Managing your Dissertation Data

Scott Summers and Maureen Haaker

Thinking Ahead - Sociology Class

30th January 2018

Overview of this session

- Managing your data – why and how
  - Assessment
  - Consent
  - Anonymisation
  - Documentation
  - Storage (security, backups, encryption, formatting, etc.)
- Resources available
- Your questions

Why is it important to manage research data well?

- Good quality data leads to good quality research
- Data underpins published findings
  - Documentation can be used in dissertation write-up
  - Documentation can be used in a viva
- Helps fodder discussion in dissertation supervision about how to collect and analyse data
- To protect data from loss, destruction and potential exposure
- Enables compliance with ethical codes, and data protection laws*
- Enhances transparency of research and can authenticate your dissertation progress

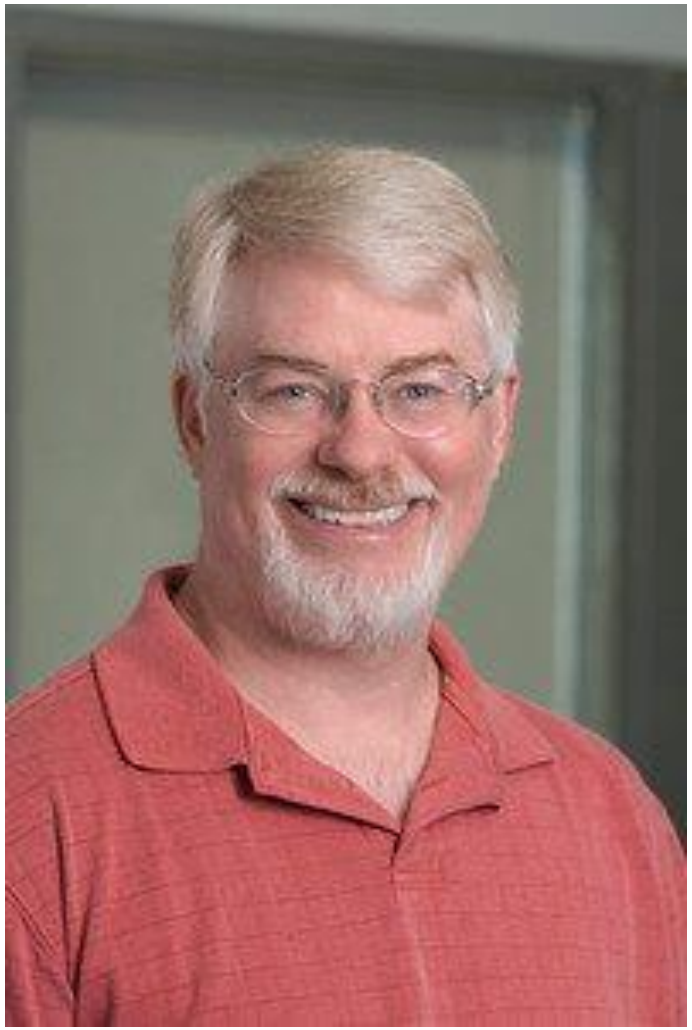
*As you go on to further research, journals and funders often have requirements and/or policies about data management.

University of Essex

UK · DATA ARCHIVE

Meet the scientist whose ideas were stolen at least three times

Jeff Offutt, a professor of software engineering at George Mason University, has some stories to tell. He says that when one of his students wrote his first paper, the student reused four paragraphs from another source, not knowing he couldn’t do that



http://retractionwatch.com/2018/01/15/meet-scientist-whose-ideas-stolen-least-three-times/

Someone Has Stolen My Project :-(

posted
06-Sep-09, 02:42
by birdie

postgrad forum

posted about 9 years ago

Dear Everyone,
I am a second year PhD student and I wanted to ask if anyone has experienced someone else stealing your ideas? Basically, I wrote a grant proposal 9 months ago and emailed it to someone who said they might be able to help me. A few weeks later I emailed to ask if I hadbeen successful and I had a one sentance email to say that 'sorry, I do not have the funds'. I have just found out, by chance, they the same study is now being carried out by this person, who is not academic, but has recruited several academics to do analysis etc. It is the same work I was going to do (and have started) and would comprise 2 chapters of my thesis. The people that are involved with his study and the funds he has for equipment means that it is impossible for me to compete and get a paper out before his.

Has anyone else experienced
second year now :-(

A professor may have stolen an idea I discussed with him: how to proceed?

17

2

I'm really not sure how to proceed from here.

I'm an undergraduate Computer Science student, planning to earn my MSc in the next couple of years.

A couple of weeks ago, I spoke with a professor (I'll call him A) on an informal setting, and, since he was my teacher on an area relevant to the subject, I discussed with him an idea I had, and about how I was trying to make it my own thesis project.

I have also previously "pitched" this idea to another professor (B). He liked it, and we agreed that once I graduated, he was willing to be my advisor, and I could work on it.

However, I was recently shown the list of thesis proposals for current students, and my idea was among them - being supervised by prof. A, and an unrelated professor, C.

The details are so close to what we have discussed that this is almost surely no coincidence.

Now, I can't say I'm 100% sure he stole my idea, but he didn't mention any of this when we talked, and this thesis list was made *after* our conversation.

Anything could have happened: either I'm right, or prof. C actually came up with the same idea, or prof. A had a similar idea in the past but didn't tell me about it.

Practical steps you can take

- Consider how to manage your data early

- Make sure you can understand your data and it is protected:
  - obtain consent to share data with your supervisor
  - do not disclose identities without consent
  - provide clear documentation
  - create a datalist
  - store your data safely at all stages

Assessment

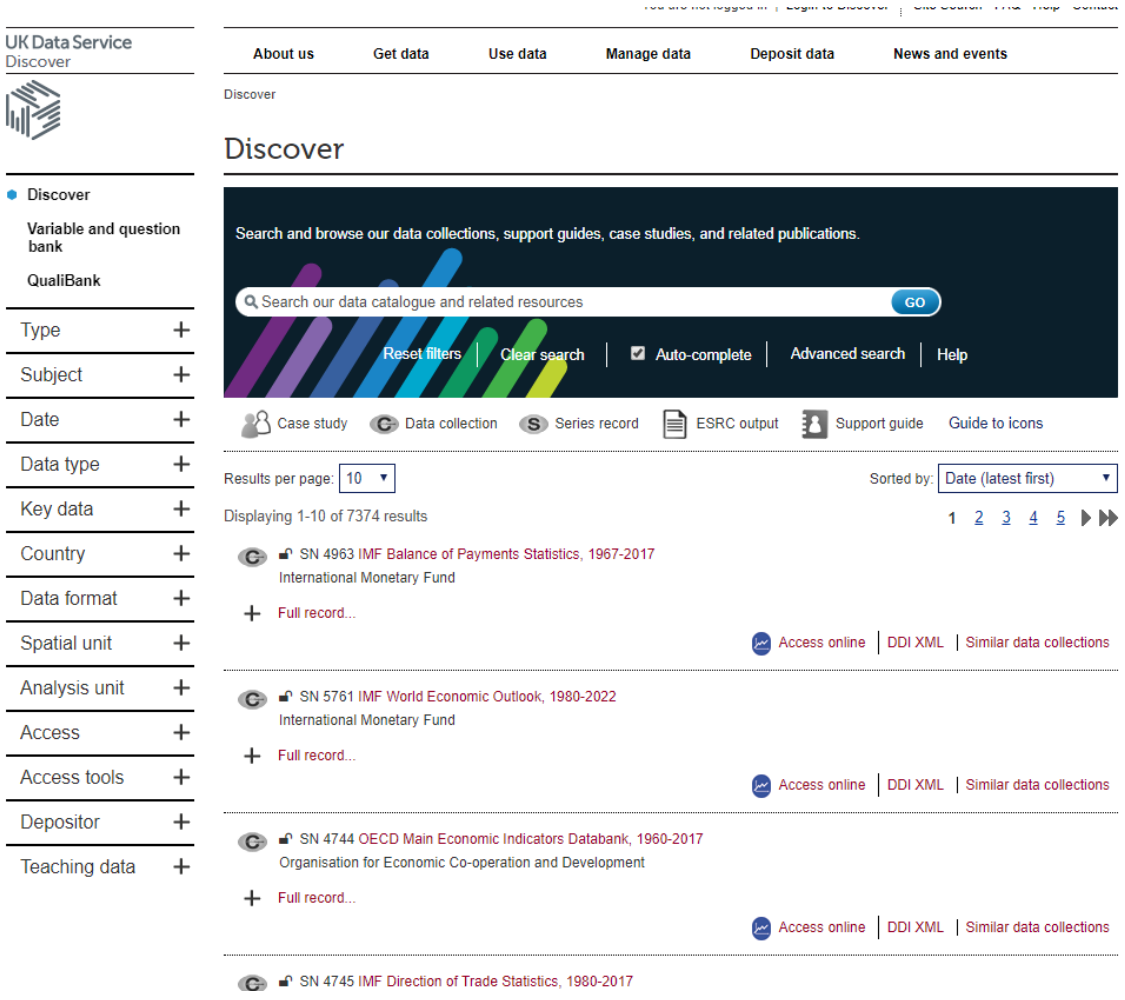
Before undertaking data collection always check…

