



Facilitate Open Science Training for European Research

RESEARCH DATA MANAGEMENT AND OPEN DATA

6-7 October 2015

University of Manchester, UK



ANONYMISATION

Veerle Van den Eynden, UK Data Service



UK Data Service



Options for sharing research data that may contain confidential information

- Obtain informed consent, also for data sharing and preservation / curation
- Protect identities e.g. anonymisation, not collecting personal data
- Regulate access where needed (all or part of data) e.g. by group, use, time period

Why anonymise research data?

Ethical reasons

- protect people's identity (sensitive, illegal, confidential info)
- disguise research location

Legal reasons

- not disclose personal data

Commercial reasons

Discuss with your research participants



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 (or replace with pseudonyms)
e.g. names, address, institution, photo
- Reduce precision/detail through aggregation
e.g. birth year vs. date of birth, occupational categories, area rather than village
- Generalise meaning of detailed text
e.g. occupational expertise
- Restrict upper lower ranges to hide outliers
e.g. income, age

Anonymising qualitative data

- Remove direct identifiers, or replace with pseudonyms - often not essential research info
- Avoid blanking out; use pseudonyms or replacements
- Identify replacements, e.g. with [brackets]
- Plan or apply editing at time of transcription
- Avoid over-anonymising - removing information in text can distort data, make them unusable, unreliable or misleading; so balance anonymisation with the need to preserve context
- Consistency within research team and throughout project.
- Keep anonymisation log of replacements or removals made - keep separate from anonymised data files

Anonymisation log

Example: Anonymisation log interview transcripts		
Interview / Page	Original	Changed to
Int1		
p1	Spain	European
p1	E-print Ltd	Printing
p2	20 th 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##@@

Audio-visual data

Digital manipulation of audio and image files can remove personal identifiers

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

Labour intensive, expensive, may damage research potential of data

Better:

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

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 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.

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
- regulating/restricting user access, e.g. at UK Data Archive:
 - archived data NOT in public domain
 - use of data for specific purposes only after user registration
 - data users sign legally binding End User Licence - e.g. not identify any potentially identifiable individuals
 - stricter access regulations for confidential data (case by case basis):
 - access to approved researchers only
 - requiring data access authorisation from data owner prior to data release
 - confidential data under embargo for given time period
 - secure access to data
- researchers - consider access to data and safe storage

Managing access to data

Open

- available for download/online access under open licence without any registration

Safeguarded

- available for download/online access to logged-in users who have registered and agreed to an End User Licence

Controlled

- available for remote or safe room access to authorised and authenticated users whose research proposal has been and who have received training

Can such research data be open ?

- ESRC research data policy:
 - *Publicly-funded research data are a public good, produced in the public interest, which shall be made openly available and accessible with as few restrictions as possible in a timely and responsible manner that meets a high ethical standard and does not violate privacy or harm intellectual property.*
 - Openly available research data, with as few restrictions as possible, means in the ESRC context that research data will be made available for re-use free of charge, as open data, safeguarded data or controlled data; the access category being selected to minimise the risk of disclosing personal information

Open about data with restricted access

Publish:

- Which data exist
- Where data are kept, e.g. which repository
- Who can access them
- For which purpose can they be used
- Under which conditions

In practice: data with access conditions

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.

- Interviews (audio + transcript) and written diaries with 54 people
- 40 interview and diary transcripts are archived and available for re-use by registered users
- 3 interviews and 5 diaries are embargoed until 2015
- audio files archived and only available by permission from researchers

discover.ukdataservice.ac.uk/catalogue/?sn=5407

doc.ukdataservice.ac.uk/doc/5407/mrdoc/pdf/q5407userguide.pdf

In practice: access conditions ReShare

Global Uncertainties: Security In an Africa of Networked, Multi-Level Governance

Leonard, David 2014. Global Uncertainties: Security In an Africa of Networked, Multi-Level Governance. [Data Collection]. Colchester, Essex: Economic and Social Research Council. <http://reshare.ukdataservice.ac.uk/id/eprint/851233>

The programme of research centres on how the various institutions responsible for the production of security and the management of conflict in Sub-Saharan African societies do, could and should evolve in response to the presence of violent conflict. These institutions include: the armed forces, police, courts (civil as well as criminal), 'traditional' institutions of local governance, elections (and other methods of choosing leadership), international organisations (such as the United Nations, African Union, UN High Commission for Refugees, the World Court.), and the community of international donor nations. The programme is built on the observation that all governance (especially in Africa) is multi-leveled and networked - from the village to the international organisation, and well beyond what is specified in formal government structures. Thus the focus will be not only on the ways in which key conflict-management institutions evolve themselves but also on the changing ways in which the networks in which they are embedded actually operate. This leading edge research challenges theories about the state, state formation, and sovereignty as well as the dynamics of violent conflict as presented in the dominant current work on Africa. The primary methods of research will be network analysis and qualitative interviewing.

