



# FOSTER

---

## Facilitate Open Science Training for European Research

Scottish Graduate School of Social Science Workshop:  
**Overcoming obstacles to sharing data about human subjects**

Edinburgh, 10 June, 2015

Robin Rice  
EDINA and Data Library  
University of Edinburgh






# Overview

- What are personal / sensitive data?
- Legal and ethical principles for working with personal / sensitive research data
- Disclosure control overview
- Governance for making data safe

# What are sensitive research data?





Generally -

-  Data relating to people.
-  Data generated or used under a commercial research funding agreement.
-  Any data likely to have significant negative public impact if released.

# What are personal data?

*Any information about a living, identifiable individual*

Typical information that can be used to directly identify person in isolation or in combination:

-  Name
-  Address
-  Date of Birth
-  National Identification Number (NIN)

# What are sensitive personal data?

Information that can be used to discriminate requires extra protection:

- 🌸 Racial or ethnic origin
- 🌸 Political opinions
- 🌸 Membership of a political association
- 🌸 Religious beliefs or affiliations
- 🌸 Membership of a professional or trade association
- 🌸 Membership of a trade union
- 🌸 Sexual preferences or practices
- 🌸 Criminal record
- 🌸 Health and genetic information

# Managing personal / sensitive data legally

- The Data Protection Act outlines two classes of information to be protected by data controllers and data processors:
  - personal data and
  - sensitive personal data
- Research projects typically apply same (high) level of protection to both information types.

# 8 Principles of Data Protection Act, 1998

Anyone who processes personal data must make sure that personal data are:

1. fairly and lawfully processed;
2. processed for limited purposes;
3. adequate, relevant and not excessive;
4. accurate and up to date;
5. not kept for longer than is necessary;
6. processed in line with data subject's rights\*;
7. secure; and
8. not transferred to other countries without adequate protection.

\* Includes right to request what data is held; right to correct data held.

<https://ico.org.uk/for-organisations/guide-to-data-protection/>

# Research Exemption?

The Scottish DPA has a 'research exemption' under Clause 3.3.

 Applies when personal data will not be processed

- In a way that affects the individual
- Causes distress

 IF the exemption applies

- Data collected for one purpose may be used for other purposes
- May be kept indefinitely
- Exempt from subject's rights of access, where statistically anonymised

 Caveats

- Always consult your university's Data Protection Officer
- The law is due to change in 2017: research exemption may be revoked
- Students process data on their own behalf, not Uni's [except in certain circumstances](http://www.ed.ac.uk/schools-departments/records-management-section/data-protection/guidance-policies/students-processing-data) (<http://www.ed.ac.uk/schools-departments/records-management-section/data-protection/guidance-policies/students-processing-data>)






# Ethics of collecting data on human subjects

Ethical principals are in line with, or go beyond legal principals. Ethics cover:

- 🌻 The purpose and nature of the research itself
- 🌻 The nature of consent obtained (e.g. opt-in versus opt-out participation)
- 🌻 What data needs to be safeguarded during analysis and destroyed after its use







# Which research is subject to ethical review?

-  Using human participants or live animals.
-  Referencing individual subjects (people).
-  Keeping identifiers for individuals.

# Which research is exempt from ethics review?

- Freely available in the public domain, or
- Obtained through pure observational studies of public behaviour:
  - are of human action that occurs in a forum open to the general public
  - are non-invasive
  - require no interaction with participants
  - do not identify participants
- Does this apply to internet/social media?
  - Not necessarily! Researchers need to consider the expectations of the providers of the information and seek consent where appropriate.

# Creating / collecting sensitive data

-  Don't ask those personal questions if you don't need to!
-  Familiarise yourself with the Data Protection Act (DPA).
-  Get advice on DPA from Records Management\* at an early stage.
-  Go through your School's ethical approval process.
-  Use an information sheet to fully inform participants on how you will make use of their data.
-  Obtain consent with a consent form for data publication and sharing from research participants before the data are collected.

\* Records Management:

<http://www.ed.ac.uk/schools-departments/records-management-section/data-protection>

# What to include in a consent documentation?

A written consent documentation includes:

- 🌻 an information sheet and
- 🌻 a consent form signed by the participant.

# Information sheets

An information sheet should cover:

- 🌻 Purpose of the research
- 🌻 What's involved in participating
- 🌻 Benefits and risks
- 🌻 Terms for withdrawal
- 🌻 Usage of the data
- 🌻 Strategies for ethical use of data
- 🌻 Details of the research

# Consent forms

A consent form should cover the following points:

- 🌸 The participant has read and understood information about the project.
- 🌸 The participant has given the opportunity to ask questions.
- 🌸 The participant voluntarily agrees to participate in the project.
- 🌸 The participant understands that s/he can withdraw at any time without giving reasons without penalty.
- 🌸 Procedures regarding confidentiality are explained.
- 🌸 Separate terms for consent for interviews, audio, video or other forms of data collection.
- 🌸 Use of the data in research, publications, sharing / archiving explained.
- 🌸 Signatures and dates of signing for the participant and the researcher.

