

# CESSDA

---

## Data Management Expert Guide: Protect Chapter

*Dr Scott Summers*

*UK Data Service*

*University of Essex*

*12-13th April 2018 Train the Trainers Workshop*



 [cessda.eu](http://cessda.eu)

 [@CESSDA\\_Data](https://twitter.com/CESSDA_Data)

# Citation

Please cite this presentation as:

Scott, S. (2020). The CESSDA Data Management Expert Guide. Chapter 5. Protect [presentation]. Bergen: CESSDA ERIC

# Disclaimer

- ◇ The information in this presentation is based on our current interpretation of the legislation and its implications for research and the archiving of research data
- ◇ This is a very fluid area and thus changes are still possible
- ◇ This presentation **does not constitute, or should not be construed as, legal advice and / or guidance**

# Overview

*This part of the tour guide focuses on key legal and ethical considerations in creating shareable data*

Areas to be covered:

- ◊ Ethical Review
- ◊ Data Protection
- ◊ Informed Consent
- ◊ Anonymisation
- ◊ Copyright

# Ethical Review



# Ethical Review Process

- ◊ Ethical review is about helping you as a researcher to think through the ethical issues surrounding your research
- ◊ The principles of good research practice encourage you to consider the wider consequences of your research and engage with the interest of your participants
- ◊ Ethics review by a Research Ethics Committee (REC) is typically required when (sensitive) personal data are being collected
- ◊ The role of a REC is to protect the safety, rights and well-being of research participants and to promote ethically sound research
- ◊ Among other duties, this involves ensuring that research complies with national and international data protection laws regarding the use of personal information collected in research



# Ethical Review Process

- ◇ Performing an ethical self-assessment:
  - ◇ Question 1: The project's aim and method
  - ◇ Question 2: Research involving identifiable persons
  - ◇ Question 3: Whistle-blowing



# Ethical Arguments for Archiving Data

- ◊ Not burden over-researched, vulnerable groups
- ◊ Make best use of hard-to-obtain data, e.g. elites, socially excluded, over-researched
- ◊ Extend voices of participants
- ◊ Provide greater research transparency

In each, ethical duties to participants, peers and public may be present



# Data Protection



# The General Data Protection Regulation (GDPR)

- ◊ The GDPR applies from the 25 May 2018
- ◊ The GDPR applies to any EU researcher or researcher in the European Economic Area (EEA) who collects personal data and any researcher worldwide who collects personal data on EU citizens. to any data controller or data processor in the EU who collects personal data about a data subject of any country, anywhere in the world
- ◊ A data controller or data processor that is based outside the EU but collects personal data on EU citizens will also be covered by the GDPR
- ◊ This means that a researcher (data controller) based within the EU who collects personal data about a participant, from any other country within the EU, or the world, needs to comply with the GDPR
- ◊ Also means a researcher (data controller) outside the EU who collects personal data about a participant in the EU will be covered when this relates to offering goods/services or the monitoring of their behaviour within the EU

# The GDPR applies to...

- ◊ The GDPR applies only to 'personal data' and data of 'living persons'
- ◊ Data which do not count as personal data do not fall under data protection legislation
- ◊ Though there may still be ethical reasons for wanting to protect this information!



# Principles Relating to Processing of Personal Data

## 1) Process lawfully, fair and transparent

- ◊ The participant is informed of what will be done with the data and data processing should be done accordingly

## 2) Keep to the original purpose

- ◊ Data should be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes.  
\*Note Article 5(1)(b), Article 89\*

## 3) Minimise data size

- ◊ Personal data that are collected should be adequate, relevant and limited to what is necessary

# Principles Relating to Processing of Personal Data

## 4) Uphold accuracy

- ◆ Personal data should be accurate and, where necessary kept up to date. Every reasonable step must be taken to ensure that personal data that are inaccurate are erased or rectified without delay

## 5) Remove data which are not used

- ◆ Personal data should be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed. \* Note Article 5(1)(e), Article 89 \*

## 6) Ensure data integrity and confidentiality

- ◆ Personal data are processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures

