS	56	HUMBERSTC			LEICESTER	LE5 3AP	##########	Female	English / \	lone parent family with dependent fa
34932	MADISON	ZAMORA	68	CARNWADR			GLASGOW	G46 8HR	##########	Female	Any other	lone parent family with dependent fa
39751	BELLA	CORDOVA	29	STATION RC			PRESTON	PR4 2HD	##########	Female	Pakistani	lone parent family with dependent fa
39745	ABIGAIL	JENNINGS	65	HIGH ROAD			WOODFORD	IG8 0PR	##########	Female	English / \	lone parent family with dependent fa
39728	GEORGIA	WIGGINS	100	BURY NEW F			SALFORD	M7 2YJ	##########	Female	English / \	lone parent family with dependent fa
39717	ELIZABETH	STROUD	68	GOWER STR			BOLTON	BL4 7EY	##########	Female	English / \	lone parent family with dependent fa
39710	ROSE	REECE	76	MAYFIELD R			ASHBOURNI	DE6 1AR	##########	Female	English / \	lone parent family with dependent fa
39685	WILLOW	COTE	37	UPWELL STR			SHEFFIELD	S4 8AJ	##########	Female	African	lone parent family with dependent fa
39681	ALICE	CHAMBERLA	11	BLAGUEGAT			SKELMERSD	WN8 8TY	##########	Female	English / \	lone parent family with dependent fa
34940	ELIZA	KERR	56	GROVEBURY			LEIGHTON B	LU7 4FF	##########	Female	English / \	lone parent family with dependent fa
39671	ELIZABETH	HAAS	77	GREAT PERC			LONDON	WC1X 9QU	##########	Female	English / \	lone parent family with dependent fa
39667	WILLOW	CHILDERS	40	HOLES BAY F			POOLE	BH15 2BD	##########	Female	English / \	lone parent family with dependent fa
39659	MADISON	HUTCHISON	45	LANGLANDS			GLASGOW	G51 4AW	##########	Female	English / \	lone parent family with dependent fa
39642	PAIGE	RAY	60	LAWFORD R			RUGBY	CV21 2EA	##########	Female	Pakistani	lone parent family with dependent fa
39637	BELLA	HOWELL	71	PARKFIELD F			BIRMINGHA	B8 3AY	##########	Female	Any other	lone parent family with dependent fa
39625	EMMA	WISE	29	ACADEMY S			DUMFRIES	DG1 1DA	##########	Female	English / \	lone parent family with dependent fa
39606	EVA	RODRIGUES	12	BUTTERLEY S			LEEDS	LS10 1AW	##########	Female	English / \	lone parent family with dependent fa
39583	DAISY	WINKLER	47	DRUMMONI			STAFFORD	ST16 3HJ	##########	Female	English / \	lone parent family with dependent fa

Figure 10: An extract of the personal data of the 518 female individuals within the lone parent family with dependent family category

Solution: Export the worksheet containing the pivot table to CSV (eg PivotTableExample.csv).

The exported pivot table can be validated by opening the CSV file and manually inspecting the data. An example of the Gender and Family type summary pivot table is shown in Figure 11 below.

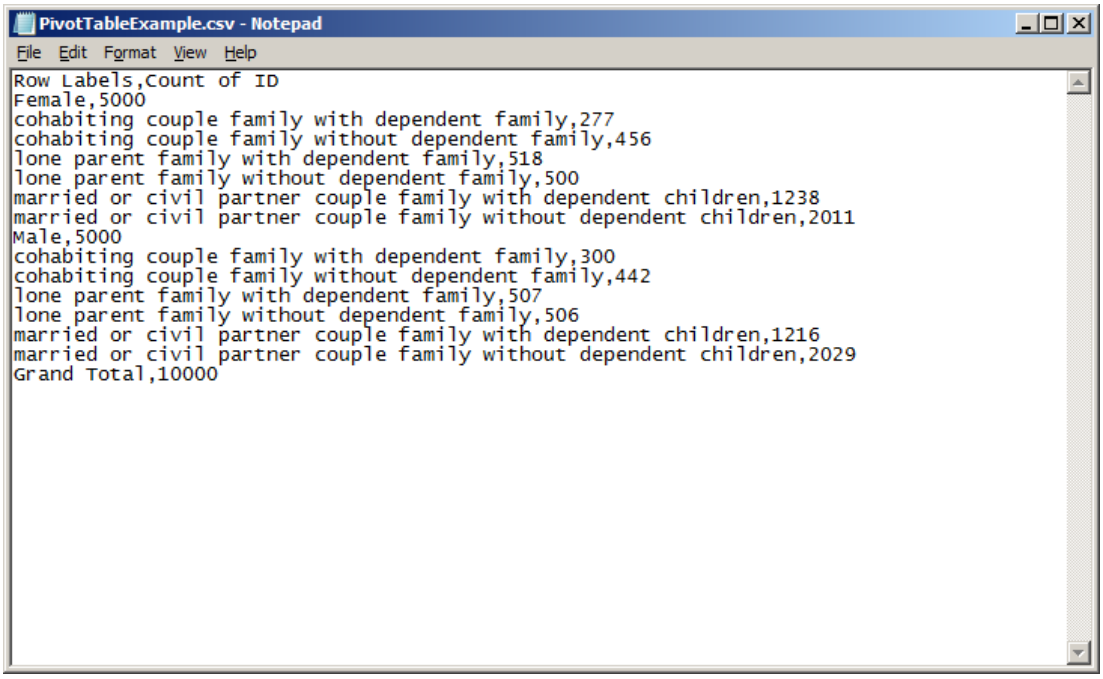


Figure 11: An example of a pivot table exported to CSV

Solution: Copy the pivot table and paste only the values to a new workbook. This is sometimes referred to as a 'paste special' operation.

