Research ethics and data sharing: governance and integrity

Louise Corti

Service Director, Collections Development and Data Publishing

UKRIO Annual Conference 2018
London
23 May 2018

UK Data Service

Overview

- The role of data in transparency in research
- Explore some of the tensions that arise between the concepts of data sharing and data protection
- Protocols and governance measures to share data safely while meeting the needs of research transparency
- Consent forms in research not precluding data sharing

To Share or Not to Share?

Consider a research team in your institution has conducted a detailed survey on incidences of violence on women

Can they share the data and what are the key issues?

Opportunities

Problems

Research Integrity through data publishing

- Open Science and replication agendas /directives
- Demonstrating value and transparency through
  - Publishing methods, code and data – small or big data project PROCESS TRANSPARENCY
  - Crediting authors of data sources through persistent referencing DATA CITATION

Perceptions and misconceptions

- Common perception: it is hard to reconcile data sharing vs. data protection
- Common misconception: Open Science = Open Data

UK Open Science agendas

- OECD Principles and Guidelines for Access to Research Data from Public Funding
- Expert Advisory Group on Data Access (EAGDA)
- Transparency Organizations (Centre for Open Science)
- Data Management and Use: Governance the 21st century– British Academy and Royal Society
- Universities UK Concordat on Research Integrity
- Welcomed campaigns: AllTrials, UKRIO
- FAIR data

  Findable Accessible Interoperable Re-usable

- Stewardship, equitable access, open publishing

Research funder policies

- Most based on the OECD Principles and Guidelines for Access to Research Data from Public Funding

- Variety of models
  - Data management plans and recommendation only
  - Dedicated data centres
  - Institutions taking responsibility
  - Some with penalties

- UK Data Service co-wrote and manages the ESRC Data Policy and acts as the data sharing advocate/negotiator/repository. Since 1995

Journal / Publisher Data Policies

- Most science journals have data sharing policies

  Science, Nature, PLOS ONE, BioMed Central

- Slower in social science

  Psychology, economics and political science

  ✓The Journal of Development Economics
  ✓Quarterly Journal of Economics
  ✓Quarterly Journal of Political Science

- Data underpinning publication accessible:
  - upon request from author
  - supplementary materials alongside publication
  - journal own repository, e.g. https://dataverse.harvard.edu/dataverse/ajps
  - public repository or specialist data centre

The case of political science journals

- Data Access and Research Transparency (DA-RT): A Joint Statement by Political Science Journal Editors

  "Journal editors commit their respective journals to the principles of data access and research transparency, and to implementing policies requiring authors to make as accessible as possible the empirical foundation and logic of inquiry of evidence-based research"

- 26 journal signed up by 2016
- Some established rules for implementation
- Offer guidance on preparing replication data
- Offer *Dataverses* for hosting replication data

Issues to consider for transparency

- Locating the data – which data? Is it FAIR?
- Are there ethical/legal issues with accessing/reusing the data?
- Is there sufficient information to reconstruct the analysis dataset
- Is the research and data documentation good enough?

Requires well-prepared, well-documented and ethically and legally-accessible research and data

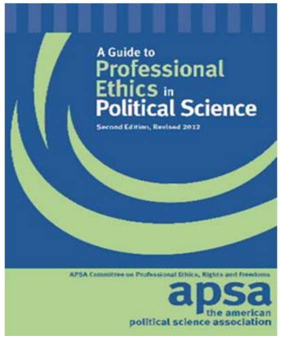
UKDS about to produce a Practical Guide for Meeting Transparency Demands in Qualitative Research

Professional bodies ethical guidance on data sharing

UKDS recent review of data sharing recommendations in Professional Bodies research ethics guidance in the social sciences

Examples of good practice

- Association of Internet Researchers, *Ethical Decision Making and Internet Research* (2012)
- British Society of Gerontology refers to *Guidelines on ethical research with human participants* (2012) - clearly points to archiving data and its secondary use as "a way of doing something useful for posterity" (p. 2)
- American Political Science Association, *A Guide to Professional Ethics in Political Science (2014)*



A Guide to Professional Ethics in Political Science

Second Edition, Revised 2012

APSA Committee on Professional Ethics, Rights and Freedoms

apsa the american political science association

Current progress..