- Is there evidence that secondary sources of data have been considered and evaluated?
- Is there evidence presented that the project is not creating new data when there are existing resources that could be re-used?
- If existing data are used, have issues such as copyright or IPR of such data been considered and possible copyright clearance obtained to be able to share data or data derived thereof?

UK Data Service Discover

- One-stop-shop for social science data

- https://discover.ukdataservice.ac.uk/



University of Essex

UK · DATA ARCHIVE

Consent

University of Essex

UK · DATA ARCHIVE

Obtaining consent from participants

- It is important to gain informed consent from participants

- Consent should be:
  - freely given
  - informed
  - unambiguous
  - specific
  - clear affirmative action

In practice: wording in consent forms / information sheets

Consent Form for [name of project]

Please tick the appropriate boxes	Yes	No
Taking Part		
I have read and understood the project information sheet dated DD/MM/YYYY and have been given the opportunity to ask questions about the project.	☐	☐
I agree to take part in the project. Taking part in the project will include being interviewed and recorded (audio or video).[1]	☐	☐
I understand that my taking part is voluntary; I can withdraw from the study at any time and I do not have to give any reasons for why I no longer want to take part.	☐	☐
Use of the information I provide		
I understand my personal details such as phone number and address will not be revealed to people outside the project.	☐	☐
I understand that my words may be quoted in publications, reports, web pages, and other research outputs.	☐	☐
I understand that the researcher's supervisors will have access to [anonymised] data.	☐	☐
*Please choose **one** of the following two options:*		
I would like my real name used in the above	☐	
I would **not** like my real name to be used in the above.	☐	
So we can use the information you provide legally		
I agree to assign the copyright I hold in any materials related to this project to [name of researcher].	☐	☐

University of Essex

UK·DATA ARCHIVE

In practice: wording in consent forms / information sheets

Any personal information that could identify you will be removed or changed before files are shared with research supervisors or results are made public.

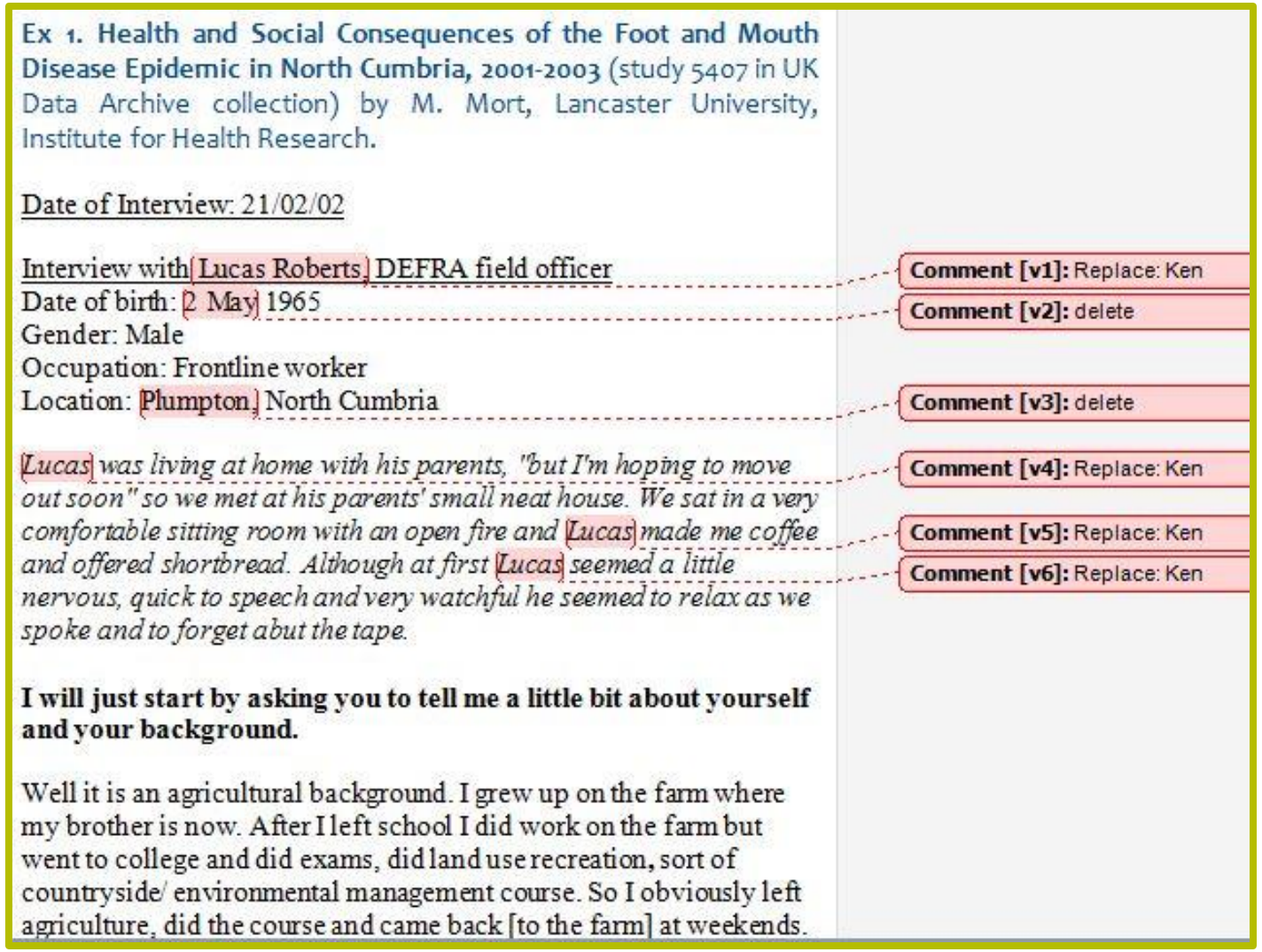
We ask you to consider the following points **before** agreeing to participate.