The copied data can also be checked by double-clicking cells within the copied pivot table to ensure there is no link back to the source data. However, exporting the data to CSV would provide a greater assurance and provide greater compatibility with other software packages.

Charts

49. As with pivot tables, charts can also contain an embedded copy of the source data. A further risk could arise when a chart is embedded into a document or presentation as the embedded chart could also contain a copy of the source data.

50. Figure 12 shows an example of a chart being created from a pivot table summarising the count of family types in the dataset (see also EmbeddedChartExample.xlsx).

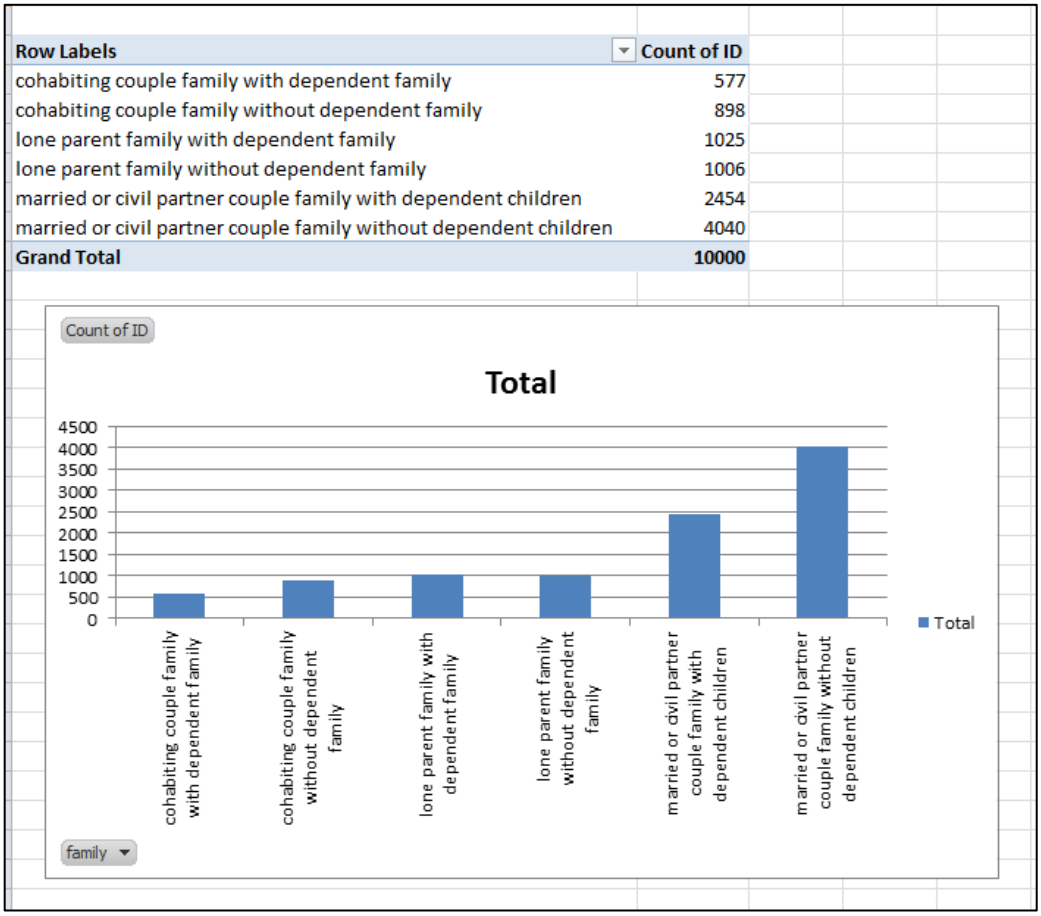


Figure 12: A chart and pivot table summarising the count of family types in the dataset

51. If the chart is embedded within a Microsoft Office Word document (see Figure 1 in EmbeddedChartExample.docx) then a copy of the underlying data is also copied across and embedded within the document. Simply double-clicking on the chart and selecting the data worksheet can reveal the underlying data (Figure 13).

