
Legal and Ethical Considerations to Creating Shareable Research Data

Dr Scott Summers

UK Data Service

University of Essex

International Open Access Week - UEA

24th October 2018



Overview

1. The UK Data Service
2. Sharing research data
3. Legal considerations
4. Three-pronged approach
 1. Consent
 2. Anonymisation
 3. Access controls
5. Questions

What is the UK Data Service?

- Funded by the ESRC
- Single point of access to a wide range of secondary social science data
- We provide support and training for data creators with accessing, managing, sharing and using data
- Delivered by staff based at universities across the UK (Essex, Manchester, Leeds, Southampton, Edinburgh & UCL)
- UK Data Archive – manages the UK Data Service and curates the data

Some statistics about the UK Data Service

- **7,277** datasets in the collection
- **1034** qualitative and mixed methods collections
- **400** new datasets added each year
- **219** case studies of data reuse
- **25,000** registered users
- **60,000** downloads worldwide per year
- **4000+** user support queries per year

Why share data? (1)

- Increasing drive for openness, research transparency and sharing (linked to journal and funder requirements)
- Maximise return on investment and making optimal use of publicly funded research
- Avoid duplication of data collection
- Not burden over-researched, vulnerable groups and make best use of hard-to-obtain data, e.g. elites and socially excluded
- Extend voices of participants

Why share data? (2)

- Better science
- Technological advances – easier for digital data to be discoverable and accessible
- Visibility for researchers
- Societal benefits

Society benefits from data sharing

F1000Research
Open for Science

Articles Collections For Authors For Referees About Blog

What is open science? New on F1000Research – 17 November 2014

Beat Ebola with better research sharing, says discoverer of virus

POSTED BY THOMAS INGRAHAM, 12 NOVEMBER 2014

COMMENTS 0 [Share](#) 418

Key points

- **F1000Research: Ebola** article collection launched to enable all Ebola-related knowledge to be made available on a single platform within days of submission. All articles submitted to this collection will be prioritized and those accepted will be published free of charge.
- **F1000Research** has set up a dedicated phone line (+442071931030) to enable those working in affected areas to dictate their reports directly to our office.
- **F1000Prime** makes all existing and upcoming **Ebola and Marburg virus Article Recommendations** free to access.

Sharing critical information faster

Peter Piot (Director the London School of Hygiene & Tropical Medicine and co-discoverer of the Ebola virus) has published an **Editorial** in **F1000Research** appealing for everyone working on the current epidemic to publish all their findings and experiences rapidly and openly.

Since the onset of this year's West African outbreak* there have been over 13,250 suspected cases and 4,960 deaths, the majority of these in the last two months. Depending on how effective control efforts are over the next few weeks the number of cases will likely swell to between 16,500 and 30,000 by December.

Speed is required at all levels if this epidemic is to be successfully controlled; identification and isolation of cases, clinic construction, and community action campaigns must all be carried out rapidly and on a large scale. Speed is just as vital to Ebola research; we need knowledge to spread faster than the virus if Ebola's impact is to be mitigated. For this to happen, the whole research cycle has to start spinning faster.

Some major funders have responded quickly by releasing emergency research grants, such as the NSF's **RAPID** awards and the Wellcome Trust's **Emergency Ebola Initiative**. Science publishing needs to be just as nimble, but in the current system it can take several months for findings to be made publically available; with Ebola cases doubling every 20 days or so, such delays are unacceptable.



BBC NEWS

Home UK World Business Politics Tech Science Health Education Entertainment

Asia China India

How 'crisis mapping' is helping relief efforts in Nepal

By Saira Asher
BBC News

© 6 May 2015 Asia



Kathmandu, and nearby areas before and after mapping efforts were ramped up after the quake

Thousands of people in remote parts of Nepal are still in need of medical help and basic supplies. But with roads damaged and buildings collapsed, knowing what aid is needed and where, is a challenge. One group of Nepalis, backed by a global community, is trying to change that by "crisis mapping" Nepal.