- British Psychological Society *Code of Human Research Ethics (2014)*
- No explicit reference to sharing research data, principle of open access or research transparency
- Six separate mentions of data destruction! (p. 9, 15, 19 and 30)

Ethical arguments *for* sharing data

- Not burden over-researched, vulnerable groups
- Make best use of hard-to-obtain data, e.g. elites, socially excluded, over-researched
- Extend voices of participants
- Provide greater research transparency

In each, ethical duties to participants, peers and public may be present

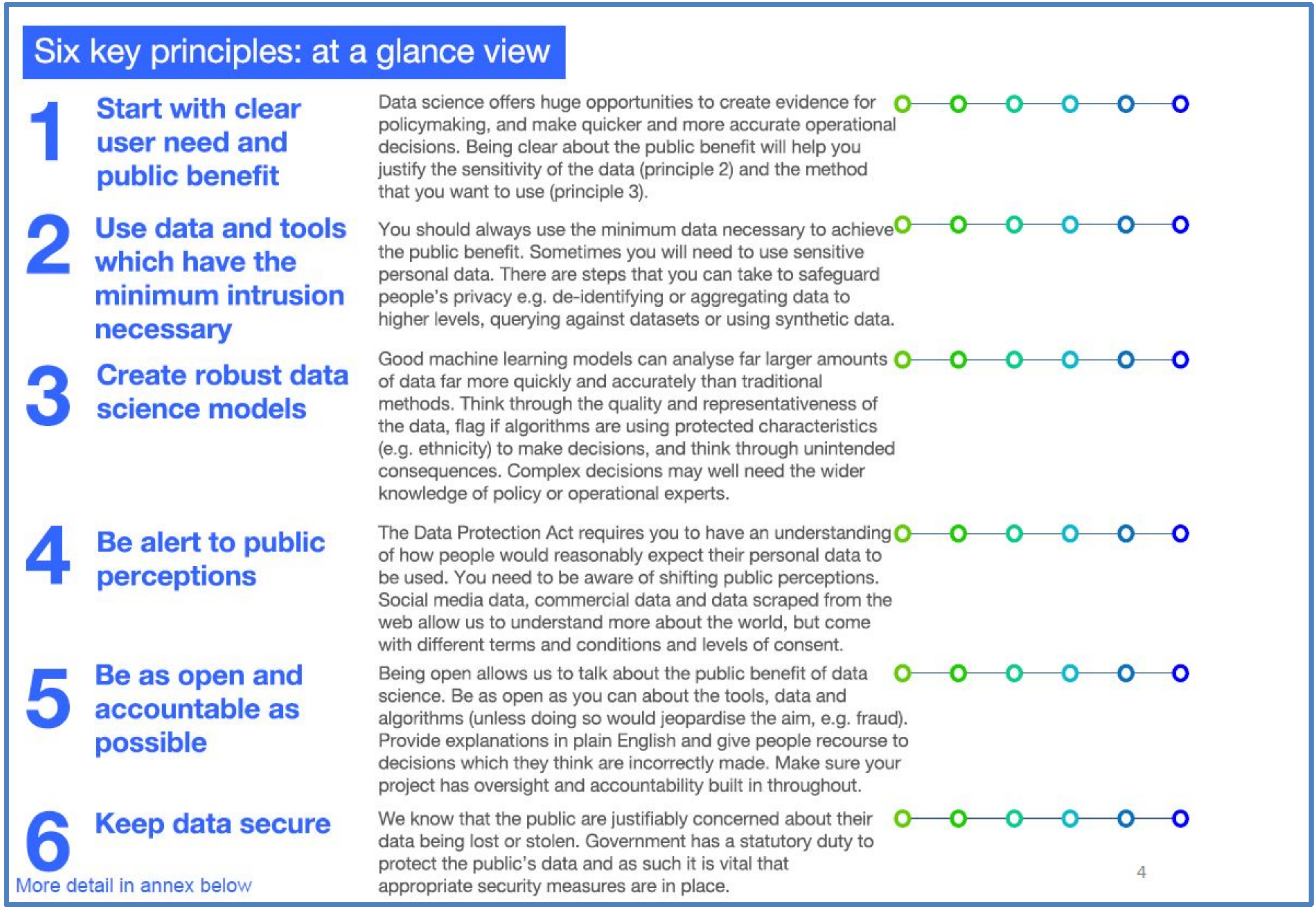
Ethical, legal and research integrity challenges