# GDPR Archiving and Research Exemption

*Processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, shall be subject to appropriate safeguards, in accordance with this Regulation, for the rights and freedoms of the data subject. Those safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner*

- ◊ Principle 2. and 5. are less strict
- ◊ 2. Purpose: further processing allowed
- ◊ 5. Personal data may be stored for longer periods

# The Grounds for Processing Personal Data

- ◇ There are 6 grounds for the processing of personal data, and one of these must be present in order to process a data subject's personal data:
  1. Consent of the data subject
  2. Necessary for the performance of a contract
  3. Legal obligation placed upon controller
  4. Necessary to protect the vital interests of the data subject
  5. Carried out in the public interest or is in the exercise of official authority
  6. Legitimate interest pursued by controller

# GDPR - Data Subject Rights

- ◊ The right to be informed
- ◊ The right of access
- ◊ The right to rectification
- ◊ The right to erasure - the 'right to be forgotten'
- ◊ The right to restrict processing
- ◊ The right to data portability
- ◊ The right to object
- ◊ Rights in relation to automated individual decision-making and profiling





# Processing Personal Data in Research

In research, the three most applicable bases for processing personal data:

- ◇ Consent of the data subject
  - ◇ [for example\*: A oral history project where people's real names are used]
- ◇ Public interest (public task)
  - ◇ [for example\*: A longitudinal study of people living with dementia and their carers, to identify how people would like to be supported. Findings inform and support the caring strategy and public advocacy]
- ◇ Legitimate interest
  - ◇ [for example\*: A research project that is fully funded and is being undertaken by a private corporation to look at the effects of smoking on car passengers]

**For each research project, if personal data will be collected and processed, the most appropriate legal basis needs to be decided and recorded (and should not be changed at a later date)**

- ◇ \*The examples are provided by UK Data Service, which has published examples of where a legal basis may be applied in research

# Research and the GDPR in Practice

If you will collect personal data, then:

- ◊ Determine who will be the data controller (possibly your institution)
- ◊ Decide which legal basis will apply
- ◊ If collaborative partners need access to personal data, then make sure agreements are in place
- ◊ Consider whether a Data Protection Impact Assessment (DPIA) is needed
- ◊ Communicate to research participants how personal data collected about them will be used, stored, processed, transferred, who the data controller is (with their contact details), the legal ground and purpose of the processing, the period of retention and their rights; this can be done via an information sheet or a webpage (e.g. privacy notice)
- ◊ Consider where to store personal data securely
- ◊ Minimise the personal data to collect and pseudonymise where possible

# Data Protection Impact Assessment

A DPIA is required for data processing that is likely to result in a high risk to the rights and freedoms of individuals. In practice this means if at least two of these criteria apply (examples can be found in the Data Protection Working Party 248 guidelines):

- ◊ Evaluation or scoring
- ◊ Automated-decision making with legal or similar significant effect
- ◊ Systematic monitoring
- ◊ Sensitive data
- ◊ Data processed on a large scale
- ◊ Datasets that have been matched or combined
- ◊ Data concerning vulnerable data subjects
- ◊ Innovative use or applying technological or organisational solutions
- ◊ Data transfer across borders outside the European Union
- ◊ When the processing in itself prevents data subjects from exercising a right or using a service or a contract

# How to Implement a DPIA across Different Institutions, for Research

If research is done as a collaboration of more than one institution, with shared responsibilities, one DPIA done by one of the institutions should be sufficient, and the other partner institutions can apply that same DPIA

Problems might arise when research involves institutions that are implementing a DPIA in different countries, whereby policies or *requirements may vary across those countries*, such as for data security, ownership of the data, different understandings on gaining consent and which legal basis to use for processing personal data

# The GDPR and...(1)

## Open Data

- ◊ For personal data, the open access motto 'as open as possible, as closed as necessary' is important
- ◊ A political or societal drive for open access and open science does not mean that individual rights granted by legislation can be overruled. Therefore, for personal data, '*as closed as necessary*' is the key