- Your contribution to the research will take the form of a focus group participant. This will be digitally video recorded and transcribed.
- **Your name and any information which may directly or indirectly identify you will be altered to protect your anonymity**.
- Any recordings of the discussions will be kept securely, and only authorised to other researchers on the condition they preserve your anonymity.
- The transcriptions (*excluding* names and other identifying details) will be retained by the researcher and analysed as part of the study. They will also be deposited with the UK Data Archive which has strict regulations about accessing data for research and protecting participant confidentiality.

ukdataservice.ac.uk/manage-data/legal-ethical/consent-data-sharing/consent-forms.aspx

Anonymisation

In practice: example anonymisation



Ex 1. Health and Social Consequences of the Foot and Mouth Disease Epidemic in North Cumbria, 2001-2003 (study 5407 in UK Data Archive collection) by M. Mort, Lancaster University, Institute for Health Research.

Date of Interview: 21/02/02

Interview with Lucas Roberts, DEFRA field officer
Date of birth: 2 May 1965
Gender: Male
Occupation: Frontline worker
Location: Plumpton, North Cumbria

Lucas was living at home with his parents, "but I'm hoping to move out soon" so we met at his parents' small neat house. We sat in a very comfortable sitting room with an open fire and Lucas made me coffee and offered shortbread. Although at first Lucas seemed a little nervous, quick to speech and very watchful he seemed to relax as we spoke and to forget abut the tape.

I will just start by asking you to tell me a little bit about yourself and your background.

Well it is an agricultural background. I grew up on the farm where my brother is now. After I left school I did work on the farm but went to college and did exams, did land use recreation, sort of countryside/ environmental management course. So I obviously left agriculture, did the course and came back [to the farm] at weekends.

Comment [v1]: Replace: Ken
Comment [v2]: delete
Comment [v3]: delete
Comment [v4]: Replace: Ken
Comment [v5]: Replace: Ken
Comment [v6]: Replace: Ken

University of Essex

UK · DATA ARCHIVE

Anonymising qualitative data

- plan or apply editing at time of transcription
- avoid blanking out; use pseudonyms or replacements
- identify replacements, e.g. with [brackets]
- avoid over-anonymising – removing / aggregating information in text can distort data or make it misleading
- consider keeping an anonymisation log of all replacements, aggregations or removals made and keep it *separate* from anonymised data files*

Anonymising quantitative data

- remove direct identifiers
  - e.g. names, address, institution and photos
- reduce the precision / detail of a variable through aggregation
  - e.g. birth year instead of date of birth; occupational categories rather than job; and, area rather than village
- generalise meaning of detailed text variable
  - e.g. occupational expertise
- restrict upper lower ranges of a variable to hide outliers
  - e.g. income and age
- combining variables
  - e.g. creating non-disclosive rural / urban variable from place variables

Audio-visual data

Digital manipulation of audio and image files can remove personal identifiers

e.g. voice alteration and image blurring (e.g. of faces)

Labour intensive, expensive, may damage research potential of data

Better alternatives:

- obtain consent to use and share data unaltered for research purposes
- avoid mentioning disclosive information during audio recordings

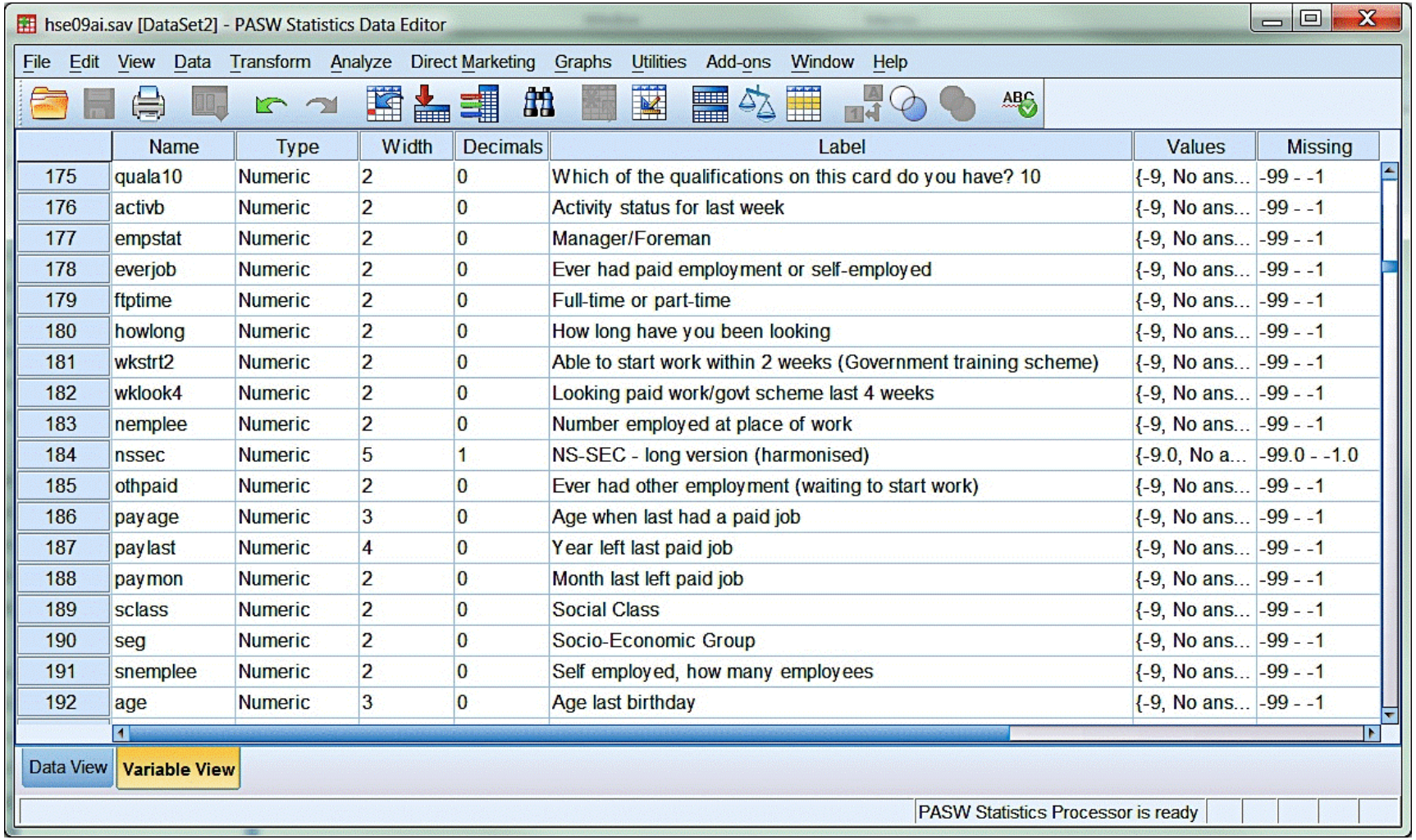
Documentation

- Enables you to understand data when you return to it
- If someone else was looking at your data for the first time, what would they need to know to make sense of it?
- Different documentation may be needed for data-level and project-level
- Every project will have different kinds of documentation

Include as documentation

- Data collection methodology and processes: sampling methods, sampling size, fieldwork protocol and interviewer instructions
- Information sheet / consent form
- Fieldwork tools: questionnaire, showcards and interview schedule
- Data list: overview of key information about each interview, as ‘at-a-glance’ summary of the data collection
- Analysis tools: codebook, memos, variable listing