- Concerns about appropriate re-use of data

- Personal and confidential information
  - identifiers can be difficult to conceal e.g. identity of participants /fieldwork locations
  - risk of disclosure increases as data are linked e.g. social with geo-located data, administrative or biomedical information
  - big data – unknown provenance

- Efforts and funding for preparing data for sharing compete with ‘the science

UK Cabinet Office Data Science Ethical Framework



Six key principles: at a glance view

1 Start with clear user need and public benefit

Data science offers huge opportunities to create evidence for policymaking, and make quicker and more accurate operational decisions. Being clear about the public benefit will help you justify the sensitivity of the data (principle 2) and the method that you want to use (principle 3).

2 Use data and tools which have the minimum intrusion necessary

You should always use the minimum data necessary to achieve the public benefit. Sometimes you will need to use sensitive personal data. There are steps that you can take to safeguard people's privacy e.g. de-identifying or aggregating data to higher levels, querying against datasets or using synthetic data.

3 Create robust data science models

Good machine learning models can analyse far larger amounts of data far more quickly and accurately than traditional methods. Think through the quality and representativeness of the data, flag if algorithms are using protected characteristics (e.g. ethnicity) to make decisions, and think through unintended consequences. Complex decisions may well need the wider knowledge of policy or operational experts.

4 Be alert to public perceptions

The Data Protection Act requires you to have an understanding of how people would reasonably expect their personal data to be used. You need to be aware of shifting public perceptions. Social media data, commercial data and data scraped from the web allow us to understand more about the world, but come with different terms and conditions and levels of consent.

5 Be as open and accountable as possible

Being open allows us to talk about the public benefit of data science. Be as open as you can about the tools, data and algorithms (unless doing so would jeopardise the aim, e.g. fraud). Provide explanations in plain English and give people recourse to decisions which they think are incorrectly made. Make sure your project has oversight and accountability built in throughout.

6 Keep data secure

We know that the public are justifiably concerned about their data being lost or stolen. Government has a statutory duty to protect the public's data and as such it is vital that appropriate security measures are in place.

More detail in annex below

GDPR what’s new for data sharing?

- Principles, rights of data subjects and processing grounds for processing personal data similar but much more explicit than under DPA
- Emphasis on transparency, clear information, clear documentation
- Extra rights of subjects: erasure, data portability, automated profiling
- Reuse for research allowed with safeguards

Applies only to ‘personal data’, not anonymised or de-identified data

Data Protection legislation is not intended to, and does not, inhibit ethical research

GDPR research exemption - archiving

Further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes is not considered to be incompatible with the initial purposes

Appropriate safeguards, e.g.

- data minimisation , pseudonymisation

Two Principles less strict:

- Purpose: further processing of personal data allowed (2)
- Personal data may be stored for longer periods (5)

Transparency when collecting personal data

	Personal data obtained directly from participants	Personal data obtained indirectly
Name and contact details of data controller (entity that determines the reason for processing personal data) and data protection officer	✓	✓
Purposes of the processing and legal basis	✓	✓
Categories of personal data concerned		✓
Who will receive or have access to the personal data	✓	✓
How long the personal data will be stored	✓	✓
The data subject's rights (access, correction, removal,…)	✓	✓
The right to lodge a complaint with the ICO	✓	✓
Source from which the personal data originate, and if applicable, whether it came from publicly accessible sources		✓
Whether providing personal data is part of a statutory or contractual requirement or obligation and possible consequences of failing to provide the personal data	✓	
Any automated decision-making, and, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject	✓	✓
Safeguards for transfers out of Europe	✓	✓

rvice

Informed consent and research data

- Identify and explain possible future uses of data
- Offer participant option to freely consent on a granular level
- Must be done across the research lifecycle e.g. for new types of data collection in a longitudinal study
    - e.g. participant to consent (or not) separately for survey data, bloods, and so on
- Indicate steps that will be taken to safeguard anonymity and confidentiality
- Consent must be documented
- Future governance for data access should recognise these consents

In practice: wording in consent form / information sheet – interviews, photos

We expect to use your contributed information in various outputs, including a report and content for a website. Extracts of interviews and some photographs may both be used. We will get your permission before using a quote from you or a photograph of you.