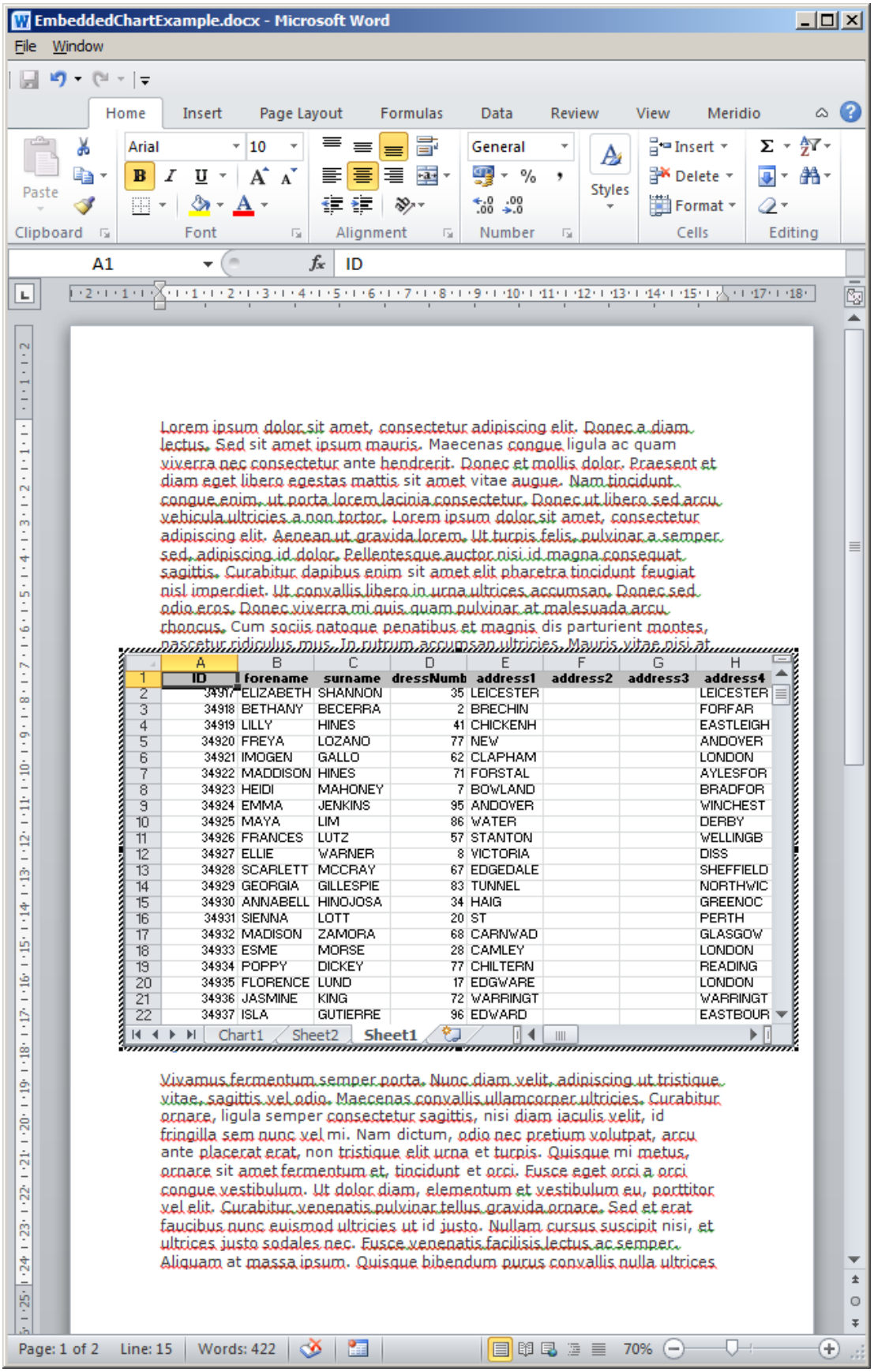


Figure 13: After double-clicking the embedded graph object the underlying data can be revealed by selecting the data worksheet

Solution: Copy the chart and paste as an image file (eg jpg or png) into the destination file. An example of inserting a chart as an image file can be seen in Figure 2 of EmbeddedChartExample.docx). Exporting the document into a format such as PDF will also remove the underlying source data from the graph.

Solution: Create the chart using a summarised version of the data (eg a pivot table with source data removed). Providing a copy of the anonymised chart source data would therefore not be a disclosure of personal data.

Functions

52. Functions such as LOOKUP and VLOOKUP also create and store a cache of the source data that can be exposed through careful manipulation of the function.

53. Figure 14 shows the VLOOKUP function that can be found in VLookupExample.ods. The source data was located in a different file (in this example in C:\Users\Public\VLookupDataset.ods) but a cache of the dataset is also stored within the current file.