## Social Media Data

Gaining consent would be the best approach when using social media data. So even for social media data in the public domain, researchers should ask the people whose social media content they mine for their consent when possible. In some cases public task may be used as a legal basis

# The GDPR and...(2)

## Administrative or Register Data

When data that contain personal information, if consent is not collected from the individuals when the administrative or register data are collected, then the most common legal basis for further use is public task

## Qualitative Data

Despite the use of pseudonyms, interviewed people may still be identifiable from the story they tell. This would constitute personal information. In this case, the legal ground could be consent, which would need to be sought from the study participants

Another aspect to keep in mind here is data collected which would allow identification of other people who may not have been asked for consent, for example people involved in the story told by the interviewed. So you may also be processing personal data from people who have not been asked for consent

In that case, the processing ground could be public interest and the argument would be that the research has value for society

# Diversity in Data Protection

- ◊ One of the key aims of the GDPR is to harmonise laws across the EU regarding data protection legislation
- ◊ In addition to the GDPR, each EU Member State has rules on data protection and legislation that you have to familiarise yourself with if you collect personal data
- ◊ Because of this, some Member States have more restrictive data protection legislation than others
- ◊ It is important researchers familiarise with the local laws, rules and ethical requirements for their projects
  - ◊ When research crosses legal and jurisdictional boundaries researchers should always seek to apply the requirements of the legislation that has the most stringent requirements of the whole project
  - ◊ Where this is unclear, you should obtain advice from your institute, ethical committees or qualified legal professionals



# Best Practice for Legal Compliance

- ◊ Investigate early which laws apply to your data
- ◊ Do not collect personal or sensitive data if not essential to your research
- ◊ Seek advice from your research office
- ◊ Plan early in research
  
- ◊ If you must deal with personal or sensitive data
  - ◊ Inform participants about how their data will be used
  - ◊ Remember: not all research data are personal (e.g. anonymised data are not personal)



# Strategy for Sharing Data

1. Obtain informed consent, also for data sharing and preservation or curation
2. Protect identities e.g. anonymisation and not collecting personal data
3. Regulate access where needed (all or part of data) e.g. by group, use or time period
4. Securely store personal or sensitive data

# Informed Consent



# Consent is Needed Across the Data Lifecycle

- ◇ Engagement in the research process
  - ◇ What activities are involved in participating in the project?
- ◇ Dissemination in presentations, publications, the web
  - ◇ Consent for use of quotes for articles and video publicity
- ◇ Data sharing and archiving
  - ◇ Consider future uses of data

\* Consent is always dependent on the research context – special cases of covert research and verbal consent

# Informed Consent (Broadly)

- ◆ Consent needs to be freely given, informed, unambiguous, specific and by a clear affirmative action that signifies agreement to the processing of personal data
- ◆ When special categories data are processed – and the processing grounds for this is consent – there is a further requirement to the above that this must be based on explicit consent



# Informed Consent – Data Sharing (1)

- ◇ Gaining informed consent for data sharing is seen as 'one more small step' to gaining consent from participants to partake in a research project
- ◇ Adding the discussion of data sharing and archiving permits the participant to make an informed decision. This empowers them and puts them in charge of choosing whether they wish for their contribution to the research project – and their data – to be available for use in future research projects

# Informed Consent – Data Sharing (2)

- ◊ The best way to achieve informed consent for data sharing is to **identify** and **explain the possible future uses of their data** and offer the participant the option to consent on a **granular level**
- ◊ For example, in a qualitative study, this may involve allowing the participant to consent to data sharing of the anonymised transcripts, the non-anonymised audio recordings and the photographs

## Informed Consent for [name of study]

Please tick the appropriate boxes

Yes No

### 1. Taking part in the study

I have read and understood the study information dated [DD/MM/YYYY], or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.  Yes  No

I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.  Yes  No

I understand that taking part in the study involves [.....].  Yes  No

Describe in a few words how information is captured, using the same terms as you used in the information sheet, for example: an audio-recorded interview, a video-recorded focus group, a survey questionnaire completed by the enumerator, an experiment, etc.].