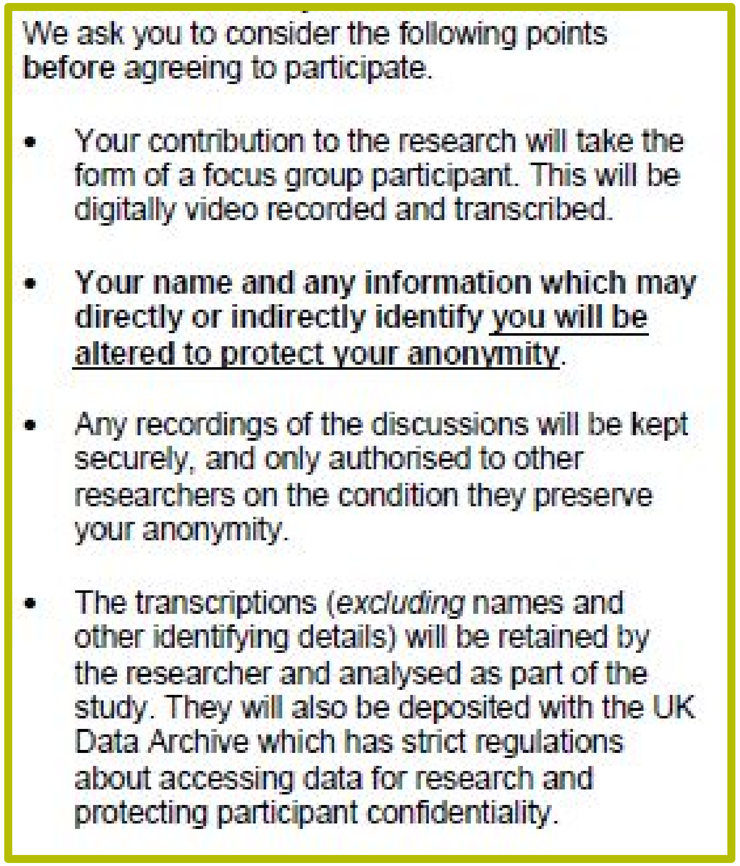
After the project has ended, we intend to archive the interviews at …. Then the interview data can be disseminated for reuse by other researchers, for research and learning purposes.

The interviews will be archived at ……. and disseminated so other researchers can reuse this information for research and learning purposes:

- ❑ I agree for the audio recording of my interview to be archived and disseminated for reuse
- ❑ I agree for the transcript of my interview to be archived and disseminated for reuse
- ❑ I agree for any photographs of me taken during interview to be archived and disseminated for reuse

In practice: wording in consent form / information sheet – focus group

Any personal information that could identify you will be removed or changed before files are shared with other researchers or results are made public.



We ask you to consider the following points before agreeing to participate.

- Your contribution to the research will take the form of a focus group participant. This will be digitally video recorded and transcribed.

- **Your name and any information which may directly or indirectly identify you will be altered to protect your anonymity**.

- Any recordings of the discussions will be kept securely, and only authorised to other researchers on the condition they preserve your anonymity.

