iSkills: Working with Sensitive or Confidential Research Data

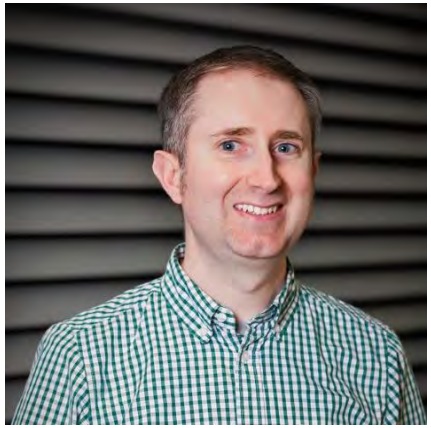
BODLEIAN LIBRARIES
UNIVERSITY OF OXFORD

Hilary Term

February 2024

BODLEIAN LIBRARIES

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



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

Data Storage

BODLEIAN LIBRARIES
UNIVERSITY OF OXFORD

- 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

Data Preservation

BODLEIAN LIBRARIES
UNIVERSITY OF OXFORD

- 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

BODLEIAN LIBRARIES
UNIVERSITY OF OXFORD

- 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

BODLEIAN LIBRARIES
UNIVERSITY OF OXFORD

- 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."
- https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/key-definitions/

Legal Regulation - GDPR

BODLEIAN LIBRARIES
UNIVERSITY OF OXFORD

- … 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

BODLEIAN LIBRARIES
UNIVERSITY OF OXFORD

- 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

BODLEIAN
LIBRARIES
UNIVERSITY OF OXFORD

- 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

BODLEIAN LIBRARIES
UNIVERSITY OF OXFORD

- 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

- "The Anonymisation Decision Making Framework" Elliot, Mackay et al (2016)

- Second edition is downloadable with additional templates and materials

https://ukanon.net/framework/

- UK Anonymisation Network https://ukanon.net/

Blurring, Masking or Anonymisation

BODLEIAN LIBRARIES
UNIVERSITY OF OXFORD

- Perhaps best used for **particular content**
  - Removing columns from spreadsheets
  - Specific names/words in transcripts
- Allows preservation in open access archives like ORA
- Dangers of data degradation or distortion
- ICPSR guidance on RDM and confidentiality
  - www.icpsr.umich.edu/web/pages/datamanagement/index.html
- UK Data Service guidance
  - ukdataservice.ac.uk/learning-hub/research-data-management/

BODLEIAN LIBRARIES
UNIVERSITY OF OXFORD

Subject and research guides

Depositing data in the Oxford Research Archive (ORA): Getting Started

Purpose of this guide

This guide is intended for students and researchers at the University of Oxford who wish to deposit their research data in the Oxford Research Archive (ORA).

Use this guide to find out about the University's policies regarding research data, the value of archiving completed research data, and how to deposit data in ORA..

Research data management

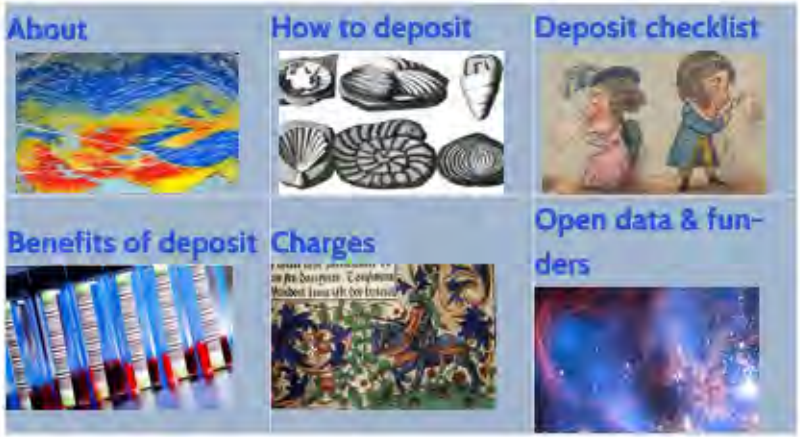
RESEARCH DATA OXFORD | UNIVERSITY OF OXFORD

Research Data Oxford is the University of Oxford's main portal for advice about research data management.

