
UK Data Service: Data Access Categories

Matthew Woollard

COPAFS

Washington DC

30 January 2019

UK Data Service



Our organisation: the UK Data Archive

- Department of the **University of Essex**
- Founded 1967 by ESRC as a '**Data Bank**' for science science
- **Over 50 years** of curating and providing access to data for research and teaching
- A Trusted Digital Repository (TDR) and accredited to ISO27001 Info Security Management standard
- **Lead the national UK Data Service**
- Work closely with **research funders** and **data producers**



Types of data held

Social survey

One off large scale survey e.g. Health Survey for England

Longitudinal/cohort

Detailed surveys following people over time, e.g British birth cohorts

Qualitative

Interviews, focus groups, diaries

Aggregate data

UK census counts/tables, country level statistics

Historical data

Digitised databases



Funders and data sharing: ESRC

- ESRC --- robust Data Policy – *Since 1990s*
 - optimised model, matured over time
- Data Management Plan required with applications
- Support data infrastructure and a support service
 - Applicant and grant holder support/monitoring
 - Advocacy and early intervention work
 - Explicit strategy for non-compliance
- UK Data Service - mutual working relationship with data owners



Basic data publishing requisites

Findable

Accessible

Interoperable

Re-usable

Persistent identification of data collections



UK Data Service: selection, appraisal and management of data

- Selection managed through a robust and auditable appraisal and selection process
- Data acquisition managed through a rights framework
- Ability to manage different access conditions through a robust legal and technical framework
- Ability to store, curate and host data through a trusted digital repository framework (CoreTrustSeal, ISO 27001)



Some statistics about our UK Service

- **c.7,500** datasets in the collection
- **c.250** new datasets and new editions added each year
- **c.30,000** registered users
- **c.60,000** downloads worldwide p.a
- **>100+** applications for restricted data p.m.
- **c.4000+** user support queries p.a



UKDS strategies for enabling safe access to data

- ✓ **Protection of identities** when promised
- ✓ **Processing ground** for personal data
- ✓ **Regulated access** where needed (all or part of data) by group, use, time
- ✓ **Safeguards and security**

Open where possible, closed when necessary

Incoming dataset assessment at UKDS

- All data coming in **may be potentially be personal**
- Staff processing data have NDA and DBS
- ISO 27001 Information Security Management
- Security protocols for processing spaces/secure servers

- Check consent status – forms /information sheets
- **Assessment of whether data are disclosive or not**
- **Treat (minimise) data for widest possible audience**
- Ensure appropriate legal gateways /safeguards in place for access to identifiable data
- **Accountability, transparency and documentation**



Spectrum of Access: UK Data Service

Open

- No disclosure risk
- Open licence; few restrictions on reuse

Safeguarded

- Zero to low disclosure risk
- User agreement, optional extra safeguards
- Authentication, authorisation & auditing

Controlled

- Disclosure risk/personal data/legal gateway
- Authentication, authorisation & auditing
- Added safeguards (Five Safes)

Generic Three Tier Access Mechanism

Open

Safeguarded

Controlled



Variables and geography ensure impossibility of disclosure	Most data, limited geographies	All data, complete geographies
Any purpose	Research purposes only	Approved Research
Open	Safeguarded	Controlled
No Restrictions (Open Licence)	Some restrictions (EUL)	Many restrictions and training required (EUL+)
Not applicable	Researcher's responsibility	Complete checking



UK Data Service





Safe Settings	DISTRIBUTE DATA		DISTRIBUTE ACCESS	
Safe Data	Variables and geography ensure impossibility of disclosure	Most data, limited geographies	All data, complete geographies	
Safe Projects	Any purpose	Research purposes only	Approved Research	
	Open	Safeguarded	Controlled	
Safe People	No Restrictions (Open Licence)	Some restrictions (EUL)	Many restrictions and training required (EUL+)	
Safe Outputs	Not applicable	Researcher's responsibility	Complete checking	



UK Data Service



Five Safes in terms of Access Control

- Safe Settings [Distribute Data => Distribute Access]
- Safe Data [Appropriately controlled for disclosure risk]
- Safe Projects [Appropriate use of data]
- Safe People [Appropriately trained users]
- Safe Outputs [Appropriate levels of disclosure control]



The challenges of personal identification

- Identifiability of personal information is on a scale (**Anonymous** – **Identifiable**) but the **law is binary!**
- **Content:** the strength of identifiers
- **Context:** the access environment
 - Who can access the data?
 - What other information might a user access?
 - Is the risk adequately controlled?



The problem with classifying risk

- Data access must involve **reduction of risk** in a manner acceptable to the data owner
- Risk is mitigated by legal gateways and appropriate safeguards
- We seek to control **context** making identification **not 'reasonably likely' by any likely means**
- ICO: '**Reasonable likelihood**' is the test vs. hypothetical risk



Safeguards: use of the Fives Safes Framework

Fives Safes enables **safe access to data** that meet the needs of **data protection** yet fulfils the demands for **open science and transparency**

- **Safe data** - treat data to protect confidentiality
- **Safe people** - educate researchers to use data safely
- **Safe projects** - research projects for 'public good'
- **Safe settings** – Secure Lab environment for personal data
- **Safe outputs** – Secure Lab projects outputs screened