For interviews, focus groups and observations, specify how the information is recorded (audio, video, written notes).

For questionnaires, specify whether participant or enumerator completes the form.  
For audio or video recordings, indicate whether these will be transcribed as text, and whether the recording will be destroyed.

If there is a potential risk of participating in the study, then provide an additional statement:

I understand that taking part in the study has [.....] as potential risk.  Yes  No

### 2. Use of the information in the study

I understand that information I provide will be used for [.....].  Yes  No

List the planned outputs, e.g. reports, publications, website, video channel etc., using the same terms as you used in the study information sheet.

Consider whether knowledge sharing and benefits sharing needs to be considered, e.g. for indigenous knowledge.

I understand that personal information collected about me that can identify me, such as my name or where I live, will not be shared beyond the study team.  Yes  No

At times this should be restricted to the researcher only.

### Potential additional statements

i) If you want to use quotes in research outputs: I agree that my information can be quoted in research outputs.  Yes  No

ii) If you want to use named quotes: I agree that my real name can be used for quotes.  Yes  No

iii) If written information is provided by the participant (e.g. diary): I agree to joint copyright of the [DD/MM/YYYY] to [name of researcher].  Yes  No

### 3. Future use and reuse of the information by others

I give permission for the [specify the data] that I provide to be deposited in [name of data repository] so it can be used for future research and learning.  Yes  No

Specify in which form the data will be deposited, e.g. anonymised transcripts, audio recording, survey database, etc.; and if needed repeat the statement for each form of data you plan to deposit.

Specify whether deposited data will be anonymised, and how. Make sure to describe this in detail in the information sheet.

Specify whether use or access restrictions will apply to the data in future, e.g. exclude commercial use, apply safeguarded access, etc.; and discuss these restrictions with the repository in advance.



# In Practice: Wording in Consent Forms / Information Sheets

The interviews will be archived at ..... and disseminated so other researchers can reuse this information for research and learning purposes:

- I agree for the audio recording of my interview to be archived and disseminated for reuse
- I agree for the transcript of my interview to be archived and disseminated for reuse
- I agree for any photographs of me taken during interview to be archived and disseminated for reuse

We expect to use your contributed information in various outputs, including a report and content for a website. Extracts of interviews and some photographs may both be used. We will get your permission before using a quote from you or a photograph of you. After the project has ended, we intend to archive the interviews at .... Then the interview data can be disseminated for reuse by other researchers, for research and learning purposes.



# Timing and Form of Consent

	Advantage	Disadvantage
Written consent	<p>More solid legal ground, e.g. participant has agreed to disclose confidential info</p> <p>Often required by Ethics Committees</p> <p>Offers more protection for researcher (as they have written documentation of consent)</p>	<p>Not possible for some cases: infirm, illegal activities</p> <p>May scare people from participating (or have them think that they cannot withdraw their consent)</p>
Verbal consent	<p>Best if recorded</p>	<p>Can be difficult to make all issues clear verbally</p> <p>Possibly greater risks for researcher (in regards to adequately proving participant consent)</p>

	Advantage	Disadvantage
<p><b>One-off consent:</b> participant is asked to consent to taking part in the research project only once.</p>	<p>Simple</p> <p>Least hassle to participants</p>	<p>Research outputs not known in advance</p> <p>Participants will not know all info they will contribute</p>
<p><b>Process consent:</b> participant's consent is requested continuously throughout the research project</p>	<p>Ensures 'active' consent</p>	<p>May not get all consent needed before losing contact</p> <p>Repetitive, can annoy participants</p>



# Documenting Consent

- ◊ Under the GDPR, consent needs to be documented, which means (in the context of research) it will be important for researchers to maintain documented and accurate records of the consent obtained from their participants
- ◊ This could, for example, be written consent or audio recorded oral consent
- ◊ Though the GDPR does not require this consent to be in a written form, many UK research ethics committees and professional bodies do require this or recommend it as best practice

# Consent Exercises

UK Data Service



## Exercise: Assessing Consent Form Statements

All of the statements below have appeared on real consent forms or are very minor modifications of statements that have been used. Please read each statement and consider the implications for data sharing for any data generated using this consent statement. Do you have any suggestions for alternative wording or other changes you would wish to make?