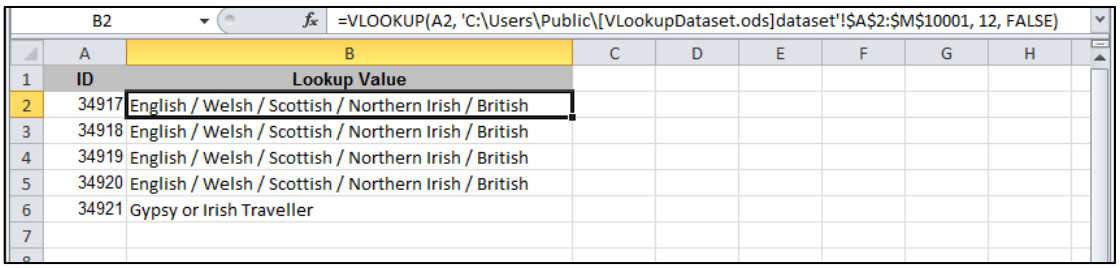


Figure 14: A VLOOKUP function referencing a dataset in a different file

54. This can be demonstrated by editing the column index in the VLOOKUP formula. By manually changing the column index from 12 to 2 the formula will return data from the second column in the cache without access to the source file as can be seen in Figure 15.

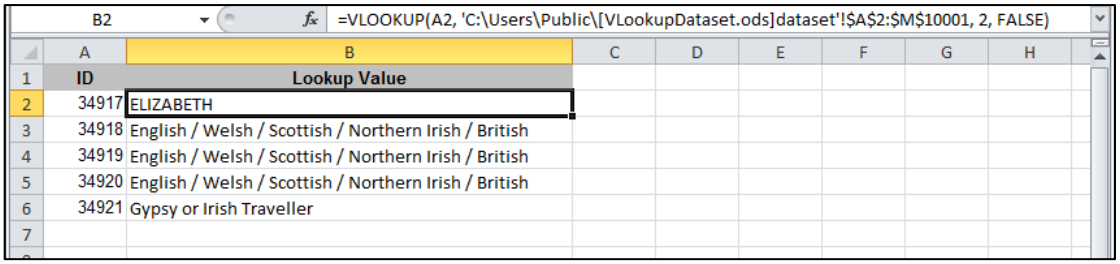


Figure 15: Manually editing the VLOOKUP function can return a different value from the cache

Solution: Export the VLOOKUP values to CSV (VLookupExample.csv.

The exported values can be validated by opening the CSV file and manually inspecting the data.

Ineffective redaction

55. When disclosing information under FOIA or in response to a subject access request, it may be necessary to remove or redact certain information.

56. You can only withhold an entire document under FOIA if all the information is exempt from disclosure under an exemption or the redaction renders the document meaningless. The ICO's Guide to freedom of information provides further guidance on what to consider when redacting documents.

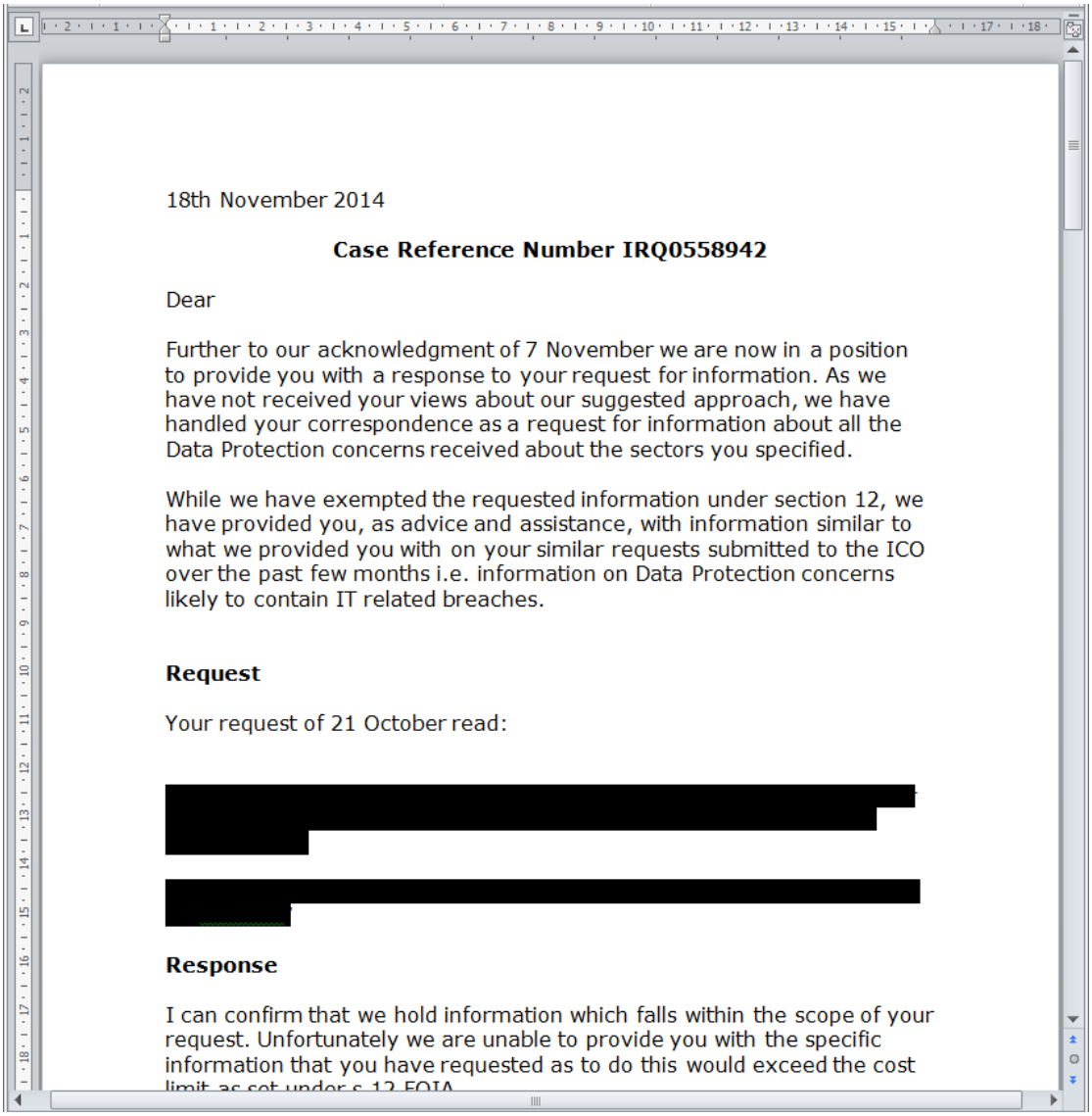
57. Public authorities should also consider the redaction toolkit provided by the National Archives. The toolkit defines redaction:

> "Redaction is the separation of disclosable from non-disclosable information by blocking out individual words, sentences or paragraphs or the removal of whole pages or sections prior to the release of the document. In the paper environment some organisations will know redaction as extracts when whole pages are removed, or deletions where only a section of text is affected."

58. The purpose of redaction is to irreversibly remove the exempt information from the redacted copy of the information. Care should be taken to protect against deleting data from the original file.

59. When it is necessary to provide a copy of a report, memo or email it will be important to ensure that, where appropriate, personal data is redacted from the file.

60. In a similar manner to 'hiding data in plain sight' described earlier, an author might be tempted to use the highlighter tool to add a black box around text marked for redaction. Using the ICO's response to freedom of information request IRQ0558942 as an example, attempting to redact the original request with the black highlighter tool will give the result shown in Figure 16 below and in file IneffectiveRedactionExampleIneffectiveRedaction.docx.



18th November 2014

Case Reference Number IRQ0558942

Dear

Further to our acknowledgment of 7 November we are now in a position to provide you with a response to your request for information. As we have not received your views about our suggested approach, we have handled your correspondence as a request for information about all the Data Protection concerns received about the sectors you specified.

While we have exempted the requested information under section 12, we have provided you, as advice and assistance, with information similar to what we provided you with on your similar requests submitted to the ICO over the past few months i.e. information on Data Protection concerns likely to contain IT related breaches.

Request

Your request of 21 October read:

[redacted]

Response

I can confirm that we hold information which falls within the scope of your request. Unfortunately we are unable to provide you with the specific information that you have requested as to do this would exceed the cost limit as set under s.12 FOIA

Figure 16: Freedom of Information request IRQ0558942 with the original request hidden from view using the black highlighter tool in Microsoft Word 2010

61. If the Word file was disclosed as a printed copy (or printed, scanned and emailed) this would prevent disclosure of the

desired information. However it is important to recognise that the information still exists underneath the black box in the original electronic file. If this file was retained this may not be clear to future readers.

62. Even using a 'Save as PDF' or 'Print as PDF' function is unlikely to provide effective redaction because the PDF file format can support formatting marks such as the highlighter. As an example, the Microsoft Office Word 2010 document was exported to PDF format and made available as IneffectiveRedactionExampleIneffectiveRedaction.pdf.

63. Simply copying the highlighted text and pasting to a text editor will reveal the content because the formatting will not be copied across as shown in Figure 17 below.