- The transcriptions (*excluding* names and other identifying details) will be retained by the researcher and analysed as part of the study. They will also be deposited with the UK Data Archive which has strict regulations about accessing data for research and protecting participant confidentiality.

ukdataservice.ac.uk/manage-data/legal-ethical/consent-data-sharing/consent-forms.aspx

ESRC research data sharing problems

The UK Data Service regularly finds a mismatch between:

- intended plans for archiving/publishing data set out/and often promised in a DMP
- and the actual consent agreements in use in the field, that are often prohibitive
- Role of Research Offices to help with this problem (deception?) - better scrutiny up front needed

Practical challenges

- Managing different forms of consent for different materials, e.g. audio recordings, transcripts, photos
- Special cases of consent, e.g. children etc.
- Right to withdraw – what to do with already collected data?
- Informed consent for ‘unknown future data uses’

Solution for data sharing/archiving

- Is it personal data or not?
- Provide maximum information to participants about re-use
    - Benefits of data sharing in lay terms
    - Who can access which data and how – authenticated researchers, End User Agreements
    - Purposes of reuse – research, teaching, both
    - Confidentiality protections in place
- Set up appropriate safeguards…

Pathways to access for data: The 5 Safes framework

Strategies for enabling safe access to data

- ✓ Informed consent for long-term data sharing
- ✓ Protection of identities when promised
- ✓ Regulated access where needed (all or part of data) by group, use, time
- ✓ Securely store personal data (separately)

Open where possible, closed when necessary

Data Access Policy: our spectrum

Open	• No risk. Under open licence; almost no restrictions on reuse
Safeguarded	• Zero to low risk. Requires authentication and authorisation e.g. registered user and End User Agreement
Controlled	• Risk. Requires project approval, user vetting and training; access via a safe setting; output checking

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
  Become an **Approved Researcher** (ONS, DEA)
- Safe Projects Research projects assessed for ‘public good’
- Safe Settings Safe haven/Secure Lab environment for accessing personal data
- Safe Outputs Secure Lab projects outputs screened

5 Safes Animation

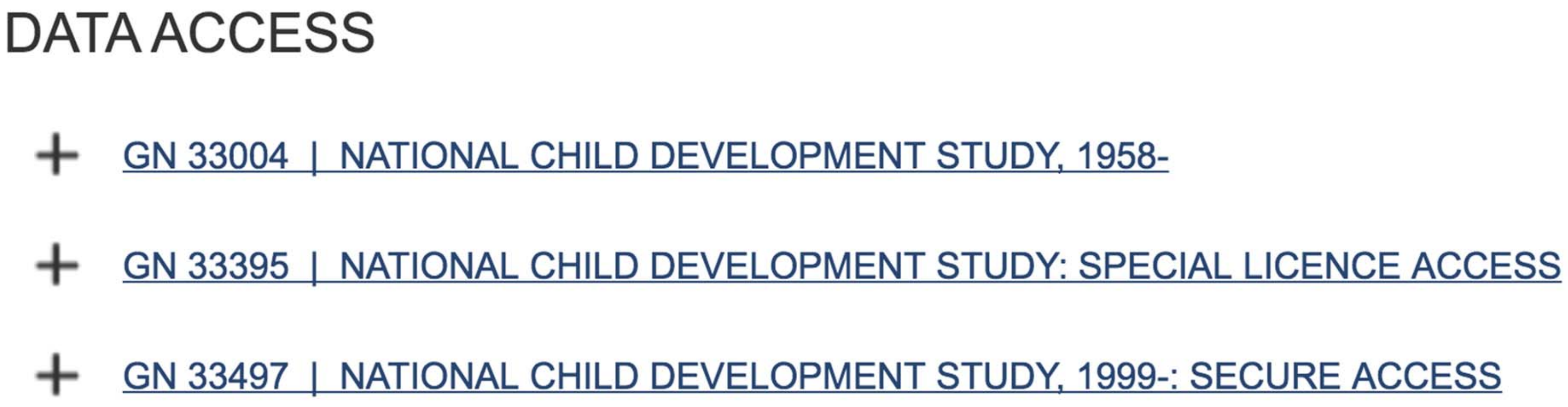
How ‘safe’ is Safe?