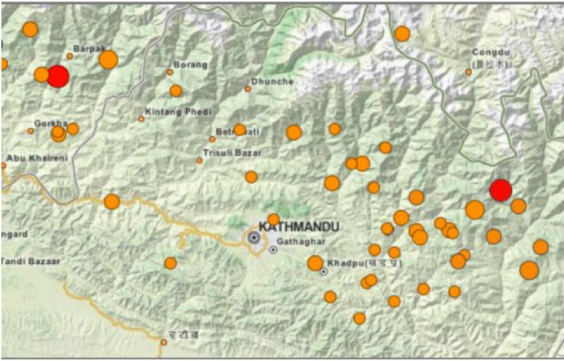
Nepal quake

Unsettled Earth continues to rattle Nepal

Nepal quake causes

HOT Humanitarian OpenStreetMap Team

About Updates Projects Get Involved Donate



FEATURED PROJECT

Nepal 2015 Earthquake Response

The 2015 Nepal earthquake (Wikipedia) struck on 25th April with a magnitude of 7.8, followed by many aftershocks including a large magnitude 7.3 quake on 12 May. The quake hit the city of Kathmandu and the mountainous regions to the north.

The Humanitarian OpenStreetMap Team [HOT] applies the principles of open source and open data sharing for humanitarian response and economic development.

Why is (open) data access important?

- Reproducibility of research
- Openness, transparency, research integrity
- Risk management
- Re-use and re-purposing
- Expectations on data sharing from:
 - Funders
 - Publishers





Enter search term



Variable and question bank > QualiBank >

STUDIES (7359)

SERIES (73)

Date from:

Date to:

1753

2018

Refine date

Topic:

Data Type:

Access:

Safeguarded (6554)

Open (546)

Controlled (151)

Reset filters

Displaying 1 - 10 of 7359 results

Page 1 of 736

Results per page:

10

Sort by:

Most recently released

SN 853258 | [Countries at risk of electoral violence 2016-2018](#)

Birch, S, King's College London | Muchlinski, D, Georgia Tech

SN 853370 | [Remembered meal satisfaction and later snack food intake: A laboratory study 2016-2018](#)

Whitlock, V, University of Liverpool | Robinson, E, University of Liverpool

SN 853371 | [A smartphone based attentive eating intervention for energy intake and weight loss: A randomised controlled trial 2017-2018](#)

Whitlock, V, University of Liverpool | Kersbergen, I, University of Liverpool
Higgs, S, University of Birmingham | Aveyard, P, University of Oxford
Halford, J, University of Liverpool | Robinson, E, University of Liverpool

SN 8391 | [Active Lives Survey, 2016-2017](#)

Sport England

SN 853342 | [The impact of coloured hyperlinks when reading text, experimental data 2018](#)

British Premier league fans, survey data 2014

Newton, Martha (2018). *British Premier league fans, survey data 2014*. [Data Collection]. Colchester, Essex: UK Data Archive. [10.5255/UKDA-SN-853014](https://beta.ukdataservice.ac.uk/datacatalog/studies/study?id=10.5255/UKDA-SN-853014)

Some of the greatest atrocities have been caused by groups defending or advancing their political aspirations and sacred values. In order to comprehend and address the wanton violence of war, terrorism and genocide, it is necessary to understand the forces that bind and drive human groups. This five year programme of research investigates one of the most powerful mechanisms by which groups may be formed, inspired, and coordinated: ritual. Studying how children learn the rituals of their communities will shed light on the various ways in which rituals promote social cohesion within the group and distrust of groups with different ritual traditions. Qualitative field research and controlled psychological experiments will be conducted in a number of troubled regions (including Northern Ireland, the Middle East, Nepal, and Colombia) to explore the effects of ritual participation on ingroup cohesion and outgroup hostility in both general populations and armed groups. New databases will be constructed to explore the relationship between ritual, resource extraction patterns, and group structure and scale over the millennia. These interdisciplinary projects will be undertaken by international teams of anthropologists, psychologists, historians, archaeologists, and evolutionary theorists.

DATA DESCRIPTION (ABSTRACT)

British football fans completed this correlational survey. Willingness to lay down one's life for a group of non-kin, well documented historically and ethnographically, represents an evolutionary puzzle. Building on research in social psychology, we develop a mathematical model showing how conditioning cooperation on previous shared experience can allow individually costly pro-group behavior to evolve. The model generates a series of predictions that we then test empirically in a range of special sample populations (including military veterans, college fraternity/sorority members, football fans, martial arts practitioners, and twins). Our empirical results show that sharing painful experiences produces "identity fusion" – a visceral sense of oneness – which in turn can motivate self-sacrifice, including willingness to fight and die for the group. Practically, our account of how shared dysphoric experiences produce identity fusion helps us better understand such pressing social issues as suicide terrorism, holy wars, sectarian violence, gang-related violence, and other forms of intergroup conflict.