Data Seal of Approval

2014 DSA 2017

Oxford Research Archive - Data



Related Guides

Oxford Research Archive (ORA) for publications and theses

Oxford University Research Archive

ORA
OXFORD UNIVERSITY RESEARCH ARCHIVE

ORA (Oxford University Research Archive) is the institutional repository for the University of Oxford and is home to the scholarly output of its research members.

Contact us at: ora@bodleian.ox.ac.uk, or via our contact form

- https://libguides.bodleian.ox.ac.uk/ora-data

Restricting Access

BODLEIAN LIBRARIES
UNIVERSITY OF OXFORD

- 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, ONS or UKDS etc.
  - Or introduce time limited embargoed deposits (last resort)

Restricting Access

BODLEIAN LIBRARIES
UNIVERSITY OF OXFORD

- 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
- One application of this known as ‘The Five Safes’

UK Data Service

Login

Find data | Deposit data | Learning Hub | Training and events | About | News | Impact | Help | Contact

Home › Help › What is the Five Safes framework?

What is the Five Safes framework?

in **SecureLab**

The Five Safes framework is a set of principles which enable data services to provide safe research access to data. The framework originated from the ONS and was developed by them and other data providers in the 2010s. The framework has become best practice in data protection whilst fulfilling the demands of open science and transparency.

Five Safes was adopted more recently in 2020 by a range of other Trusted Research Environments (TREs) across the UK including Health Data Research-UK (HDR-UK) and National Institute for Health Research Design Service (NIHR).

Following the Five Safes, the UK Data Service Secure Lab provides Approved Researchers with controlled access to sensitive or confidential data, enabling researchers to access and use datasets in a secure and responsible way.

The Five Safes

Safe data: data is treated to protect any confidentiality concerns.

Safe projects: research projects are approved by data owners for the public good.

Safe people: researchers are trained and authorised to use data safely.

Safe settings: a SecureLab environment prevents unauthorised use.

Safe outputs: screened and approved outputs that are non-disclosive.

- https://ukdataservice.ac.uk/help/secure-lab/what-is-the-five-safes-framework/

Five safes

From Wikipedia, the free encyclopedia

The **Five Safes** is a framework for helping make decisions about making effective use of data which is confidential or sensitive. It is mainly used to describe or design research access to statistical data held by government agencies, and by data archives such as the UK Data Service.[1]

Two of the Five Safes refer to statistical disclosure control, and so the Five Safes is usually used to contrast statistical and non-statistical controls when comparing data management options.

Contents hide

Concept [edit]

The Five Safes proposes that data management decisions be considered as solving problems in five 'dimensions': projects, people, settings, data and outputs. The combination of the controls leads to 'safe use'. These are most commonly expressed as questions, for example:[2][3]

Safe projects	Is this use of the data appropriate?
Safe people	Can the users be trusted to use it in an appropriate manner?
Safe settings	Does the access facility limit unauthorised use?
Safe data	Is there a disclosure risk in the data itself?
Safe outputs	Are the statistical results non-disclosive?

These dimensions are scales, not limits. That is, solutions can have a mix of more or fewer controls in each dimension, but the overall aim of 'safe use' independent of the particular mix. For example, a public use file available for open download cannot control who uses it, where or for what purpose, and so all the control (protection) must be in the data itself. In contrast, a file which is only accessed through a secure environment with certified users can contain very sensitive information: the non-statistical controls allow the data to be 'unsafe'. One academic likened the process to a graphic equalizer,[4] where bass and treble can be combined independently to produce a sound the listener likes.

There is no 'order' to the Five Safes, in that one is necessarily more important than the others. However, Ritchie[5] argued that the 'managerial' controls (projects, people, setting) should be addressed before the 'statistical' controls (data, output).

The Five Safes concept is associated with other topics which developed from the same programme at ONS, although these are not necessarily implemented. Safe people is associated with 'active researcher management',[6] while safe outputs is linked with principles-based output statistical disclosure control.