Embedded metadata for a quantitative study in SPSS file

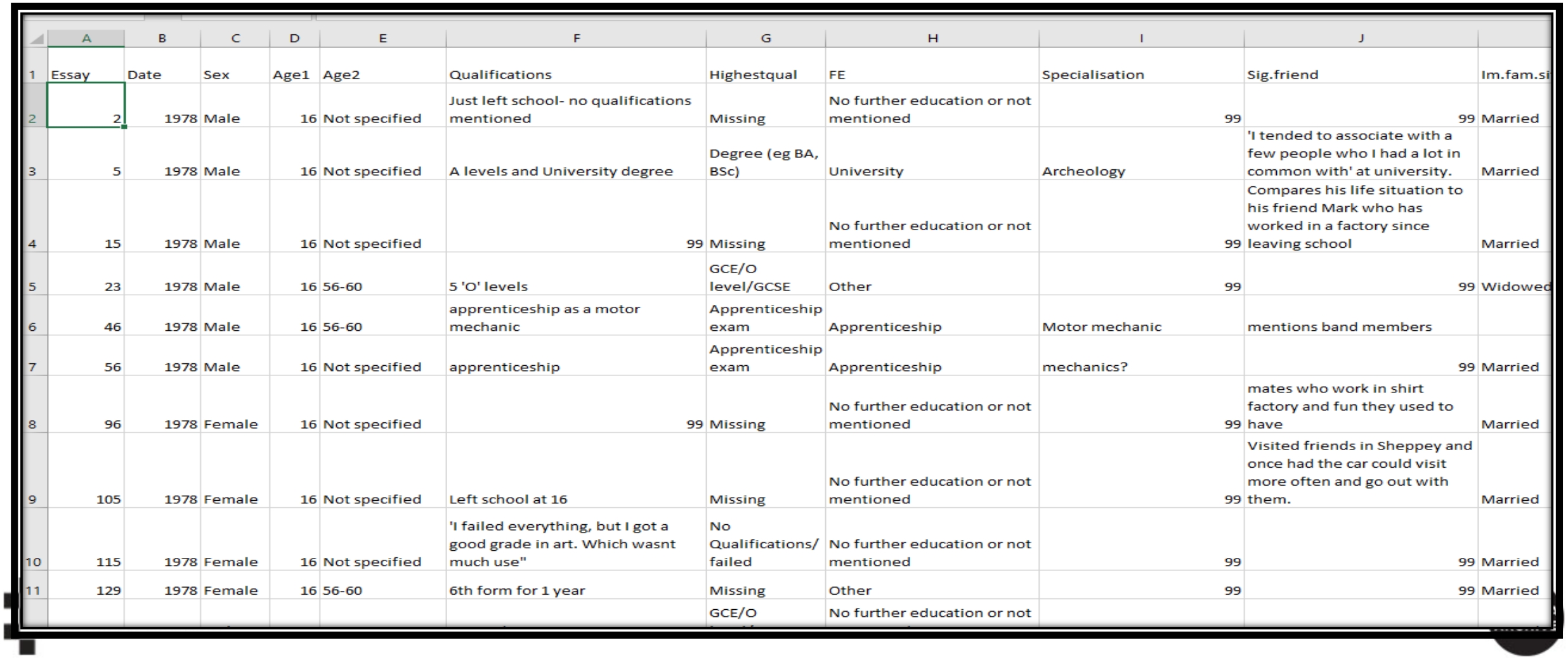


	Name	Type	Width	Decimals	Label	Values	Missing
175	quala10	Numeric	2	0	Which of the qualifications on this card do you have? 10	{-9, No ans...	-99 - -1
176	activb	Numeric	2	0	Activity status for last week	{-9, No ans...	-99 - -1
177	empstat	Numeric	2	0	Manager/Foreman	{-9, No ans...	-99 - -1
178	everjob	Numeric	2	0	Ever had paid employment or self-employed	{-9, No ans...	-99 - -1
179	ftptime	Numeric	2	0	Full-time or part-time	{-9, No ans...	-99 - -1
180	howlong	Numeric	2	0	How long have you been looking	{-9, No ans...	-99 - -1
181	wkstrt2	Numeric	2	0	Able to start work within 2 weeks (Government training scheme)	{-9, No ans...	-99 - -1
182	wklook4	Numeric	2	0	Looking paid work/govt scheme last 4 weeks	{-9, No ans...	-99 - -1
183	nemplee	Numeric	2	0	Number employed at place of work	{-9, No ans...	-99 - -1
184	nssec	Numeric	5	1	NS-SEC - long version (harmonised)	{-9.0, No a...	-99.0 - -1.0
185	othpaid	Numeric	2	0	Ever had other employment (waiting to start work)	{-9, No ans...	-99 - -1
186	payage	Numeric	3	0	Age when last had a paid job	{-9, No ans...	-99 - -1
187	paylast	Numeric	4	0	Year left last paid job	{-9, No ans...	-99 - -1
188	paymon	Numeric	2	0	Month last left paid job	{-9, No ans...	-99 - -1
189	sclass	Numeric	2	0	Social Class	{-9, No ans...	-99 - -1
190	seg	Numeric	2	0	Socio-Economic Group	{-9, No ans...	-99 - -1
191	snemplee	Numeric	2	0	Self employed, how many employees	{-9, No ans...	-99 - -1
192	age	Numeric	3	0	Age last birthday	{-9, No ans...	-99 - -1

