Working with Sensitive or Confidential Research Data for MSD

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

Cathy Scutt
Subject Consultant for Education

Michaelmas Term 2017
Our Focus Today

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

• When is this a concern?
  • During creation and execution of a project
  • Afterwards
    • Solutions need to cover both

• What kind of data or outputs are you producing?
Consider Reasons

• Why confidential?
• Because of content
  • Particular Content
  • General Content
  • Which is more prevalent in MSD?
Particular Content

• Consent agreement or legislation define it
• Specific sections are confidential
• Opens up possibility of tagging
• Editing / ‘Anonymisation’
• Examples?
Handling Particular Content

In practice;

• Security during collection and handling paramount
• More options on reducing sensitivity of future versions
• ‘Safer’ preservation copy
  • Minor editing
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
- Examples?
Handling General Content

In practice;

• Security during collection and handling paramount
  • More so perhaps – throughout data use
  • Finishing with data destruction?

• Fewer options on reducing sensitivity of future (preserved) versions
Secure storage & 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
  - Storage becomes active preservation
Content and Stakeholders

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

• Framework in which data is used
  • Undertakings given to research participants
  • To research partners
  • To funder or institutional expectations
  • Publisher expectations

• Consent agreements or legislation play a role?
Protection and reassurance

• Intention is to protect and reassure all parties

• Who are the concerned parties?
  • Project participants – supplying the information
  • The audience – for the information and related analysis
    • Reproducibility
    • Verification
  • The researcher – who will gain or lose based on how it is used and received
  • Examples in MSD?
Consent Undertakings

• Help ensure participation BUT also define interests of stakeholders

• What is typical for MSD?
Consent Undertakings

• Avoid agreements that are too restrictive
  • “only to be used by this researcher” - “will be destroyed” - “no one else will read”

• Helpful but not too restrictive
  • Protection not a strait-jacket
  • Look for examples
    • Past projects
    • Oxford Curec – resources page
    • UKDA, ICPSR
Useful Informed Consent?

• Undertakings need to encourage participation
• Protect *everyone* involved
• BUT also produce usable data
  • Avoid unnecessary promises
  • [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)
  • Pilot/ trial your agreements
• Based on three traditions
Three Main Approaches

Managing data during *and* after a project

• Destruction
  • Straightforward
  • Contradiction of newer funder policy?
  • Requires good reasons
  • Wasteful

• Most common in MSD
  • True or false?
  • Examples
Three Main Approaches

Managing data during *and* after a project

- Anonymisation
  - Clearly defined
  - Additional work?
  - Time consuming and imperfect
  - Examples?
Three Main Approaches

Managing data during *and* after a project

• Access Restriction
  • Leaves content intact
  • Needs active management throughout
  • MSD examples?
Data Destruction

• Useful for all types of content or just general?
• Protocols in place to ensure done
• Clarity on documentation and working practices
• Demonstrable
  • Compliance with other research partners
  • Other benefits?
Blurring, Masking or Anonymisation

• During and after a project
• **Light** touch; limited key identifiers e.g. Names and addresses only
• Aggregation – fine grain detail/numbers removed
Blurring, Masking or Anonymisation

- 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
- MSD examples?
Access Restrictions

- Anonymisation allows wide access to less data (ie by removing content) post project
- Alternative approach leaves content but makes access harder
  - Requires clear access and usage conditions
  - Vetting of access
  - Restrict reproduction or analysis
  - Introduce embargoes
  - MSD examples?
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
All Approaches Need...

- Effective use and handling of data and metadata
- Document the research process
  - Metadata that captures decisions and clear requirements
  - How sensitive data is managed and processed
- Pilot consent paperwork
- Consider digital and analogue security
  - Virtual and physical security
Effective Handling and Use

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

• Build on this session
• 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

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