- All information gathered will be totally anonymised, dealt with in the strictest of confidence and used at an aggregated level.
- Any information I give will be used for research only and will not be used for any other purpose.
- I agree to participate in this study and to have my audio and visual data used for analysis, reports and presentations.
- I understand that information generated by the study may be published. I was guaranteed that no personal details will be disclosed from which my identity could be traced back.
- In accordance with the requirements of our University's Research Ethics Committee, all sensitive data from this project will be destroyed after funding concludes.
- I understand that the researchers and I will be walking around some of the rooms in my home to make a list of the toys my child has and their location. I also understand that they will be talking digital photos of the room, to help them remember what they have seen, as well as to use with my child on follow up visits. This information will be kept confidential and fully anonymised.
- I understand that only the research team will have access to the data I will provide.
- I understand that the information provided will be used in a report and other publications likely to be read by other parents of young children and by teachers and others working on educational issues.
- The funders ask that information from the project is made available for other researchers to use. Only information that is fully anonymised will be made available. Please tick the box if you agree to this. If you choose not to tick the box you can still be a part of the project.

I, ....., agree to take part in the research study being carried out by the researchers named above of the XXXX at the XXXXX.

I understand that:

- I do not have to take part in the research if I don't want to.
- If I change my mind and decide to withdraw from the research at any stage after signing this form, I can. I do not have to give any reason or sign anything to do so.
- Any information I give will be used for research purposes only.
- Your interview will be treated confidentially. Data will be stored electronically in a secure format. Transcripts will be anonymised and will not be archived for use by other researchers

If you do not want your name to be used in the study please tick here:

If you would like to see a copy of the research prior to submission and/or publication, please tick here:

SIGNATURE.....

DATE.....



## Communicating Chronic Pain Workshop Consent Form: Imaging and Imaging Chronic Pain

I have read the accompanying information sheet and have been sent a copy to keep. I understand what is being asked of me and have been given the opportunity to ask questions. I understand that I am entitled to withdraw from the research at any time without penalty and without needing to give a reason. I hereby consent to participating in this research.

I wish to have my words remain anonymous in all outputs from the

project:

I wish to have only my creative outputs remain anonymous:

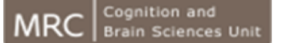
Name: \_\_\_\_\_ Date: \_\_\_\_\_

Researcher's Name: \_\_\_\_\_ Date: \_\_\_\_\_

### Creative Commons Outputs

In most research projects where creative materials are produced, researchers ask participants to sign over their copyright to the research team. Using Creative Commons licensing allows you to retain the copyright of the materials while still allowing us to use your materials as part of our research. It also means that materials can be re-shared and used by others, subject to certain conditions which you choose below. These conditions include 'attribution', 'no derivatives', 'share alike' and 'non-commercial'. We hope that this re-use will enable others to find new ways of communicating about pain.

- I am willing to share the materials produced in this workshop under Creative Commons License
- I would like my work to be attributed to me (to have my name on it)



Principal Investigator:
Project reference number:
LREC/CPREC code
Scan ID:

### INFORMED CONSENT (THIS FORM MUST BE COMPLETED PRIOR TO THE TEST)

Initials of Participant

I confirm that I have read the CBSU Guide for Volunteers, understand the volunteer Information Sheet provided to me for the above study and have had the opportunity to ask questions.

I understand that my participation is voluntary and that I am free to withdraw at any time, without giving a reason, without my medical care or legal rights being affected.

I understand that this is not a diagnostic scan by that if something abnormal should be noticed, I will be informed, as will my GP if I so wish.

I understand that, where the MRC is the sponsor, there are volunteer indemnity arrangements to cover negligent harm. Where the MRC is not the sponsor, insurance indemnity arrangements are in place.

I understand that my personal data, which link me to the research data, will be kept securely in accordance with data protection guidelines, and only available to the immediate research team.

I understand that the research data, which will be anonymised (not linked to me), may be shared with others.