Creators:	Creator Name	Email	Affiliation	ORCID
	Leonard, David	Unspecified	Institute of Development Studies	Unspecified
Research funders:	Economic and Social Research Council			
Grant reference:	RES-071-27-0048			
Subjects:	Law, crime and legal systems Politics			
Date deposited:	31 Jan 2014 19:00			
Last modified:	22 Oct 2014 19:27			

In practice: access conditions ReShare

— Coverage and Methodology

Collection period:	Date from: 1 April 2009	Date to: 31 March 2013
Country:	Sierra Leone	
Data collection method:	Group interviews in 39 villages and towns in rural Sierra Leone	
Observation unit:	Groups	
Kind of data:	Alpha-numeric, Textual	
Type of data:	Qualitative and mixed methods data	
Resource language:	English	

— Access and Administration

Data sourcing, processing and preparation: Citizen interviews

Copyright holders:	Name	Email	Affiliation	ORCID
	Leonard, David	Unspecified	Institute of Development Studies	Unspecified
	,	Unspecified	Unspecified	Unspecified
Contact:	Name	Email	Affiliation	ORCID
	Leonard, David	leonard@berkeley.edu	Institute of Development Studies	Unspecified

Notes on access: The names of individual respondents have been removed from the data set.

Publisher: Economic and Social Research Council

Last modified: 22 Oct 2014 19:27

In practice: access conditions ReShare

AVAILABLE FILES

Data

- Security_%26_Networks.xlsx
 - Accessible to:** Registered users only (safeguarded data)
 - File or bundle content:** Data
 - File or bundle description:** Security & Networks
 - File format:** application/octet-stream
 - License:** UK Data Service End User Licence
 - File size:** 10Kb
- + Sierra_Leone_Security_%26_Networks__Coded.xls
- + Copy_of_Somalia_SC_peace_initiatives.xlsx

Documentation

- Sierra_Leone_methods.doc
 - Accessible to:** Anyone (open data)
 - File or bundle content:** Documentation
 - File or bundle description:** Sierra Leone methods
 - File format:** application/msword
 - License:** UK Data Service End User Licence
 - File size:** 64Kb

In practice: access conditions ReShare

The Impact of Living Abroad: Stress, Adaptation, and Intergroup Contact

Geeraert, Nicolas 2014. The Impact of Living Abroad: Stress, Adaptation, and Intergroup Contact. [Data Collection]. Colchester, Essex: Economic and Social Research Council. <http://reshare.ukdataservice.ac.uk/id/eprint/850827>

Whether as migrant, expatriate, or overseas student, living abroad has an impact on individuals in terms of well-being, task-oriented functioning, and intergroup behaviour. This research examines the impact of intercultural contact by comparing such sojourners with a matched group of peers. The study will monitor a large cohort of international exchange students (sojourners) before, during and after intercultural contact. The research has four objectives. A first objective is to research the occurrence of acculturative stress, specifically examining when, and why stress occurs, and how it evolves over time. A second objective is to examine cultural learning and adaptation. In particular, the research will explore whether sojourners benefit in terms of improved culture specific knowledge (eg language, customs), general cultural awareness, and personal development (eg self-confidence). A third objective is to study the impact of intercultural contact on intergroup behaviour. The study will appraise whether and how intergroup contact influences perception of ingroup and outgroups. Finally, the study will investigate how differences in cultural distance (ie the extent to which cultures are different) affect these issues of stress, adaptation, and intergroup contact.

Creators:	Creator Name	Email	Affiliation	ORCID
	Geeraert, Nicolas	Unspecified	University of Essex	Unspecified
Research funders:	Economic and Social Research Council			
Grant reference:	RES-062-23-1211			
Subjects:	Psychology			
Date deposited:	04 Apr 2013 13:42			
Last modified:	01 Oct 2014 13:22			

+ Coverage and Methodology

+ Access and Administration

AVAILABLE FILES

Data

- data.csv

Accessible to:	Registered users requesting permission access from depositor until 1 January 2015
File or bundle content:	Data
File or bundle description:	data



Questions ?

- Veerle Van den Eynden
- veerle@essex.ac.uk

