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
(June 2021)
Data Storage

• Securely storing the data is key
• During research data is active / live
• Requires appropriate secure handling and storage
  • Use approved tools – OneDrive for Business
  • 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
Legal Regulation - GDPR

- General Data Protection Regulation (GDPR)
- Addresses handling/processing of personal data
- Information Commissioners Office ICO definition of personal data
  - “If it is possible to identify an individual directly from the information you are processing, then that information may be personal data.”
GDPR Exemptions

• Non-commercial / Non administrative use
• “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).
Three General Approaches

Whether Personal, Confidential or Sensitive

• Destroy
• Anonymise
• Restrict
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)
• Requires good reasons
• Wasteful
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
- Requires **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
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

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