The Five Safes is a positive framework, describing what is and is not. The FDRU ('evidence-based, default-open, risk-managed, user-centred')

Criteria for selecting a data archive

Options for preserving your data

While archives are generally the preferred option, in some cases, researchers may find that no suitable archive is available, or the data is subject to particular regulations concerning preservation and sharing which restrict where it can be deposited. The sections below therefore explore both archives and some alternatives.

If you would like to talk about selecting the most suitable option for your own data, please contact the Research Data Oxford team by emailing researchdata@ox.ac.uk.

+ Expand All

University of Oxford options

ORA: the University of Oxford's repository for research outputs, including data

The Oxford Research Archive (ORA) is an archiving service provided by the University of Oxford. It also functions as a catalogue of data produced by Oxford researchers and deposited either in ORA or elsewhere.

ORA accepts data from any discipline, and especially data that underpins publications. It can provide a home for datasets that must be deposited to comply with a funder's policy, but where there is no suitable national or discipline-specific archive. However, it is currently unable to accept deposits of sensitive or non-anonymised personal data.

ORA preserves stable versions of data and can assign DOIs to data collections if desired, making them citable. Each collection has a freely available online record, to aid data discovery. Data creators can assign rich metadata to their dataset, allowing them to meet funder and publisher requirements, and to receive proper credit and acknowledgement for their work.

ORA does not aim to hold all research data produced by Oxford researchers: it will co-exist with disciplinary and general archives. However, researchers depositing data elsewhere are strongly encouraged to create at least a metadata record in ORA.

DigiSafe

DigiSafe is an opt-in subscription service designed to provide secure storage for data which needs to be preserved for short or long periods, typically a year or longer. It has strong features for adding metadata and preserving access to file formats even when the original software used to create the data is no longer available. Data access is comprehensively logged and there is regular integrity checking of all data on the platform. Jupyter notebooks can be run to analyse data directly on the platform.

It is most useful for categories of research data which are not suitable for sharing (for example, identifiable participant records from medical research projects). Stored data can be easily searched and retrieved by users with the appropriate permissions. Built-in functions allow easy management of retention schedules - where material has to be deleted after a set amount of time, for example. DigiSafe is offered on a subscription basis to departments, colleges, and other units, so access to the service is dependent on whether your unit has opted to subscribe. Individual research groups who have secured funding are also welcome to sign up for the services in their own right. Whilst data can be shared directly from the platform, the functionality is quite basic.

Sustainable Digital Scholarship service

The Sustainable Digital Scholarship service is designed to allow researchers to store, work with, preserve, and share research data. The SDS platform, which is provided by Figshare, can be used both for collecting and editing data, and as a way of keeping research data safe for the long term and making it available to a wider public.

The service launched in the Humanities Division, but is available to researchers from across the University. Support and hosting are available free of charge to most pre-existing research projects seeking a more sustainable long-term home. For new projects which have not yet applied for funding, charges apply: quotations can be provided for support and hosting. These fees are only applicable during the funded phase of the project: once the active research period concludes, the data will be maintained on the system indefinitely without further charge.

Departmental data stores

Some Oxford departments have well established data stores that have served their research groups for a significant time. Because these are locally maintained, provision varies greatly: consult your local IT support staff to find out if your department is able to offer long term data storage. A departmental data store may be a good option in some circumstances (for example, if data needs to be preserved but is unsuitable for sharing, if you have very specialist requirements, or if

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.

SDS service in Numbers

Projects supported	107
Total items	584,892
Total views	1,872,891
Total downloads	211,601

Metadata and Planning

BODLEIAN LIBRARIES
UNIVERSITY OF OXFORD

- 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

BODLEIAN LIBRARIES
UNIVERSITY OF OXFORD

- Pilot consent paperwork
  - Does it protect you and participants?
- Think about what could go wrong!
  - Collect unnecessary data
  - Hardware /software failure
  - Security – breaches - theft
- Put in place procedures to manage accusations of disclosure (actual or mistaken)

What next?

BODLEIAN
LIBRARIES
UNIVERSITY OF OXFORD

- Seek support and advice

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