University of Essex

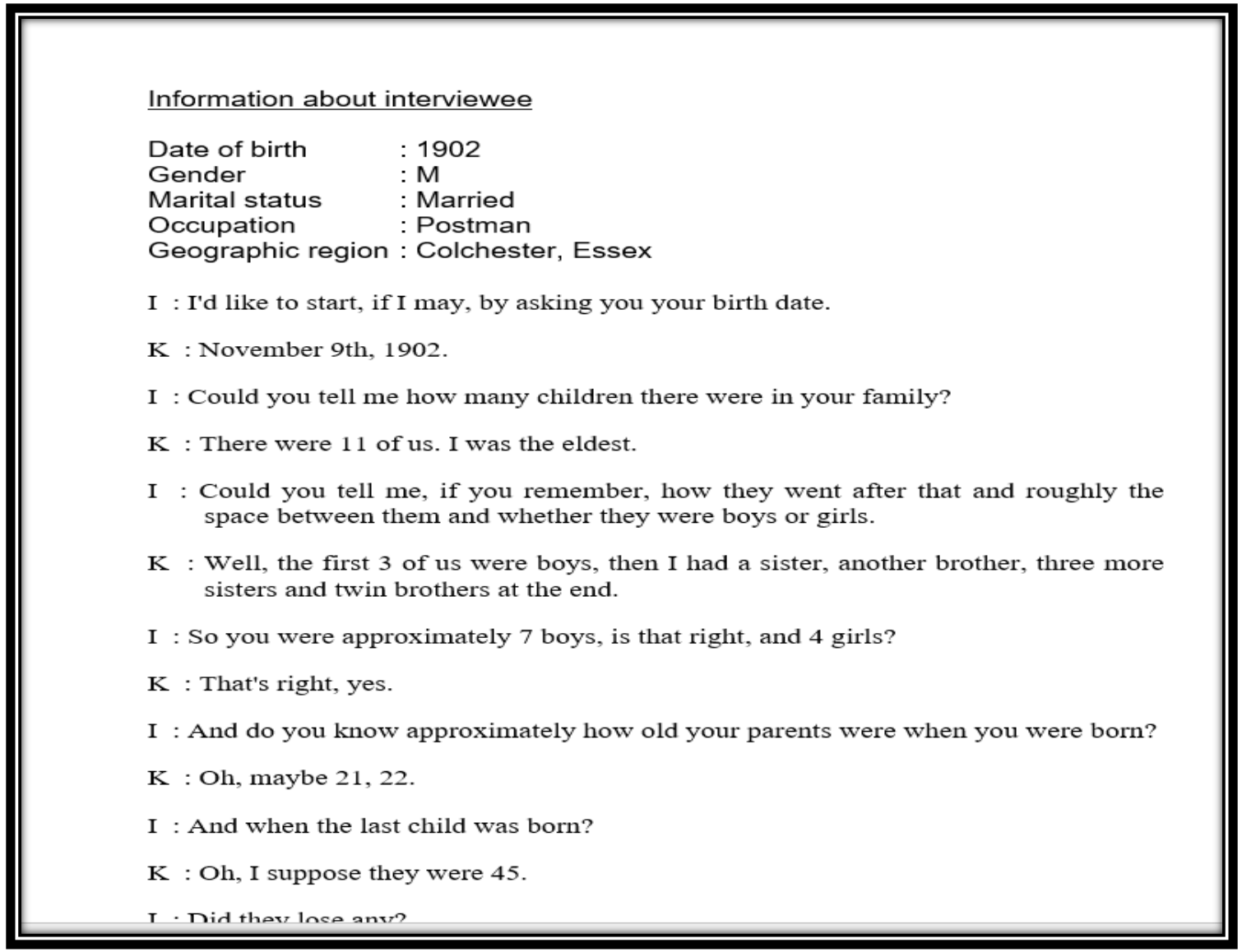
UK · DATA ARCHIVE

Documenting metadata for a qualitative study in Excel



	A	B	C	D	E	F	G	H	I	J	
1	Essay	Date	Sex	Age1	Age2	Qualifications	Highestqual	FE	Specialisation	Sig.friend	Im.fam.si
2	2	1978	Male	16	Not specified	Just left school- no qualifications mentioned	Missing	No further education or not mentioned	99	99	Married
3	5	1978	Male	16	Not specified	A levels and University degree	Degree (eg BA, BSc)	University	Archeology	'I tended to associate with a few people who I had a lot in common with' at university.	Married
4	15	1978	Male	16	Not specified	99	Missing	No further education or not mentioned	99	Compares his life situation to his friend Mark who has worked in a factory since leaving school	Married
5	23	1978	Male	16	56-60	5 'O' levels	GCE/O level/GCSE	Other	99	99	Widowed
6	46	1978	Male	16	56-60	apprenticeship as a motor mechanic	Apprenticeship exam	Apprenticeship	Motor mechanic	mentions band members	
7	56	1978	Male	16	Not specified	apprenticeship	Apprenticeship exam	Apprenticeship	mechanics?	99	Married
8	96	1978	Female	16	Not specified	99	Missing	No further education or not mentioned	99	mates who work in shirt factory and fun they used to have	Married
9	105	1978	Female	16	Not specified	Left school at 16	Missing	No further education or not mentioned	99	Visited friends in Sheppey and once had the car could visit more often and go out with them.	Married
10	115	1978	Female	16	Not specified	'I failed everything, but I got a good grade in art. Which wasnt much use"	No Qualifications/ failed	No further education or not mentioned	99	99	Married
11	129	1978	Female	16	56-60	6th form for 1 year	Missing	Other	99	99	Married
							GCE/O	No further education or not			

Documenting metadata on interviews



Information about interviewee

Date of birth : 1902
Gender : M
Marital status : Married
Occupation : Postman
Geographic region : Colchester, Essex

I : I'd like to start, if I may, by asking you your birth date.

K : November 9th, 1902.

I : Could you tell me how many children there were in your family?

K : There were 11 of us. I was the eldest.

I : Could you tell me, if you remember, how they went after that and roughly the space between them and whether they were boys or girls.

K : Well, the first 3 of us were boys, then I had a sister, another brother, three more sisters and twin brothers at the end.

I : So you were approximately 7 boys, is that right, and 4 girls?

K : That's right, yes.

I : And do you know approximately how old your parents were when you were born?

K : Oh, maybe 21, 22.

I : And when the last child was born?

K : Oh, I suppose they were 45.

I : Did they lose any?

University of Essex

UK · DATA ARCHIVE

Transcription template

Should:

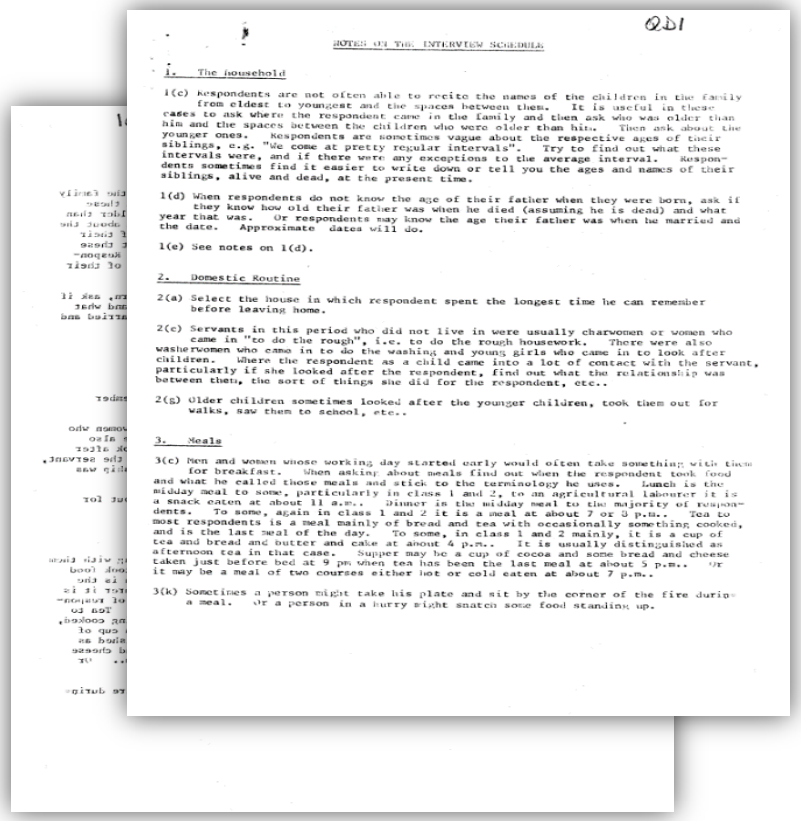
- possess a unique identifier
- adopt a uniform layout throughout the research project
- make use of speaker tags - turn-taking
- carry line breaks
- be page numbered
- carry a document header giving brief details of the interview: date, place, interviewer name, interviewee details, etc.

Other considerations:

- cover page or header
- compatibility with import features of Computer Assisted Qualitative Data Analysis Software (CAQDAS)

In practice: user guide and documentation

- A user guide could contain a variety of documents that provide context: interview schedule, transcription notes, even photos



QD1

NOTES ON THE INTERVIEW SCHEDULE

1. The household

1(c) Respondents are not often able to recite the names of the children in the family from eldest to youngest and the spaces between them. It is useful in these cases to ask where the respondent came in the family and then ask who was older than him and the spaces between the children who were older than him. Then ask about the younger ones. Respondents are sometimes vague about the respective ages of their siblings, e.g. "We come at pretty regular intervals". Try to find out what these intervals were, and if there were any exceptions to the average interval. Respondents sometimes find it easier to write down or tell you the ages and names of their siblings, alive and dead, at the present time.

1(d) When respondents do not know the age of their father when they were born, ask if they know how old their father was when he died (assuming he is dead) and what year that was. Or respondents may know the age their father was when he married and the date. Approximate dates will do.