# Video

- 🌻 Lynn Jamieson, CRFR, on written consent (from Research Data MANTRA): <http://edin.ac/1JrrkQ1>



# Sharing sensitive data: Why share?

## Funder Expectations:

-  Research Councils UK “Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property.”
-  Economic & Social Research Council “ ... publicly funded research data ... valuable, long-term resources that, where practical, must be made available for secondary, scientific research.”
-  Medical Research Council “... publicly-funded research data ... should be openly available to the maximum extent possible.”




## Why data should be shared

<b>University of Edinburgh policy</b>	Where appropriate, research data management plans must ensure that research data are available for access and re-use.
<b>UK research funding council mandate</b>	To avoid duplication of effort, maximise funding and extends its reach.
<b>Academic integrity</b>	Data sharing facilitates transparency, accountability and scrutiny, and prevents fraud.
<b>Increased research impact</b>	Both within and beyond the discipline in which the data were created.
<b>Publicly funded research</b>	Should be available to the public.
<b>Preservation</b>	Assists with long term storage, for retrieval and future use by yourself and others.
<b>Teaching resource</b>	Help students to collect and manage data.

*From Data Sharing Guide on UoE [RDM website](#)*

# Sharing sensitive data: Why not share?

Sharing is encouraged but not suitable for all data. Must be balanced with other obligations:



-  Ethical duty of confidentiality and protect participants from harm.
-  Allow participants to make their own decisions on how their information can be used, shared, and made public (through informed consent).
-  Enable researcher to gain maximum benefit from findings.

## Data sharing - your concerns?

<b>Unrealised commercial value?</b>	Edinburgh Research and Innovation (ERI) can help you determine the value of your research and identify any patent potentiality.
<b>Sensitivity of data?</b>	If there are ethical concerns, it may be possible to anonymise data to remove personal information, so that it may be shared.
<b>Issues of data ownership?</b>	If parts of the data are owned by others and/or your data is derived from the original source, you may not have the rights to share it.
<b>Controlling access to data?</b>	If data are sensitive and anonymisation to permit open sharing is not possible, then it is acceptable to limit access to the data to researchers who sign a suitable confidentiality agreement. Metadata about this data should be openly shared so that other researchers know what it contains and how to request access.

*From Data Sharing Guide on UoE [RDM website](#)*

# Sharing Sensitive Data: Risks

-  1. Triangulation: When the identity or sensitivity of a participant or subject can be determined by combining several pieces of non-sensitive information. For example, with human data, if you have information about a person's age, occupation and family composition, it may not be difficult to identify them in a relatively small sample.
-  2. Data Linkage is bringing together two or more datasets that include the same person or subject of research. Alone, a dataset may not contain enough information to identify individuals or make place subjects at risk, but when combined with two or more datasets, this may now be achievable.

# Disclosure scenario

- Dataset A describes the medical history of a group of non-identifiable patients with a cognitive disorder.
- Dataset B includes employment information and public transport usage in the same, moderately small population.
- When these datasets are linked, there could be sufficient information about where patients work and live such that they may be identified.

# Sharing sensitive data: textual data

## Advice from UK Data Archive:

- 🌸 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 file

# Sharing sensitive data: audio-visual data

## Advice from UK Data Archive:

Digital manipulation of audio and image files can remove personal identifiers

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

[http://ukdataservice.ac.uk/media/440279/ethicallegal\\_anonymisation.pdf](http://ukdataservice.ac.uk/media/440279/ethicallegal_anonymisation.pdf) (page 8)



# Sharing sensitive data: quantitative data

## Aggregation: release data in table format only

- Cell thresholds
- Permutation
- Cell swapping

## Generalisation:

- limit geographic details
- use standard terminology such as standard occupation codes

## Banding and top-coding variables

- age groups
- income

# Sharing sensitive data: De-identifying data

1. Re-sort and renumber cases
2. Remove direct identifiers
3. Remove or recode specific days and dates
4. Examine geographic variables, recode or remove based on analytic purposes of data
5. Examine variables that could provide links to external datasets

# Governance options for making data safe

- Safe data
  - Public use dataset (fully anonymised)
- Safe researcher
  - Accredited training process
- Safe project
  - Approved research project; in public interest
- Safe settings & safe outputs
  - Use of secure access facility or remote secure access.
  - Outputs checked by responsible staff-person.
  - Data linkage may be offered as part of the service.

# Safe haven examples for social science and medical data

-  [ONS data via UK Data Service Secure Lab](#)
-  [Administrative Data Research Network](#)
-  [The Scottish Longitudinal Study Development and Support Unit \(SLS-DSU\)](#)
-  [National Safe Haven and Electronic Data Research and Innovation Service \(eDRIS\)](#)
-  [Farr Institute of Health Informatics Research](#)