Working with Sensitive or Confidential Research Data in MPLS and Medical Sciences

John Southall
Bodleian Data Librarian
Subject Consultant for Economics, Sociology, SP&I

Cathy Scutt
Subject Consultant for Education

Trinity Term 2019
Our Focus Today

• Are you creating or using existing data
• Protecting all your data
• Aware of appropriate strategies
• Aware of the support at Oxford
Types of Data

• Numerical - Text based - Audio Visual
• Each are representations of information that may be confidential
• Goal: understand
  • Why your data is considered confidential or sensitive in some way
  • Impact of this on research process
  • How you manage and preserve it
Consider Reasons

• Why confidential?
• Because of content?
  • Broadly two types

OR

• Framework in which data is used
  • Undertakings given to research participants
  • Funder or institutional requirements

All of the above?
Particular Content

• Consent agreement or legislation define it
• Specific sections are confidential
  • Example; survey data where post codes have been collected as one of the variables
  • Example; Interview data that includes names and accounts of illegal activity
  • Example; Information defined as confidential by the participant
• Opens up possibility of tagging
• Editing / ‘Anonymisation’ (framework)
Handling Particular Content

In practice;

• Security during collection and handling paramount
• More options on reducing sensitivity of future versions
• ‘Safer’ preservation copy
  • ‘safer’ in what sense?
General Content

• Subject matter as a whole is sensitive or confidential
• Non-sensitive elements cannot be separated
• Everything has to be dealt with in the same way
• Editing not appropriate
• Consent agreement or legislation play a role
Handling General Content

In practice;
- Security during collection and handling paramount
  - More so perhaps – throughout data use
- Fewer options on reducing sensitivity of future versions
- Judgement of researcher
- Greater responsibility on researcher
  - Demonstrate this is the case
General or particular?

What applies for the group today?

When does it apply?
Content and Stakeholders

• Not just about content but also stakeholders in the research process

• Who are the concerned parties?
  • Project Participants – supplying the information
  • The audience – for the information and related analysis
  • The researcher – who will gain or lose based on how it is used and received
Why Important?

• Understanding reasons your data is confidential helps you make best use of it
  • In current project
  • Gain ethical approval/ curec
  • In the future
    • Retaining
    • Depositing
    • Sharing
Consent Undertakings

• Help ensure participation BUT also define interests of stakeholders

• Avoid agreements that are too restrictive – don’t make unnecessary promises – negotiate!
  • “only to be used by this researcher” - “will be destroyed” - “no one else will read”

• Useful Informed Consent?
  • Undertakings need to encourage participation
  • Protect everyone involved
  • Create trust
  • Pilot/ trial your agreements
  • www.admin.ox.ac.uk/curec/resources/informed-consent/
  • www.data-archive.ac.uk/create-manage/consent-ethics
Welcome to the Research Data Oxford website

About RDM
Overview of research data management and funder policies.

Working with data
Data management day-to-day and at the project planning stage.

Sharing data
Sharing, licensing, depositing, and citing your data.

Tools, services, and training

Research data glossary

Oxford research data blog

ORA-Data

Deposit your data
Not sure if you’re ready?
See the Pre-deposit checklist

University research data policy
What does my funder expect?
Data management planning
A to Z site index
Contact us

Recent blog posts
Data Storage or preservation?

- Securely holding the data is key
  - But only **one** part
- Enabling efficient access for you
  - Short term
  - Long term
  - Encryption / disaster planning
- Managing data
  - Version control
  - Honouring agreements made
  - Storage becomes active preservation
Legal Regulation

• Data Protection Act
• General Data Protection Regulation (GDPR) May 2018
  • Research occupies a privileged position within the Regulation. Organizations that process personal data for research purposes may avoid restrictions on secondary processing and on processing sensitive categories of data (Article 6(4); Recital 50). As long as they implement appropriate safeguards, these organizations also may override a data subject’s right to object to processing and to seek the erasure of personal data (Article 89).
  • Useful to bear in mind
Three Main Approaches

Managing data during *and* after a project

- **Destruction**
  - Requires good reasons
  - Wasteful

- **Anonymisation**
  - Clearly defined
  - Time consuming and imperfect

- **Access Restriction**
  - Leaves content intact
  - Needs active management
Blurring, Masking or Anonymisation

- During and after a project
- **Light** touch; limited key identifiers e.g. Names and addresses only
- Replacement / Pseudonyms – data blurring
- Aggregation – fine grain detail/numbers removed
- Randomised sampling
- Effectiveness tests?
  - Singling out
  - Linkability
  - Inference
Blurring, Masking or Anonymisation

• Perhaps best used for **particular content**
  • Removing columns from spreadsheets
  • Specific names/words in transcripts
• But an imperfect solution – too blunt a tool?
• Dangers of data degradation or distortion
• ICPSR Guidance - [www.icpsr.umich.edu/icpsrweb/deposit/index.jsp](http://www.icpsr.umich.edu/icpsrweb/deposit/index.jsp)
• UK Data Archive Guidance - [www.data-archive.ac.uk/create-manage/consent-ethics/anonymisation](http://www.data-archive.ac.uk/create-manage/consent-ethics/anonymisation)
• UK Anonymisation network - [//ukanon.net/](http:////ukanon.net/)
Restricting Access

- Anonymisation allows wide access to less data (i.e. by removing content) post project.
- An alternative approach is to leave content but make access harder.
  - E.g. Microdata from Eurostat.
  - Vetting of access from UKDS.
  - Requires clear access and usage conditions.
  - Restrict what content may be reproduced.
  - Introduce embargoes (last resort).
Restricting Access

• During and after a project?
  • Data security

• Best used for general content confidentiality?

• Effective or credible policing of restrictions needed

• Both approaches can increase usage potential of data but require planning from the beginning
Effective Handling and Use

• Document the research process
  • Metadata that captures decisions and clear requirements
  • How sensitive data is managed and processed

• Pilot consent paperwork

• Think about what could go wrong!
  • Collecting inappropriate data
  • Hardware /software failure
  • Security – breaches - theft
  • Managing accusations of disclosure
In conclusion?

• Seek support and advice
• Become familiar with services
Welcome to the Research Data Oxford website

About RDM
Overview of research data management and funder policies.

Working with data
Data management day-to-day and at the project planning stage.

Sharing data
Sharing, licensing, depositing, and citing your data.

Tools, services, and training

Research data glossary

Oxford research data blog

ORA-Data

Deposit your data

Not sure if you're ready?
See the Pre-deposit checklist

University research data policy
What does my funder expect?
Data management planning
A to Z site index
Contact us

Recent blog posts