1(e) See notes on 1(d).

2. Domestic Routine

2(a) Select the house in which respondent spent the longest time he can remember before leaving home.

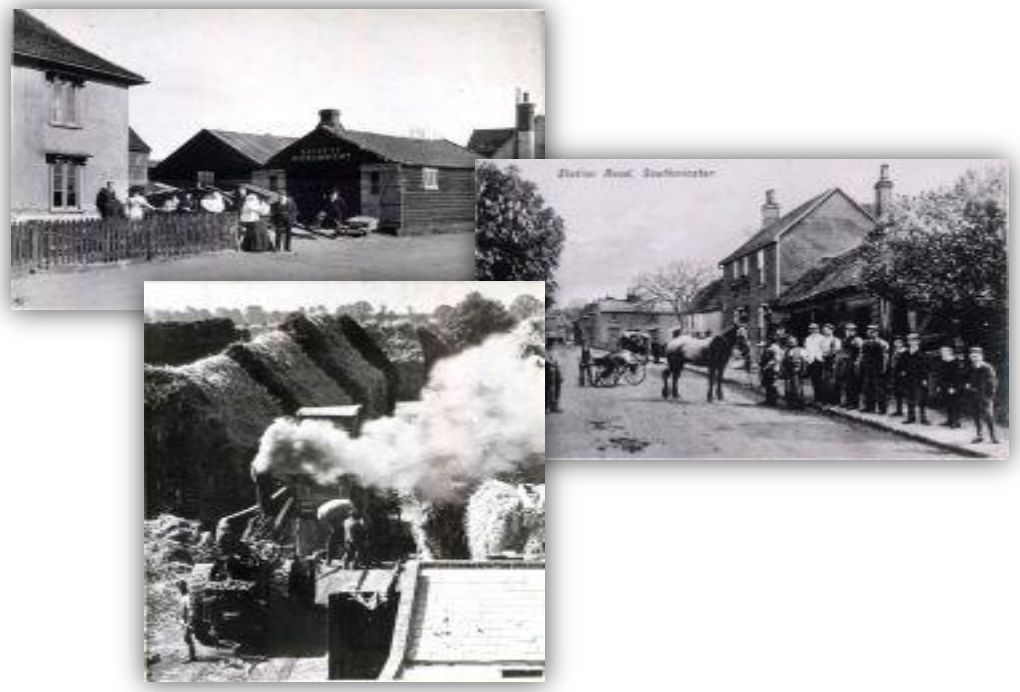
2(c) Servants in this period who did not live in were usually charwomen or women who came in "to do the rough", i.e. to do the rough housework. There were also washerwomen who came in to do the washing and young girls who came in to look after children. Where the respondent as a child came into a lot of contact with the servant, particularly if she looked after the respondent, find out what the relationship was between them, the sort of things she did for the respondent, etc..

2(g) Older children sometimes looked after the younger children, took them out for walks, saw them to school, etc..

3. Meals

3(c) Men and women whose working day started early would often take something with them for breakfast. When asking about meals find out when the respondent took food and what he called those meals and stick to the terminology he uses. Lunch is the midday meal to some, particularly in class 1 and 2, to an agricultural labourer it is a snack eaten at about 11 a.m.. Dinner is the midday meal to the majority of respondents. To some, again in class 1 and 2 it is a meal at about 7 or 8 p.m.. Tea to most respondents is a meal mainly of bread and tea with occasionally something cooked, and is the last meal of the day. To some, in class 1 and 2 mainly, it is a cup of tea and bread and butter and cake at about 4 p.m.. It is usually distinguished as afternoon tea in that case. Supper may be a cup of cocoa and some bread and cheese taken just before bed at 9 pm when tea has been the last meal at about 5 p.m.. Or it may be a meal of two courses either hot or cold eaten at about 7 p.m..

3(k) Sometimes a person might take his plate and sit by the corner of the fire during a meal. Or a person in a hurry might snatch some food standing up.



University of Essex

UK·DATA ARCHIVE

In practice: data list

- Data listing provides an at-a-glance summary of interview sets

Study Number 5407
Health and Social Consequences of the Foot and Mouth Disease Epidemic in North Cumbria, 2001
Mort, M.

The panel respondents for the study were divided into six population groups. The data list for the diary and interviews has been colour-coded accordingly for clarity, using the depositor's original colours:

Group 1: Farmers	Group 2: Rural Business	Group 3: Agricultural related occupations	Group 4: Frontline Workers	Group 5: Community	Group 6: Animal / Human Health Professionals

1. Interviews

Respondent ID	Population Group	Date of Birth	Gender	Occupation	Interview summary	Place of Interview
PM02	Group 6: Animal / Human Health Professionals	1975	M	Veterinary Surgeon	Family and background,career and work, arrangements during FMD epidemic and perceptions of situation	North Cumbria, respo home
PM03	Group 6: Animal / Human Health Professionals	1966	F	Veterinary Surgeon	Family and background,career and work, arrangements during FMD epidemic and perceptions of situation	North Cumbria
PM07	Group 6: Animal / Human Health Professionals	1964	F	Veterinary practice manager	Family and background,career and work, arrangements during FMD epidemic and perceptions of situation	North Cumbria, respo home
					Family and	

University of Essex

UK • DATA ARCHIVE

Storing your data

File formats

Choice of software format for digital data:

- software availability / cost
- hardware used – e.g. audio capture
- discipline-specific standards and customs

Best formats for long-term preservation:

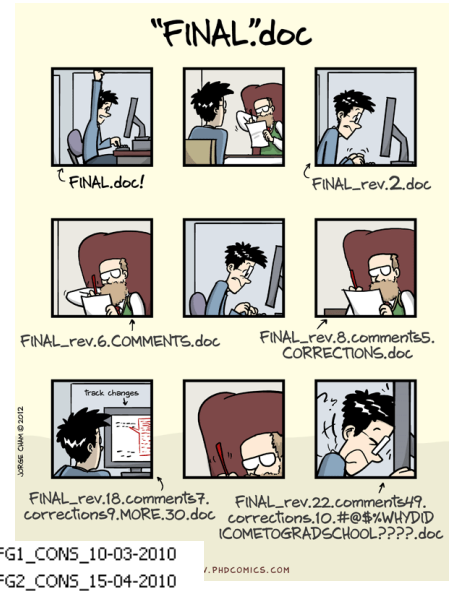
- standard, interchangeable and open
- *e.g. tab-delimited, comma-delimited (CSV), ASCII, RTF, PDF/A, OpenDocument format and XML*
- UK Data Service optimal file formats for various data types
- Digital Preservation Coalition guidance on preservation formats