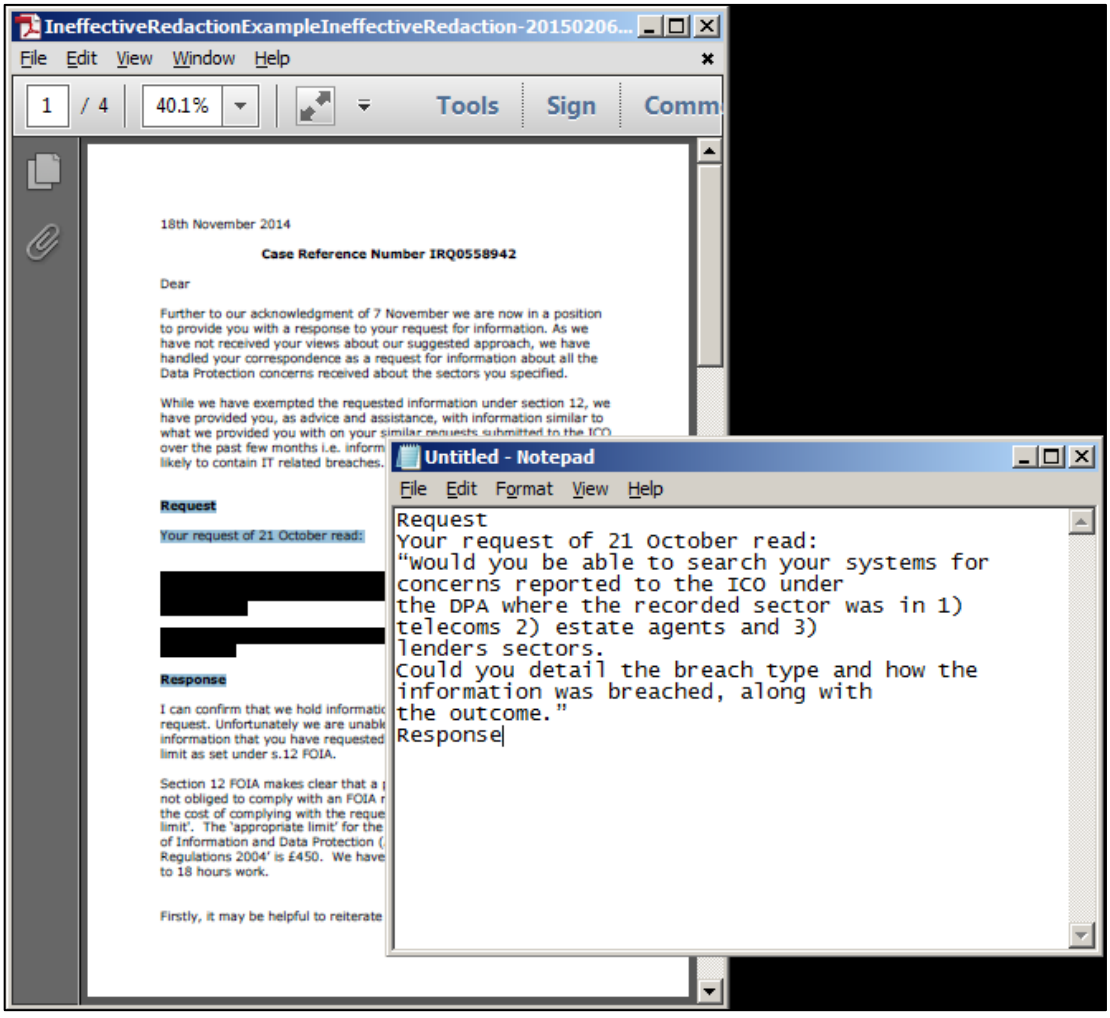


Figure 17: Extracting the hidden text from IneffectiveRedactionExampleIneffectiveRedaction-20150206.pdf

64. This problem might have arisen if one individual marked text to be redacted by someone else in the organisation (perhaps

someone who has access to appropriate software). There is a risk that the marked-up document is retained and accessed in the future by a third individual who mistakenly believes that the marked text is redacted when it is not and discloses the file.

65. Specific redaction software can be used to redact information permanently if you intend to release the file in an electronic form. This type of software generally converts the text in the PDF into an image such that it cannot be copied (for the result see IneffectiveRedactionExampleEffectiveRedaction.pdf). One consequence of performing such a step is that the information is no longer in a machine-readable format meaning that the text cannot be extracted and further processed. For some purposes, such as the example describe here, this will be an acceptable outcome.

Solution: If you are using a highlighter tool to mark text for someone else to redact, do not use a black highlighter. A different colour (eg yellow) will clearly indicate which text requires redaction yet also show that the original text remains.

Solution: For permanent redaction use specific redaction software.

Photography and video

66. A common query received by the ICO concerns the redaction of personal data from still images and video, for example, CCTV footage being disclosed as a result of a subject access request.

67. The ICO's code of practice for surveillance cameras and personal information states that:

> 'When disclosing surveillance images of individuals, particularly when responding to subject access requests, you need to consider whether the identifying features of any of the other individuals in the image need to be obscured. In most cases the privacy intrusion to third party individuals will be minimal and obscuring images will not be required. However, consideration should be given to the nature and context of the footage.'

68. Obscuring information from a single image can be a straightforward task as most operating systems include a simple image editing tool that can be used to blur or cover part

of the image, for example, with a black box. It is however important to ensure that the redacted image is exported to a simple 'un-layered' format to ensure that the redactions are permanent. This is a similar problem to the covering of text with the highlighter tool described previously. It is also worth considering whether information that you have not redacted may still result in someone being identified, clothing, skin or hair colour for example.

69. It can be more complex to obscure information from video in part due to the larger volume of data. CCTV footage stored in proprietary formats or low frame rates may also present difficulties.

70. Redacting the personal data of third-parties is likely to require the use of a specialist software tool to achieve this effectively. The task can also be contracted out to another organisation. If you do choose to contract out to another organisation you are likely to be using a data processor that will require you to have certain measures in place, such as a written contract, in order to comply with Principle 7 of the DPA.

Solution: Redact information within an individual image through the use of a simple image editing tool included with most operating systems or using a specialist redaction software tool.

Meta-data

71. Files rarely contain just the information entered by the author or just what is displayed on the screen. So-called meta-data or 'data about data' is embedded within the file and can include information such as previous authors, changes made to previous versions, comments or annotations. Photographs taken with smartphones and tablets can contain the GPS coordinates of where the image was taken, time and date or information about the type of device used. Emails contain information about the sender and recipient as well as routing information about how the message was delivered.

