iSkills: Working with Sensitive or Confidential Research Data in the Social Sciences & Humanities

Michaelmas Term
November 2022

John Southall
Bodleian Data Librarian
john.southall@bodleian.ox.ac.uk

John Pilbeam
Web/Digital Officer
john.pilbeam@sbs.ox.ac.uk
Data Storage

- Securely storing the data is key
- During research when data is active / live
- Requires appropriate secure handling and storage
  - Use approved tools – OneDrive for Business
  - https://help.it.ox.ac.uk/which-onedrive
  - Avoid common but unapproved tools – Dropbox, email
  - Seek advice from department (local solutions) and RDO
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 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
Demonstrate Steps Taken

• Be clear on your security measures
  • Well Documented
  • Open to potential audits / inquiries

• Applies to all sensitive data
  • Created by project
  • Acquired from other sources
  • Covered by external agreements
  • Terms of use
  • Data Legislation
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.”
Legal Regulation - GDPR

• ... But notice the wording of the 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.”
• Consider this carefully and be prepared to defend your definition.
• Interpretation of the regulation in the context of wider RDM decisions
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?
• Is there an intended retention period for personal data?
  • “The storage limitation principle states that we must not keep data longer than necessary for the purposes for which it was collected.”
  • https://compliance.admin.ox.ac.uk/retention-schedules#collapse1098971
• Satisfy stakeholders it is unavoidable
Data Destruction

• Use appropriate/approved (by who?) tools
  • Eraser - Blancco - Disk Utility (Mac)
• Or data/ personal data to be retained in perpetuity (ie archived)
• Planned preservation
• Even for personal data? Using exemptions
• Other strategies and approaches?
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 on RDM and confidentiality
  • [www.icpsr.umich.edu/web/pages/datamanagement/index.html](http://www.icpsr.umich.edu/web/pages/datamanagement/index.html)
• UK Data Service guidance
  • [ukdataservice.ac.uk/learning-hub/research-data-management/](http://ukdataservice.ac.uk/learning-hub/research-data-management/)
What is research data?

The University of Oxford’s Policy on the management of research data and records defines research data as ‘recorded information (regardless of the form or media in which they may exist) necessary to support or validate a research project’s observations, findings or outputs.’

What is ORA-Data?

ORA-Data is the University of Oxford's digital catalogue and repository for research data, managed by the Bodleian Libraries. It offers a service to archive, preserve and enable the discovery and sharing of data produced by Oxford researchers.

Any type of digital research data – from across all academic disciplines – may be deposited in ORA-Data. We accept any file format.

ORA-Data is aimed at researchers who:

- wish to include an entry for their dataset in the University’s catalogue of research data, irrespective of where the data is archived
- need a repository to deposit research data – especially data that underpins publications, and data where the funding body requires archiving and preservation.

ORA-Data sits within the original Oxford Research Archive (ORA) for publications, so that data can be linked easily to – and browsed alongside – related publications.

Key features of ORA-Data

Deposit
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
- Requires a host archive to act on your behalf
Welcome to the Sustainable Digital Scholarship (SDS) Service

Based in the Humanities Division and working in close partnership with Digital Scholarship @Oxford, SDS is one of the University of Oxford’s research data management services, helping projects and researchers at Oxford to store, publish and preserve their research data outputs.

Find out more about the Sustainable Digital Scholarship service and what we can offer your project, or learn why digital sustainability matters for research.
Documenting as Planning

• Document the research process
  • Metadata captures decisions with clear requirements
  • How sensitive data will be collected and handled
  • How sensitive data will be managed, preserved or destroyed

• Embrace DMP, CUREC, DPIA and similar as tools to help this
Planning for Collection, Handling and Use

• Pilot consent paperwork
  • Does it protect you and participants?
• Think about what could go wrong!
  • Collect unnecessary data
  • Hardware/software failure
  • Security – breaches - theft
• Manage actual and accusations of disclosure
What next?

• Seek support and advice

• john.southall@bodleian.ox.ac.uk
• john.pilbeam@sbs.ox.ac.uk