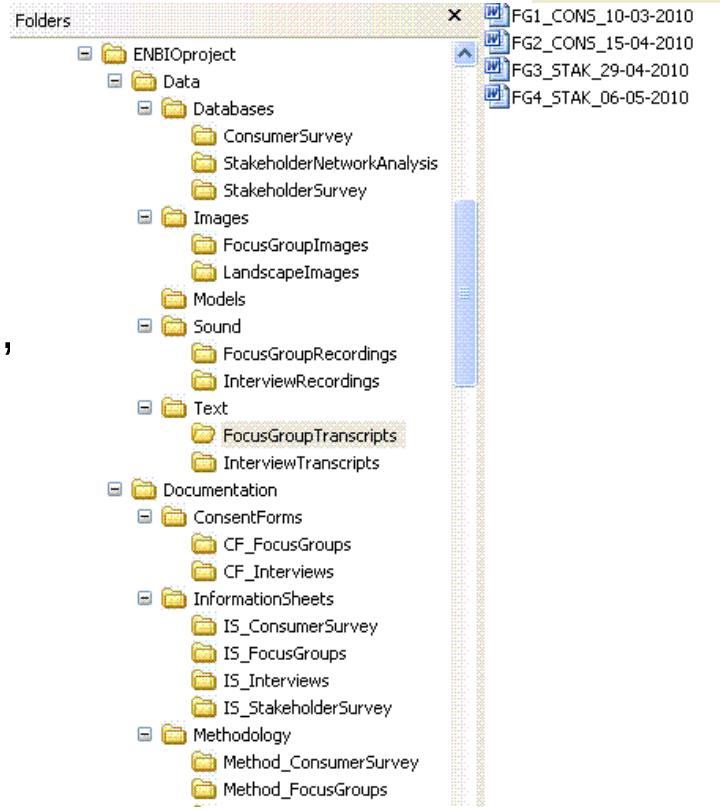
Organising data

- Plan in advance how best to organise data
- Use a logical structure
- Use logical names and version control e.g. V1.0, V2.1, ‘FINAL’
- 2018-01-30_Interview_01

Examples:

- Hierarchical structure of files, grouped in folders, e.g. audio, transcripts and annotated transcripts
- Survey data: spreadsheet, SPSS, relational database
- Interview transcripts: individual well named files





University of Essex

UK · DATA ARCHIVE

Data security and storage

Protect data from unauthorised:

- access
- use
- change
- disclosure
- destruction

Who knows who is watching, listening or attempting to access your data…

Stuff happens: data loss

- What would happen if you lost your data?
- Imagine if you left your bag on a train, containing your laptop (with all your digital research notes on) and your paper based notes too – this situation happened to Andrew Penson

- Source:
  https://twitter.com/ADPenson/status/883637257323896832

Stuff happens: data theft

- What would happen if you data was stolen?
- Imagine if seven years worth of your Ebola research was stolen – this situation happened to Dr Fitzgerald

- Source:

  https://www.standard.co.uk/news/crime/burglar-stole-laptop-with-seven-years-of-ebola-research-from-doctor-s-house-a3689406.html

Stuff happens: data theft

- Imagine if you lost four years worth of research data – this situation happened to Billy Hinchen.
- From Figshare (2014): “Biologist Billy Hinchen returned one afternoon to find that his laptop and all backup hard drives had been stolen. All that remained was a disparate collection of data, spread across numerous flash drives, email attachments and scribbled drawings.”

https://www.youtube.com/watch?v=3xlax_Iin0Y

- Source:
  https://figshare.com/blog/The_stuff_of_nightmares_imagine_losing_all_your_research_data/121

Digital back-up strategy

Consider:

- **What's backed-up?** - all, some or just the bits you change?
- **Where?** - original copy, external local and remote copies
- **What media?** - DVD, external hard drive, USB, Cloud?
- **How often?** - hourly, daily, weekly? Automate the process?
- **How many copies?** - minimum of three copies!
- **What method/software?** - duplicating, syncing or mirroring?
- **For how long is it kept?** - data retention policies that might apply?
- **Verify and recover** - never assume, regularly test and restore

Backing-up need not be expensive

- 2Tb external drives are around £70, with back-up software

Also consider non-digital storage options too!

Data security strategy

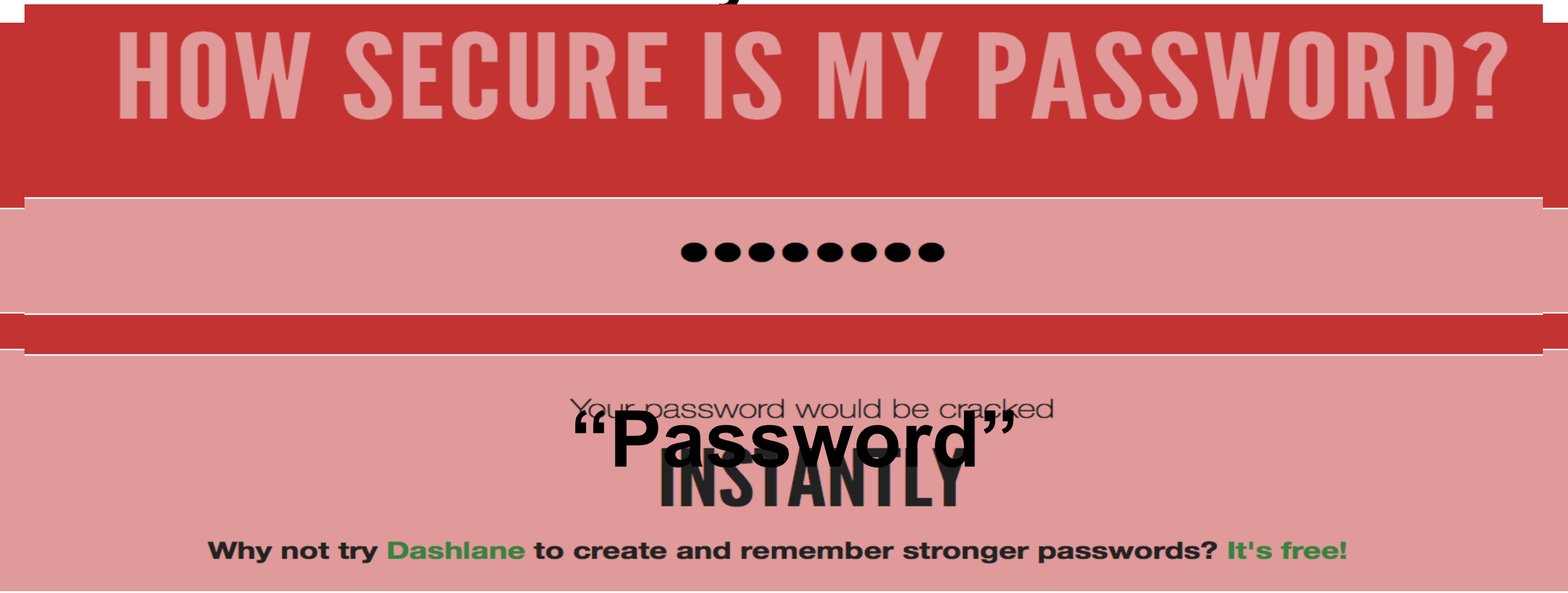
- Control access to computers:
  - use passwords and lock your machine when away from it
  - run up-to-date anti-virus and firewall protection
  - power surge protection
  - utilise encryption
    - on all devices: desktops, laptops, memory sticks and mobile devices
    - at all locations: work, home and travel
  - restrict access to sensitive materials e.g. consent forms and patient records
  - personal data need more protection – always keep them separate and secure
- Control physical access to buildings, rooms and filing cabinets
- Properly dispose of data and equipment once the project is finished

Passwords

- Strong passwords are crucial
- Avoid using weak or easy to guess passwords and reusing passwords
- Consider password managers, complex passwords or stringing words together to create stronger passwords
- But, remember that you need to be able to remember the passwords!
- **Why does this matter?**
- No matter how good the encryption is that you use if you use a weak password the encryption will offer little protection

https://howsecureismypassword.net (*Never use real passwords)

