Working with Sensitive or Confidential Research Data in the Humanities and Social Sciences

John Southall
Bodleian Data Librarian

Cathy Scutt
Bodleian Education Librarian

Trinity Term 2020
Our Focus Today

• 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
• Goal should be to 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
  • e.g. Undertakings given to research participants
  • e.g. Funder or institutional requirements

All of the above?
(Particular) Content

• Consent agreement or legislation define it
• **Specific sections** are confidential
  • e.g. survey data where post codes have been collected as one of the variables
  • e.g. Information defined as confidential by the participant
• Opens up possibility of tagging
• Editing / ‘Anonymisation’ framework
(General) Content

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

In practice;

• Influences how data analysed and shared
• Security during collection and handling paramount
  • More so for general perhaps – throughout data use
• Fewer options on reducing sensitivity for general
• Greater responsibility on researcher
Research Stakeholders

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

• 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’s handled
Consent

• 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”
Informed Consent

• *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/](http://www.admin.ox.ac.uk/curec/resources/informed-consent/)
  • [www.data-archive.ac.uk/create-manage/consent-ethics](http://www.data-archive.ac.uk/create-manage/consent-ethics)
Impact of Consent

• Informed Consent shapes the research process
  • How data collected
  • Approaches to analysis
  • Appropriate dissemination and publication
  • Data handling after project ends
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

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

Deposit your data

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

Recent blog posts
Data Storage

• Securely storing the data is key
• During research data is active / live
• Requires appropriate secure handling and storage
  • Use approved tools – One drive
  • Avoid common but unapproved tools – dropbox, email
  • Seek advice
Data Preservation

• Securely storing the data is key
  • But only one part

• In addition consider efficient access for you
  • Short term
  • Long term

• How to manage sensitive data
  • Moving material around
  • Honouring agreements made
  • Preserving the data for the future
Three Main Approaches

• Destruction
  • Requires good reasons
  • Wasteful

• Anonymisation
  • Clearly defined
  • Time consuming and imperfect

• Access Restriction
  • Leaves content intact
  • Needs active management
Data Destruction

• During or after a project
• Make a good case for this
  • Full or partial destruction?
• Satisfy stakeholders it is unavoidable
• Use appropriate tools
  • Eraser
  • DBAN
  • Disk Utility (Mac)
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
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
• UK Data Service guidance
Restricting Access

• Anonymisation allows wide access to less data (ie by removing content) post project

• An alternative approach is to leave content **but make access harder**
  • Vetting of access during a project
  • Require clear access and usage conditions when preserved
  • E.g. Microdata from Eurostat or UKDS etc.
  • Introduce embargoes (last resort)
Restricting Access

• Best used for **general content** confidentiality?
• Effective or credible policing of restrictions needed
• Both approaches can increase usage potential of data
• Require **planning from the beginning**
• Indicated in consent
Planning for Handling and Use

- Document the research process
  - Metadata captures decisions with clear requirements
  - How sensitive data will be managed and processed
- Pilot consent paperwork
- Think about what could go wrong!
  - Collecting unnecessary data
  - Hardware /software failure
  - Security – breaches - theft
  - Managing *accusations* of disclosure
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).
What next?

• Seek support and advice
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

Recent blog posts