72. Publication of such complex file types in their raw form can contain an amount of meta-data that may not be appropriate for disclosure.

File properties

73. Office suite software including word processing, spreadsheets and presentations can embed information such as the author, comments and version history into files. Some of these can be populated by the software itself such as the name of the person who is logged in may be automatically inserted into the author field.

74. The Document Inspector mentioned previously can highlight certain file properties in Microsoft Office files (ie Word, Excel and Powerpoint) including comments, annotations and version history. LibreOffice and OpenOffice can view information about the file in the Properties dialog.

Email

75. The technical specification for email defines a number of required and optional fields of meta-data. Some of these are necessary for the successful delivery of the communication. Others exist as a record of the route used for delivery and others assist in virus scanning or SPAM identification.

76. You should remember that if you intend to redact information such as the sender's or recipient's email address or part of the email subject you may also need to remove this from the meta-data or remove the meta-data entirely.

77. Releasing the original electronic version of an email may also disclose any attachments. You should make sure that these do not contain personal data that should not be disclosed.

Solution: If you need to disclose an email without meta-data you can disclose a printed version of the message (or print-as-PDF version).

If you need to redact a sender's or recipient's email address in the printed version you can use a method described earlier.

EXIF

78. The meta-data that can be contained within image file formats also deserves special mention due to its potential sensitivity. In particular, photographs taken with smart phones and tablets can include the GPS coordinates of where the image was taken as well as other data about the device used to take the

photograph including; date, time, shutter speed, ISO speed, make and model. Whilst some of these are unlikely to relate to the individual who took the photograph others may do (such as location data).

79. Settings within the device may be able to control some of these data types so that they are not included in the first place if they are not required. Specialist redaction can extract just the image leaving the EXIF data in the original file. In addition, photo-editing software can also display and provide an interface for editing and/or deleting the EXIF data. In all cases of redaction or deletion of data, care should be taken to protect against deleting data from the original file.

80. Figure 18 shows a photograph of the Alan Turing statue which can be found within Sackville Gardens in Manchester city centre (see TuringBenchWithGPS.jpg). The GPS coordinates were automatically embedded within the image by the smart phone.

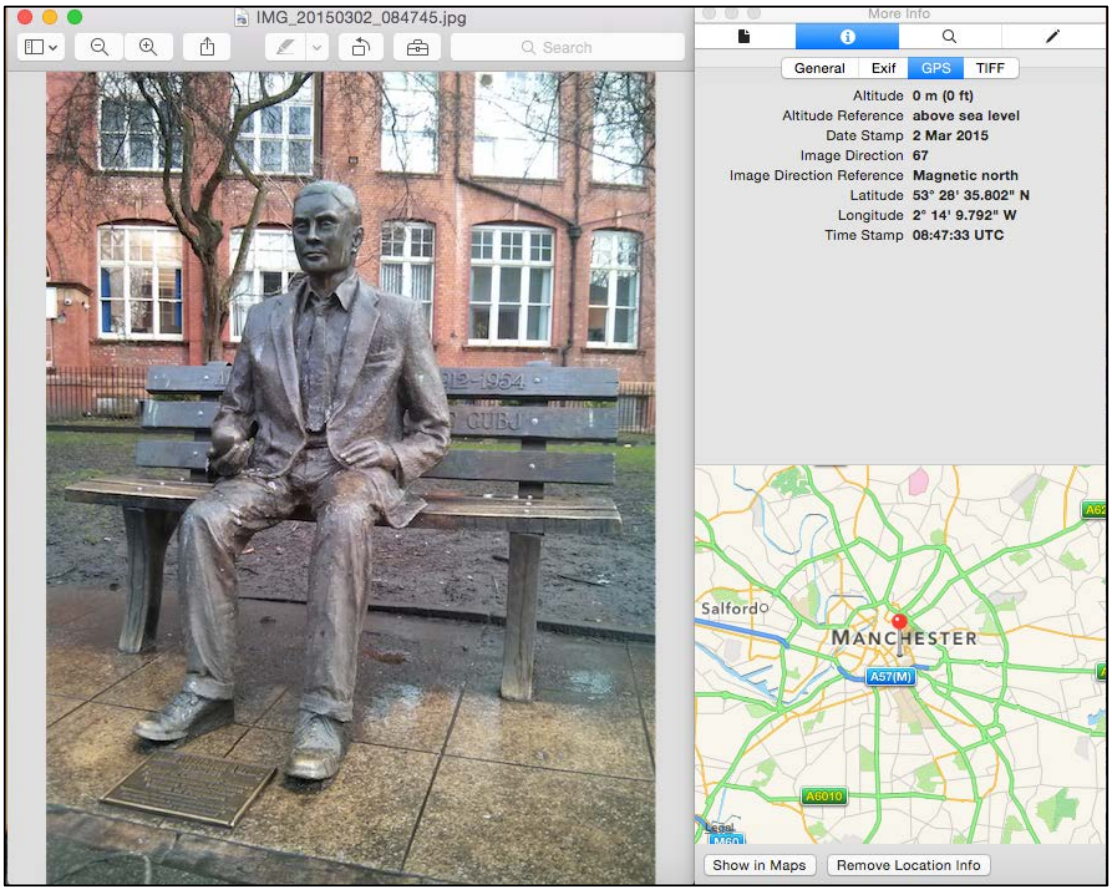


Figure 18: A photograph taken of the Alan Turing statue in Sackville Park, displayed using Apple OSX Finder. The GPS coordinates were automatically embedded by the smart phone.