Interviews of incentives for sharing research data

Van den Eynden, Veerle and **Bishop, Libby** (2014). *Interviews of incentives for sharing research data*. [Data Collection]. Colchester, Essex: UK Data Archive. [10.5255/UKDA-SN-851540](https://beta.ukdataservice.ac.uk/datacatalog/studies/study?id=10.5255/UKDA-SN-851540)

DATA DESCRIPTION (ABSTRACT)

A collection of 22 interviews with researchers in 5 European countries on their data sharing practices, their motivations to share research data, what incentivises them to share research data, and their views on how data sharing can be incentivised further.. The aim of this study was to provide evidence and examples of useful incentives for data sharing from the researchers' point of view, to inform scientists and policy makers. The study involved five research teams with an established data sharing culture in partner countries of Knowledge Exchange (Finland, Denmark, Germany, United Kingdom, and the Netherlands), and spans various academic disciplines (arts and humanities, social sciences, biomedicine, chemistry and biology). This study's aims were to: 1. identify official but especially also yet unknown or unofficial incentives of researchers for making data available with a focus on the values and intrinsic motivations of the individual as well as on the interactions within research teams and in the larger research community 2. analyse existing and possible future benefits for researchers sharing their data 3. investigate the influence of existing policies on the practice of data sharing throughout the whole lifecycle of the research process as well as the influence of existing institutions and infrastructures offering support services for data sharing 4. consider the whole research lifecycle and classify the most efficient moments in the research process for incentivising data sharing 5. provide recommendations for policy development regarding the incentivising of data access and reuse.

The role of "freedom" in EU competition law

Akman, Pinar (2015). *The role of "freedom" in EU competition law*. [Data Collection]. Colchester, Essex: UK Data Archive. [10.5255/UKDA-SN-851732](https://beta.ukdataservice.ac.uk/datacatalog/studies/study?id=10.5255/UKDA-SN-851732)

The ESRC Centre for Competition Policy (CCP) at the University of East Anglia (UEA) undertakes interdisciplinary research into competition policy and regulation that has real-world policy relevance without compromising academic rigour. It prides itself on the interdisciplinary nature of the research and the members are drawn from a range of disciplines, including economics, law, business and political science. The Centre was established in September 2004, building on the pre-existing Centre for Competition and Regulation (CCR), with a grant from the ESRC (Economic and Social Research Council). It currently boasts a total of 26 faculty members (including the Director and a Political Science Mentor), 4 full- and part-time researchers and 23 PhD students.

DATA DESCRIPTION (ABSTRACT)

The research was conducted using the official European database for case-law (EUR-Lex). Five search terms were searched for in all the documents that relate to the subject matter of 'competition' in EU case-law since 1957, as categorised by EUR-Lex (returning 2626 entries in total on 1 September 2011). These search terms were: 'freedom of competition', 'freedom to compete', 'economic freedom', 'freedom to choose' and 'freedom of choice'. They were chosen as they all relate to some aspect of freedom that is relevant for competition purposes. The database includes a total of 146 entries, comprising judgments, opinions of Advocates General (AGs) and orders in which one or more of the search terms appeared. This does not correspond to the same number of cases, since there may be more than one document in the database concerning the same case. The scope of this paper's database covers cases with subject matters falling under one or more of the substantive competition law provisions of TFEU, namely Articles 101, 102, 106 and 107 TFEU, as well as merger cases.

Change.org petitions data

Margetts, Helen Z. and **Hale, Scott A.** and **Yasseri, Taha** (2016). *Change.org petitions data*. [Data Collection]. Colchester, Essex: UK Data Archive. [10.5255/UKDA-SN-851617](https://beta.ukdataservice.ac.uk/datacatalog/studies/study?id=10.5255/UKDA-SN-851617)

DATA DESCRIPTION (ABSTRACT)

This collection contains number of signatures to 1949 petitions over time (with exact timestamp of each signature, and the city-country of the person who has signed) along with petition metadata (title, category, body, target, time of creation, and closing time), directly crawled from the website. Please see 'Related resources' section below for related data collections. This project aims to develop methodologies to study online political behaviour including use of the Internet to generate new data and experiments; to collect and analyse data on internet-mediated interactions at both individual and organisational levels; and to use this data to re-examine and where necessary develop political science knowledge and theory in light of widespread use of the Internet First, the project will re-examine the logic of collective action, assessing the impact of reduced communication and coordination costs; the changing nature of leadership; and the effects of different information environments on propensity to participate in political mobilisation. This part of the research will involve conducting laboratory and field experiments into online behaviour. Second, the research will develop the Digital-era Governance model for newer 'Web 2.0' applications and other technological developments such as cloud computing. The research will re-examine the nature of citizen-government interactions in this changing environment, examining the impact of Internet-based mediation on information exchange, organisational forms in government and citizen participation in policy-making. This part of the research will involve a comparison of government's online presence in eight countries, using webmetric techniques, and in-depth qualitative analysis of governance models, using elite interviewing and documentary analysis.