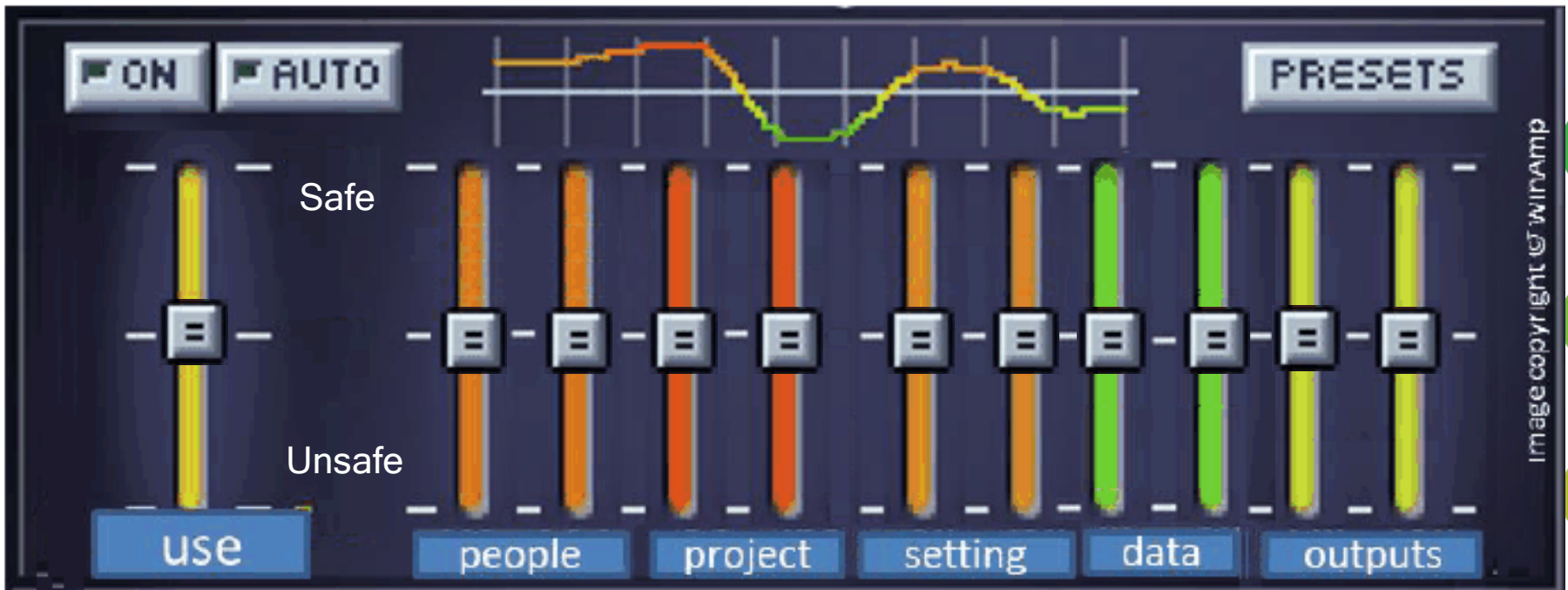
[5 Safes Animation](#)



UK Data Service



The Five Safes in operation...



Adapted by the Australian Bureau of Statistics from 'Five Safes: designing data access for research', Desai, T, Ritchie, F and Welpton, R, 2016

Access points at UK Data Service

A study can be deposited and made available as **multiple datasets** under **different access conditions**

DATA ACCESS

- + [GN 33004 | NATIONAL CHILD DEVELOPMENT STUDY, 1958-](#)
- + [GN 33395 | NATIONAL CHILD DEVELOPMENT STUDY: SPECIAL LICENCE ACCESS](#)
- + [GN 33497 | NATIONAL CHILD DEVELOPMENT STUDY, 1999-: SECURE ACCESS](#)

Controlled Access: **detailed geographies of locations and variables deemed too sensitive** for standard release

Consent forms wording – legacy studies

Processing personal data in the spirit of ‘reasonable expectations’

Tests for reasonable expectations

- Is the intended processing compatible with the consent gained and promises made?
- Implications of wording for future work/prospective studies
- Would the participants have a reasonable understanding of what was going to happen in the future?

ukdataservice.ac.uk



UK Data Service

Home > Discover data > Data access policy

Data access policy

The UK Data Service is implementing a generic, three tier access policy. The three tiers are:

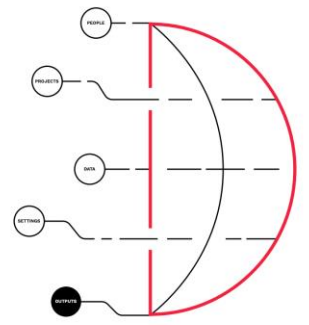
- **open**
- **safeguarded**
- **controlled**

These categories are generic because they combine modes of access and conditions of use, and there are likely to be overlaps and differences between the three major categories.

Legal definitions Open data Safeguarded data Controlled data

Open data and safeguarded data are legally 'not personal' according to the relevant legislation, e.g. Data Protection Act and Statistics and Registration Services Act. The difference between 'open' and 'safeguarded' is that safeguarded data may have a residual risk of disclosure, and open data do not. Controlled data are data which can be defined as personal.

Handbook on Statistical Disclosure Control for Outputs



UK Data Service

Home > Manage data

Prepare and manage data

"Good data habits from the moment you start planning your research"

Download OUR DATA GUIDE: Comprehensive best practice guidance for researchers on managing and sharing data

QUICK ACCESS TO: FAQ about managing data

RELATED LINKS: UK Data Archive

UK Data Service August 2018 Version 08.00



User guide

For your Secure Lab account



Secure Lab services

Enabling secure research access to data while protecting confidentiality

The Five Safes Framework

- Safe data
- Safe projects
- Safe people
- Safe settings
- Safe outputs

As promoted by the Office for National Statistics

QAMyDATA GUIDE: HOW TO INSTALL AND RUN



QAMyData User Guide

ukdataservice.ac.uk
help@ukdataservice.ac.uk





@UKDataService

@UKDSRDM

#sdc-microdata

Contact:

matthew@essex.ac.uk

UK Data Service