- When it comes to data, Safe is a relative term
- Involves reduction of risk in a manner acceptable to the data owner
- 5 Safes is therefore a balancing act of risk:
  - Open Data needs no other Safes to be put in place
  - Personal Data needs other Safes to be implemented

- ‘Treat’ data to maximise opportunities for reuse
- Use standard licence(s) for data
- Use a clear data access policy – spectrum of access
- Set up methods for enabling 5 Safes access

Access points at UK Data Service

One survey 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 respondents' locations and variables deemed too sensitive for standard release. Demonstrable research need for this more detailed data

User Access Agreements: principles

Security and storage

- No attempt to identify, individuals, households or organisations in the data
- Keep access to data secure
- Destroy data after use using best practice

Research Integrity

- Accept data are without warranty but report errors found
- Use under conditions agreed e.g. non-commercial only where specified
- Report non-compliance
- Do not republish without permission; but deposit back any derived data
- Correctly cite data using the persistent identifier provided
- Accept that non-compliance will lead to penalties

Service Usage

- Agree that personal data are correct and will be used for purposes specified

Key points for Research Ethics Committees

Better manage perceived tensions between data sharing and data protection

- Promote ethically sound research
- Ensure that research complies with data protection legislation
- Help distinguish between sharing of ‘personal data’ and ‘research data’
- Provide information to researchers at the consent and planning stages on how to share data ethically
- Include data sharing and data publishing in Ethical review procedures; and review of DMP
- Check that consent forms allow for data sharing, whilst also protecting the confidentiality of participants
- Help mediate in reconciling data sharing and data protection

Data sharing – a shared responsibility

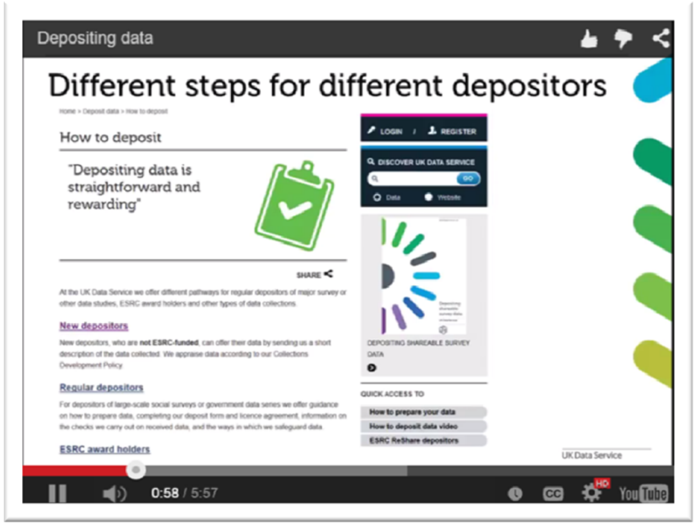
- Funders: policies, mandates, infrastructure funding
- Institutions: provide a supporting framework
  - Grants, research integrity framework, RECs, IT, data storage, RDM guidance and training
- Departments/centres: local support/infrastructure
- Funded researchers: create, manage, use, cite data
- Professional Societies: guidance on data sharing – focus on benefits, scientific rigour, ethics
- Advocacy and regulatory bodies

Resources

- UKDS: https://www.ukdataservice.ac.uk/manage-data/
- UKDS legal and ethical issues: https://www.ukdataservice.ac.uk/manage-data/legal-ethical
- UKDS research ethics review: https://www.ukdataservice.ac.uk/manage-data/legal-ethical/obligations/research-ethics-review
- *Managing and Sharing Research Data book*

  https://uk.sagepub.com/en-gb/eur/managing-and-sharing-research-data/book240297

ukdataservice.ac.uk



Depositing data

Different steps for different depositors

Home > Deposit data > How to deposit

How to deposit

"Depositing data is straightforward and rewarding"

SHARE

At the UK Data Service we offer different pathways for regular depositors of major survey or other data studies, ESRC award holders and other types of data collections.