What makes data good for sharing and reuse?

Other researchers can understand and reuse the data:

- High quality
- Accurate
- Well organised
- Easily accessible
- Well documented
- Long-term validity

or **FAIR** data:

Findable, **A**ccessible, **I**nteroperable, **R**eusable

Too difficult to share data widely?

Ethical, legal and research integrity challenges

- Personal, confidential or sensitive information
- Linkage of data in multi-disciplinary projects
 - Difficult to conceal identity of participants / fieldwork locations
- Lack of trust in others mis-using data
- Beaten to publication

Legal considerations

(i) Duty of confidentiality and data sharing

- Duty of confidentiality exists in UK common law and may apply to research data
- Information given in circumstances where it is expected that a duty of confidence applies, cannot normally be disclosed without the information provider's consent
- Disclosure of confidential information is lawful when:
 - The individual to whom the information relates has consented – **consent for data sharing**
 - Disclosure is necessary to safeguard the individual, or others, or is in the public interest
 - There is a legal duty to do so, for example a court order
- Best practice is to avoid vague or general promises in consent forms

(ii) 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

(ii) 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

(iii) Data protection considerations – the GDPR (1)

- The General Data Protection Regulation (GDPR)
- New EU-wide data protection regulation
- Came into force 25 May 2018
- The GDPR gives data subjects greater control over their personal data, whilst modernising and unifying European data protection rules
- Clarity, transparency and accountability
- The Data Protection Act 2018

(iii) Data protection considerations – the GDPR (2)

- Personal data is defined as ‘any information relating to an identified or identifiable natural person’ (‘data subject’)
- Living persons
- Though there may still be ethical reasons for wanting to protect this information
- Anonymised data is **NOT** personal data so the GDPR does **NOT** apply
- Applies to:
 - any **EU researcher** (data controller) who collects personal data about a citizen of any country, anywhere in the world
 - A data controller or data processor based outside the EU but collecting personal data on EU citizens

(iii) Principles for processing personal data

1. Process **lawfully, fair** and **transparent**

Inform participant of what will be done with the data, process accordingly

2. Keep to the **original purpose**

Collect data for specified, explicit and legitimate purposes

Do not process further in a manner incompatible with those purposes

3. **Minimise** data size

Personal data collected should be adequate, relevant and limited to what is necessary

4. Uphold **accuracy**

Personal data should be accurate and kept up to date

5. **Remove** data which are not used

6. Ensure **data integrity and confidentiality**

Protection against unauthorised or unlawful processing, accidental loss, destruction or damage, using appropriate technical or organisational measures

(iii) Data subject rights

- The right to be **informed**
- The right of **access**
- The right to **rectification** (correction)
- The right to **erasure** (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

(iii) Grounds for processing personal data

One of these must be present to process a data subject's personal data:

- **Consent** of the data subject
- Necessary for the performance of a **contract**
- **Legal obligation** placed upon controller
- Necessary to **protect vital interests** of the data subject
- Carried out in the **public interest** or is in the exercise of official authority
- **Legitimate interest** pursued by controller

(iii) The GDPR research exemption

Further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes is not considered to be incompatible with the initial purposes

Appropriate safeguards, e.g.

