

---

# Anonymisation of research data

**Veerle Van den Eynden**

UK Data Service

University of Essex

Creating Shareable Research Data: Managing and Archiving Social Science Research Data

28<sup>th</sup> and 29<sup>th</sup> November 2017

---

UK Data Service

---



---

# Scenario

- Research on asylum seekers and refugees' experiences of forced labour, using interviews. These participants can be considered vulnerable.
- We want to use the collected research data (interviews) for analysis and share afterwards with other researchers
- How should we protect participant's anonymity?
- Read about this [case study](#)



---

# Scenario - protect anonymity measures

- not record any official identifying data (e.g. Home Office numbers)
- let participants choose a pseudonym for interview
- password protect interviewee contact details
- do not connect in any way pseudonyms to the password protected interviewee contact details

---

# Protect identities

- Avoid collecting identifying information
- Anonymise other information
- Plan in advance!



---

# Identity disclosure

A person's identity can be disclosed through:

- **direct identifiers**

*e.g. name, address, postcode, telephone number, voice, picture*

often NOT essential research information  
(administrative)

- **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 vs. date of birth, occupational categories, 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*



# Examples

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 anonymisation log of all replacements, aggregations or removals made – keep separate from anonymised data files



# Anonymising qualitative data

## 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

P31. Joan → Mary

P97. Carol → {Mother}

P34. Colchester → {Town in S.E.England}

P65. Welshpool High School → @@##High School##@@

# 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 about 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



# In practice: example anonymisation

Yeah. So is part of your job to look for funding bids and to write funding bids 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.

M27UK

INT: I'd like to start with your career history really. Perhaps you can start from when you finished school?

RES: First of all I'd like to say that I do not think that my career is very traditional for someone who comes and works abroad. I graduated from mathematics in 'Country F'. That was equivalent to what is now BSc and MSc and then I did a separate MSc in Computer Science which at that time was called like Specialisation in computer science. And then I started working at the 'Country F' University as a teacher.

INT: Was that in 'City D'?

RES: No, my Specialisation in Computer Science was in 'City D' and my first degree, my mathematics degree was in another university in 'Country F' it is called [Text Cut]

So after I finished in 'City D' my specialisation in computer science I went back to [Text Cut] where I did my first degree and I started working there as an assistant Professor.

---

# What if anonymising is impossible?

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



---

# Exercise: qualitative data

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



---

# Exercise: qualitative and quantitative data

- Steven Russell, University of East Anglia. (2013). Life on antiretroviral therapy: People's adaptive coping and adjustment to living with HIV as a chronic condition in Wakiso District, Uganda. Data catalogue. UK Data Service. SN: 851094, <http://doi.org/10.5255/UKDA-SN-851094>