Solution: Manually review the EXIF meta-data prior to disclosure using photo viewing or editing software. Edit or delete the data required for redaction and share the file.

Solution: Use a bespoke redaction software tool which will extract just the image data. Such a tool may also allow redaction of individual pixels within the image data.

More information

81. Additional data protection guidance is available on our guidance pages if you need further information on other parts of the DPA. Further guidance on FOIA and the EIR is available in the Guide to Freedom of Information and the Guide to the Environmental Information Regulations.

82. This guidance has been developed drawing on ICO experience. Because of this it may provide more detail on issues that are often referred to the Information Commissioner than on those we rarely see. The guidance will be reviewed and considered from time to time in line with new decisions of the Information Commissioner, Tribunals and courts.

83. It is a guide to our general recommended approach, although individual cases will always be decided on the basis of their particular circumstances.

84. If you need any more information about this or any other aspect of data protection, please contact us, or visit our website at www.ico.org.uk.

Checklist

85. The following checklist highlights a number of things to consider when disclosing certain data types that may contain personal data. It is good practice to keep a record of all transformations or redactions you make and to retain the original records used.

File type	Considerations
Spreadsheet eg xls(x), ods	o Are you sure you know where all the data is? o Are there hidden columns? o Are there hidden rows? o Are there hidden work sheets? o Do pivot tables contain linked data? o Do charts contain linked data? o Are there formula included which link to external files? o Is there any meta-data that should be removed? o Is the file size larger than you might expect for the volume of data being disclosed?
Word processor eg doc(x), odt	o Are there any comments within the document that should be removed? o Does the document contain a version history? o Do pivot tables contain linked data? o Do charts contain linked data? o Is there any meta-data that should be removed? o Does the document title or filename contain any personal data (eg Letter to John Smith)? o Has a header or footer been automatically added to a print-out?
Presentation eg ppt(x), odp	o Are there any presenter notes which should be removed? o Do pivot tables contain linked data?

	o Do charts contain linked data? o Is there any meta-data that should be removed?
PDF	o Are there any comments which should be removed? o Are all redactions effectively applied? o Is there any meta-data that should be removed?
Email eg mbox, msg	o Is there data within any attachments that also needs to be redacted? o Is there any meta-data that should be removed?
Image and video eg jpg, avi	o Is there attached EXIF data? o Is there personal data that needs to be obscured (eg faces of third-party individuals?)

Appendix

86. In order to provide a realistic set of examples for this guidance a dataset was created to give the illusion of a disclosure of personal data.

87. A dataset containing data types which would typically be considered personal data for 10,000 individuals was created using the following data sets:

- Top 100 boys names 2013, http://www.ons.gov.uk/ons/about-ons/business-transparency/freedom-of-information/what-can-i-request/previous-foi-requests/population/baby-names-2013/2013-baby-names-boys.xls
- Top 100 girls names 2013, http://www.ons.gov.uk/ons/about-ons/business-transparency/freedom-of-information/what-can-i-request/previous-foi-requests/population/baby-names-2013/2013-baby-names-girls.xls
- Surnames occurring 100 times or more from Census 2000, http://www.census.gov/topics/population/genealogy/data/2000_surnames.html
- Ethnic origin codes and distribution as reported by the 2011 UK census, http://www.ons.gov.uk/ons/rel/census/2011-census/key-statistics-for-local-authorities-in-england-and-wales/rpt-ethnicity.html#tab-Ethnicity-in-England-and-Wales
- Gender codes Male (M) or Female (F)
- Family and household codes and distribution as reported by the 2011 UK census, http://www.ons.gov.uk/ons/rel/family-demography/families-and-households/2012/stb-families-households.html#tab-Families
- Address of UK vehicle testing stations, http://data.gov.uk/dataset/mot-active-vts

88. Surnames were filtered such that only those occurring 20,000 times or more were used

89. The address of vehicle testing stations were filtered to use only those containing words ROAD, STREET, VIEW, LANE or DRIVE and to exclude any containing GARAGE, ESTATE, UNIT, BUSINESS, SERVICE, TRADING, IND, INDUSTRIAL, COMMERCIAL, FILLING or RETAIL. Building numbers were also removed.

90. Individual records were generated by combining the following data for each gender category:

- Randomly select a gender appropriate first name
- Randomly select a surname
- Randomly select an ethnic origin (probability weighted by 2011 census distribution)
- Randomly select a family and household code (probability weighted by 2011 census distribution)
- Randomly select a building number between 1-99
- Randomly select an address

91. This process was repeated until a sufficiently large population had been created. For this guidance a dataset containing 5,000 male and 5,000 female records was generated.