- Data minimisation
- Pseudonymisation
- Technical and organisational measures

Principles 2 and 5 less strict:

- Purpose: further processing of personal data allowed (2)
- Personal data may be stored for longer periods (5)

How to share data obtained from participants in a legal and ethical way

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

- * Store personal or sensitive data **securely** and separately*
 - (Utilise encryption and consider the storage location)

Consent

Consent for sharing – one more small step

- 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

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
One-off consent: Participant is asked to consent to taking part in the research project only once	<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>
Process consent : Participant's consent is requested continuously throughout the research project	<p>Ensures 'active' consent</p>	<p>May not get all consent needed before losing contact</p> <p>Repetitive, can annoy participants</p>

Informed consent – research (1)

To obtain informed consent in practice, researchers should:

- Inform participants about the purpose of the research
- Discuss what will happen to their contribution (including the future archiving and sharing of their data)
- Indicate the steps that will be taken to safeguard their anonymity and confidentiality
- Outline their right to withdraw from the research, and how to do this

Informed consent – research (2)

- When seeking to obtain informed consent from participants, it is important for researchers to also consider the specific circumstances and needs of the participants
- This may mean, for example: pictures or diagrams are used on the consent form instead of using a lot of text or the consent form is translated into another language

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.

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.

I understand that taking part in the study involves [.....].

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.

2. Use of the information in the study

I understand that information I provide will be used for [.....].

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.

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.
- ii) If you want to use named quotes: I agree that my real name can be used for quotes.
- 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].

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.

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.

Informed consent form content

- Break down into 3 key areas:
 - I. Taking part in the study
 - II. Use of the information in the study
 - III. Future use and reuse of the information by others

(i) Taking part in the study

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.

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.

I understand that taking part in the study involves [.....].

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.

(ii) Use of the information in the study

2. Use of the information in the study

I understand that information I provide will be used for [.....].

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.

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.
- ii) If you want to use named quotes: I agree that my real name can be used for quotes.
- 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].

(iii) Future use and reuse of the information by others

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.

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.

Anonymisation

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 - tips



- 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

In practice: table (Finland)

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 or make it misleading
(<https://www.ukdataservice.ac.uk/deposit-data/stories/gush>)
- Consistency within research team and throughout project
- Identify replacements, e.g. with [brackets]
- Keep an anonymisation log of all replacements, aggregations or removals made and keep it *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 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##@@

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

What about 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 disclosing information during audio recordings

What if anonymising 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

Access Controls

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 (*e.g. not identify any potentially identifiable individuals*)
- special agreements (depositor permission; approved researcher)
- embargo for fixed time period

Controlled

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

Open where possible, closed when
necessary

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 and transcript) and written diaries with 54 people
- 40 interview and diary transcripts are archived and available for re-use by registered users (**Safeguarded**)
- 3 interviews and 5 diaries were embargoed until 2015 (**Safeguarded – Embargoed**)
- Audio files archived and only available by permission from researchers (**Safeguarded – Special Agreement**)

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

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

In practice: access & licensing ReShare

Use terms and conditions for open access data

- Data files deposited as Open Data are licensed under the Depositors choice of one of two [Creative Commons Attribution 4.0 licenses](#):
- **CC-BY-NC-SA** Creative Commons [Attribution-NonCommercial-ShareAlike 4.0 International](#)
- **CC-BY-SA** Creative Commons [Attribution-ShareAlike 4.0 International](#) (this licence allows for commercial use)
- ReShare emphasises that Creative Commons licenses can only be agreed to by the copyright holder(s) and should not be used if there are third party rights holders

Add a new file or zip bundle

Upload data files and documentation for your data collection. Click the Browse button below to select a file; and the Upload button to upload it to the repository. Large files (>2GB) make take a long time to upload, depending on your browser speed. When uploading many files (>10), please upload them in logically grouped zip bundles instead, e.g. grouping by content (data, documentation), by file format (text, images, databases) or by folder structure.

Make sure data files contain no disclosive personal information.

Include documentation on data collection methods and context to inform future reuse of the data.

Make sure file names contain no spaces.

File From URL

Choose File no file selected

+ Interview_Data_001.docx 20Kb

Hide options -

* **Accessible to:** Selecting an access level will determine who is able to download this file or bundle:

- open data are accessible to any user without registration
- safeguarded data are accessible only to users registered with the UK Data Service

Consult our detailed [data use terms and conditions for open data and safeguarded data](#). If you think that more access controls to the data are needed, please provide information to the ReShare administrator in the "Notes to administrator" text box (previous page).

Anyone (open access data)

File or bundle content: Data

File or bundle description:

Embargo date:

* **License:** Specify an explicit license for this file or bundle. This repository allows [Creative Commons](#) licences for open data (accessible to anyone) and the [UK Data Service End User Licence](#) for safeguarded data accessible to registered users of the UK Data Service only.

Creative Commons Attribution 4.0 International

Save

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:	Date to:		
	1 April 2009	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



Questions

Contact Details:

Scott Summers

UK Data Service

University of Essex

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