I have initialled the above boxes myself and I agree to take part in the study

#### SIGNATURE OF VOLUNTEER

Signature: \_\_\_\_\_

Name in block capitals: \_\_\_\_\_ Date: \_\_\_\_\_

#### SIGNATURE OF WITNESS

Signature: \_\_\_\_\_ Date: \_\_\_\_\_

MRC Cognition and Brain Sciences Unit  
15 Chaucer Road, Cambridge, CB2 7EF  
Tel: +44 (0) 1223 355294 Fax: +44 (0) 1223 359 062

The UK Data Service has permission to share this document.

# Anonymisation



# Protecting Participants

- ◊ The best way to protect your participant's privacy may be not to collect certain identifiable information at all
- ◊ The second best is anonymisation which allows data to be shared whilst protecting participant's personal information
- ◊ Anonymisation should be considered in the context of the whole project and how it can be utilised alongside, informed consent and access controls
  - ◊ For example, if a participant consents to their data being shared then the use of anonymisation may not be required

# Identity Disclosure

- ◊ A person's identity can be disclosed through:

## 1. Direct identifiers

- ◊ e.g. name, address, postcode, telephone number, voice, picture
- ◊ Often not essential research information (administrative)

## 2. Indirect identifiers

- ◊ Possible disclosure in combination with other information
- ◊ e.g. occupation, geography, unique or exceptional values (outliers) or characteristics



# Anonymising Quantitative Data

- ◇ Remove direct identifiers
  - ◇ e.g. names, address, institution, photo
- ◇ Reduce the precision / detail of a variable through aggregation / categorisation
  - ◇ e.g. birth year instead of date of birth, occupational categories rather than job titles, 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, age
- ◇ Combining variables
  - ◇ e.g. creating non-disclosive rural / urban variable from place variables



# Anonymisation in Practice - Quant

Identifier type	Direct identifier	Strong indirect identifier	Indirect identifier	Anonymisation method
Full name	x			Remove/Change
Email address	x	x		Remove
Postal code			x	Remove/Categorise
Municipality of residence			x	Categorise
Municipality type			x	
Video file displaying person(s)	x			Remove
Year of birth		x		Categorise
Age			x	Categorise
Gender			x	
Marital status			x	
Occupation		(x)	x	Categorise
Employment status			x	
Ethnic group *		(x)	x	Categorise/Remove
Crime or punishment *			x	Categorise/Remove



# Anonymising Qualitative Data

- ◊ Plan or apply editing at time of transcription
  - ◊ except: longitudinal studies - anonymise when data collection complete (linkages)
- ◊ Avoid blanking out; use pseudonyms or replacements
- ◊ Avoid over-anonymising – removing / aggregating information in text can distort data, make them unusable, unreliable or misleading
- ◊ Consistency within research team and throughout project
- ◊ Identify replacements, e.g. with [brackets]
- ◊ Keep an anonymisation log of all replacements, aggregations or removals made – keep separate from anonymised data files

# Anonymising Qualitative Data

- » P31. Joan - Mary
- » P97. Carol - {Mother}
- » P34. Colchester - {Town in S.E.England}
- » P65. Welshpool High School - @@##High School##@@

## Example: Anonymisation log interview transcripts

Interview / Page	Original	Changed to
Int1		
p1	Spain	European
p1	E-print Ltd	Printing
p2	20 <sup>th</sup> June	June
p2	Amy	Moira
Int2		
p1	Francis	my friend



# In Practice: Anonymisation - Quali

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 about the tape.*

Comment [v1]: Replace: Ken

Comment [v2]: delete

Comment [v3]: delete

Comment [v4]: Replace: Ken

Comment [v5]: Replace: Ken

Comment [v6]: Replace: Ken



# In Practice: Anonymisation - Quali

Yeah. So is part of your job to look for funding bids and to write funding bits or is that separate?

No. That was what P3 used to do and then it sort of passed down to... really it's with P1 and P4. But I don't actually think there's anything out there at the moment. I think at the moment, because there's all this money saving and things, there's nothing to... there isn't actually anything to access.