Password security

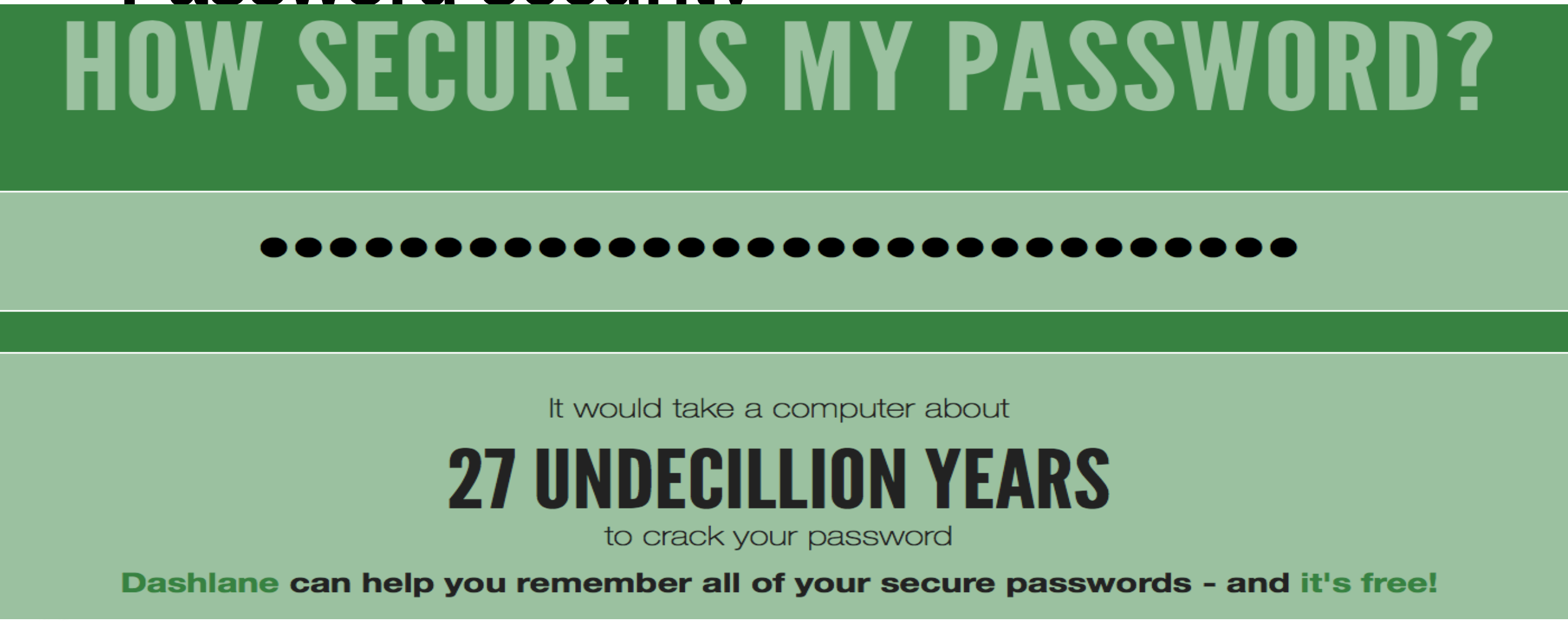


"Password"

University of Essex

UK • DATA ARCHIVE

Password security



Encryption

Encryption is the process of encoding digital information in such a way that only authorised parties can view it

Encryption software can be easy to use and offers the ability to: encrypt hard drives, partitions, files, folders and portable storage devices such as USB flash drives

[VeraCrypt](#)

[BitLocker](#)

[Axcrypt](#)

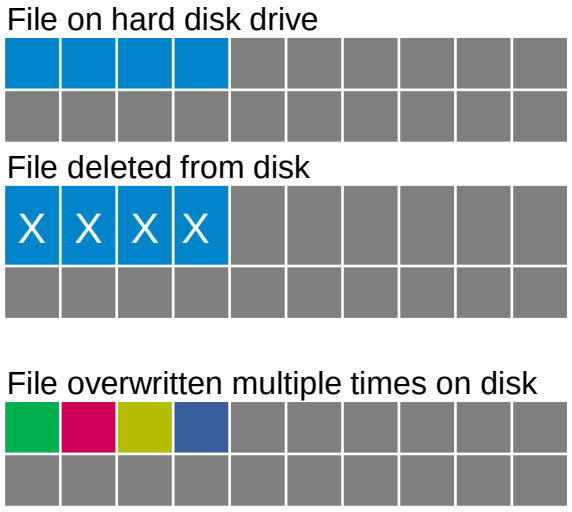
[FileVault2](#)

University of Essex

UK • DATA ARCHIVE

Data disposal

- When you delete a file from a hard drive, it is likely to still be retrievable (even after emptying the recycle bin)



- Even reformatting a hard drive is ***not*** sufficient

- Files need to be overwritten multiple times
  with random data for best chances of removal

- The **only** sure way to ensure data is irretrievable
  is to physically destroy the drive (using an approved secure destruction facility)

University of Essex

UK • DATA ARCHIVE

File sharing and cloud storage services

Think about the best way to share data (personal/sensitive) between yourself and your supervisor(s)

- Too often sent as insecure email attachments
- Physical media?

- Online or ‘cloud’ services are becoming increasingly popular (e.g. Google Drive, DropBox, Microsoft OneDrive and iCloud)
- Benefits:
  - Very convenient
  - Accessible anywhere
  - Background file syncing
  - Mirrors files
  - Mobile apps available

But,

  - These are not necessarily secure
  - Potential DPA issues
  - Limited control over where data is stored
  - Not necessarily permanent
  - Intellectual property right concerns?

Conclusion

- Documentation gaining new importance: transparency has become a new revolution in qualitative research
  - Obligation to share details about data, methods and theory – specified in academic misconduct policies
  - Better, more complete documentation = more transparency
- Legal and ethical obligation to protect data
  - Protects participants, protects you
  - “Protection” is not as simple as a password

Our data management guidance

- online best practice guidance: ukdataservice.ac.uk/manage-data.aspx
- Managing and Sharing Research Data – a Guide to Good Practice: (Sage Publications Ltd)
- helpdesk for queries: ukdataservice.ac.uk/help/get-in-touch.aspx
- training: www.ukdataservice.ac.uk/news-and-events/events

University of Essex

UK • DATA ARCHIVE

Tools & templates

- Model consent form: http://www.data-archive.ac.uk/media/112638/ukdamodelconsent.pdf
- Survey consent statement: http://data-archive.ac.uk/media/147338/ukdasurveyconsent.doc
- Transcription template: http://data-archive.ac.uk/media/136055/ukdamodeltranscript.pdf
- Transcription instructions: http://data-archive.ac.uk/media/285633/ukda-example-transcription-instructions.pdf
- Transcription confidentiality agreement: http://data-archive.ac.uk/media/285636/ukda-transcriber-confidentiality-agreement.pdf
- Data list template: http://data-archive.ac.uk/media/2989/UK%20Data%20Archive%20Example%20Data%20List.pdf
- RDM costing tool: www.data-archive.ac.uk/media/247429/costingtool.pdf
- Encryption tutorials: https://www.youtube.com/watch?v=y4Iosu-Yfsw&list=PLG87Imnep1SmnFGhAjFVHonQSVmMlpHkV

Keep connected

- Subscribe to UK Data Service list: www.jiscmail.ac.uk/cgi-bin/webadmin?A0=UKDATASERVICE
- Follow UK Data Service on Twitter: @UKDataService
- Facebook
- Youtube: www.youtube.com/user/UKDATASERVICE

Questions?

UK Data Archive
University of Essex
ukdataservice.ac.uk/help/get-in-touch.aspx
Scott Summers – ssummers@essex.ac.uk
Maureen Haaker – mahaak@essex.ac.uk

University of Essex

UK • DATA ARCHIVE