# What if Anonymisation is Impossible?

*Anonymisation should be considered in the context of the whole project and how it can be utilised alongside, informed consent and access controls*

- ◊ Obtain consent for sharing non-anonymised data
- ◊ Regulate or restrict user access

# Anonymisation Exercises

UK Data Service



## Exercise: De-identification of qualitative data

UK Data Service



1. In this example interview transcript, where would you have concerns for the risk to disclose the identity of the interviewee? What direct and indirect identifiable information do you note in the text that might concern you? Highlight any words, phrases or sections that you think need to be dealt with.
2. How might you de-identify the text to reduce the risk to disclose the identity of the interviewee?

Source: Seymour, Jane (2010-2012). *Managing suffering at the end of life: a study of continuous deep sedation until death*. [Data Collection]. Colchester, Essex: Economic and Social Research Council. 10.5255/UKDA-SN-850749

### Case and interview

Mr Tom Jeavons, aged 63, was suffering from metastatic cancer resulting from a primary site in the bladder. His wife, Sue (58), had been his main carer for many months as he struggled with severe pain, anxiety and other symptoms. Eventually, she received support from the hospice at home team, based at their nearby hospice – St Barbara. 11 days before his death, he was admitted to their inpatient unit, where he died. The case was identified by the staff there as a “critical case”, involving palliative sedation and the difficulties staff experienced in controlling his complex symptoms. Other interviews carried out were with the hospice consultant, Dr Jane O’Connor and three nurses: Elaine McDonald, Claire Smith and Mark Ferguson. Mr and Mrs Jeavons’ GP, Dr Paul Hyde, was also interviewed which added a different medical perspective, making this an unusual case.

Central themes in all of the interviews were his intractable and distressing symptoms and the repeated requests from Mr Jeavons for euthanasia. His wife mentions earlier discussions with Mr Jeavons about the possibility of going to a Dignitas clinic, but he was already too ill to travel. She also expresses how concerned she was about what Mr Jeavons’s adult children might witness when he was dying in the hospice.

INT: So, really, it’s as I said to you: I want you to tell me what you can remember about Mr Jeavons’ care in the last week of his life ... or about Mr Jeavons in the last week of his life.

RESP: Yeah, erm, 11 days, Tom was in St Barbara’s Hospice for the last 11 days of his life so...

INT: So if you’d like to talk about that period...

RESP: Yeah.

INT: ...that’d be great.

RESP: Prior to him going in, and we was coping with his care at home, but then he was becoming less and less mobile: he couldn’t go to the toilet; he had a frame, and everything that you added in that was, it was a step to help him but a downward step to the end of how he could cope. We had a Bariatric bed brought into the other room but he insisted in sleeping in his chair. We had St Barbara’s here and, erm, the GP, and, er, we also had him assessed at home as to whether or not we could care for him completely at home. And Tom was about 20-something stone, so he wasn’t easy to manoeuvre and, and the one thing that concerned me was the fact that, erm, they needed four people to move him, you know, if he wanted to go to the toilet or if he wanted to go on a bedpan or anything, and we had the bed in there – which he wouldn’t sleep in. And, erm, basically the, logistically trying to be able to do everything for him and keep him comfortable, we’d have to wait for an on-call four nurses – could be in the middle of the night – and, and sort of the idea of being able to cope, erm, for his safety and wellbeing was, was really compromised. He didn’t want to go into St Barbara’s, he didn’t want to die in hospital, erm, but I just felt I had to take that

## Exercise: De-identification of quantitative data

In order to make microdata collected as part of the Northern Ghana Millennium Villages project evaluation publicly available via the UK Data Service, to encourage research and ensure openness and transparency, all variables in the household survey were assessed for disclosure risk, with recommendations for action.

The table below shows some assessed variables (variables commonly assessed for disclosure risk such as age, community, but also variables for which local knowledge is essential to indicate risk) with identified disclosure risk. Which action would you take to reduce the disclosure risk?

Source: University of Sussex. Institute of Development Studies, Columbia University (New York). *Earth Institute*. (2017). *Millennium Village Impact Evaluation in Northern Ghana, 2012-2015: Special Licence Access*. [data collection]. 3rd Edition. UK Data Service. SN: 7734, <http://doi.org/10.5255/UKDA-SN-7734-3>

Variables	Disclosure risk	Action
Community	Low frequency counts for all named communities, respondents who gave answers very easily identifiable (especially in combination with other variables).	
Age	Low counts of older respondents over 75 years old	
Main occupation during last 12 months	Low counts of very specific occupations.	
Ethnicity of the Household Head	Low counts of specific ethnicities.	
Household’s primary type or energy/fuel used for cooking: Firewood, Electricity-based, Charcoal, Electricity-solar panel, Gas/LPG	Very low counts for ‘Gas/LPG’ and ‘Electricity-solar panel’ responses may lead to household identification (especially if combined with other datasets)	
Main material of the wall of the house, e.g. Cane/palm/trunks, Dirt/mud, Wood/bamboo with mud, Stone with mud, Uncovered adobe, Cardboard, Cement blocks, Stone with lime/cement, Bricks, ...	A number of low-frequency responses; exterior features (households/buildings easily identifiable).	
Crops grown on plots	A number of low-frequency specific responses for each variable.	



# Copyright



# Copyright

- ◊ Copyright is internationally recognised form of intellectual property right, which arises automatically as a result of original work such as research
- ◊ Copyrighted output from research could include spreadsheets (and other forms of originally selected and organised data), publications, reports and computer programs
- ◊ Copyright will not cover the underlying facts, ideas or concepts, but only the particular way in which they have been expressed
- ◊ The right will lie with the author of the work, or with their relevant institution—different universities will have different policies on intellectual property
- ◊ A copyrighted work cannot usually be published, reproduced, adapted or translated without the owner's permission





# Copyright - Key Considerations

- ◆ Questions to ask:
  - ◆ Who the copyright holder of the datasets is?
  - ◆ Are you allowed to use them and in what way?
  - ◆ Are you allowed to archive and publish them in a data repository?
- ◆ Key considerations
  - ◆ Joint ownership
    - ◆ Datasets created by multiple researchers
  - ◆ Derived datasets
  - ◆ Database rights
  - ◆ Provision in a contract
  - ◆ Repository copyright rules



# Copyright Scenarios Exercise

## ◇ **Case Study 1 – Copyright of Archived Data**

- ◇ A researcher uses International Social Survey Programme (ISSP, n.d.) data obtained from ZACAT/GESIS - Leibniz Institute for the Social Sciences in Germany. These data are freely available to registered users. The researcher incorporates some of the ISSP data within a database containing his own research data. Can this database be deposited with another archive?

## ◇ **Case Study 2 – Copyright of Data in the Public Domain**

- ◇ A researcher studies how health issues around obesity are reported in the media in the last 10 years. Freely available newspaper websites and library sources are used to obtain articles on this topic. Articles or excerpts are copied into a database and coded according to various criteria for content analysis. (i) Can the researcher use such public data without breaching copyright? (ii) Can the database be archived and shared with other researchers?

# Copyright Scenarios Exercise

## ◇ **Case Study 3 – Copyright of Survey Questions**

- ◇ A researcher wishes to reuse a set of questions from an existing survey questionnaire, to compare results between the newly proposed survey and the original

## ◇ **Case Study 4 – Copyright of Interviews with Stay-at-Home Parents**

- ◇ A researcher interviews various stay-at-home parents about their careers and produces audio recordings and near verbatim transcripts herself. The researcher analyses this material and offers it to a data archive. The researcher did not get signed copyright transfers for the interviewees' words. What are the rights issues surrounding this offer of data

# Questions

---

*Dr Scott Summers*

*Lecturer in Business Law (UEA)*

*S.Summers@uea.ac.uk*

*UK Data Service, University of Essex*

*ukdataservice.ac.uk/help/get-in-touch*

 [cessda.eu](https://cessda.eu)

 [@CESSDA\\_Data](https://twitter.com/CESSDA_Data)