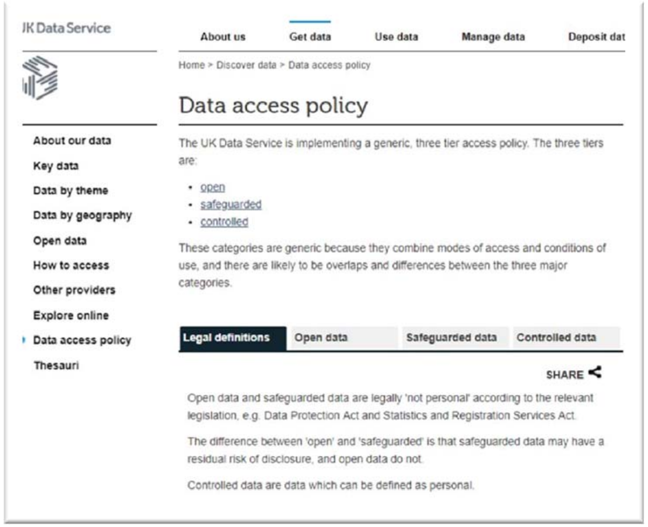
New depositors

New depositors, who are **not ESRC-funded**, can offer their data by sending us a short description of the data collected. We appraise data according to our Collections Development Policy.

Regular depositors

For depositors of large-scale social surveys or government data series we offer guidance on how to prepare data, completing our deposit form and licence agreement, information on the checks we carry out on received data, and the ways in which we safeguard data.

ESRC award holders



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:

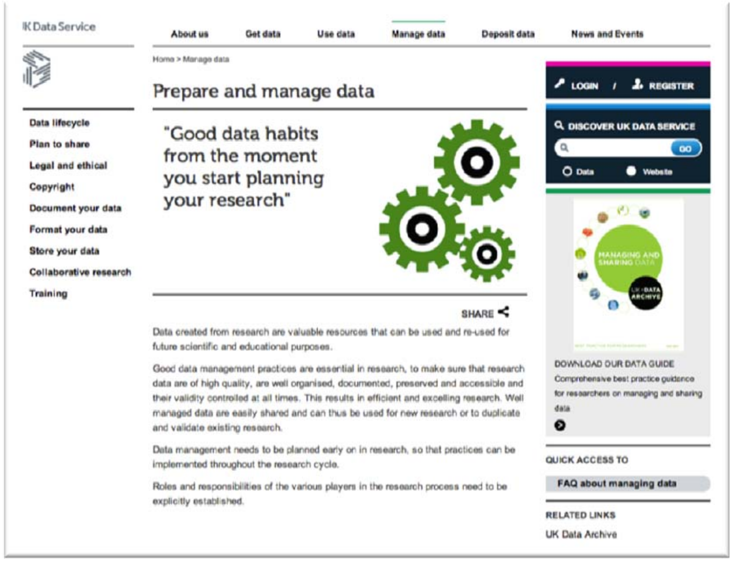
- open
- safeguarded
- controlled

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.

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.



Home > Manage data

Prepare and manage data

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

SHARE

Data created from research are valuable resources that can be used and re-used for future scientific and educational purposes.

Good data management practices are essential in research, to make sure that research data are of high quality, are well organised, documented, preserved and accessible and their validity controlled at all times. This results in efficient and excelling research. Well managed data are easily shared and can thus be used for new research or to duplicate and validate existing research.

Data management needs to be planned early on in research, so that practices can be implemented throughout the research cycle.

Roles and responsibilities of the various players in the research process need to be explicitly established.

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

What is the UK Data Archive?

- UK Data Archive – manages the UK Data Service and curates and manages the data. Since 1967
- Single point of access to a wide range of secondary social science data
- Put together a collection of the most valuable data
- Make data and documentation available for reuse
- Data and support for research, teaching and learning
- Speciality in social survey data, qualitative data, historical databases, some biomedical and big data!
- Funded by the Economic and Social Research Council

Summary: sharing data safely

- Collect data ethically and legally
- ‘Treat’ data to maximise opportunities for reuse
- Use standard licence(s) for data
- Use a clear data access policy – spectrum of access
- Set up methods for enabling appropriate access – 5 Safes
- Make application processes fair and transparent
- Use standard end user agreements
- Enable access to personal data through appropriate legal gateways
- Provide accountability across the access lifecycle

Grounds for processing personal data

- One of these must be present to process a data subject's personal data:
- Consent of the data subject
- Necessary for the performance of a contract
- Legal obligation placed upon controller
- Necessary to protect vital interests of the data subject
- Carried out in the public interest or is in the exercise of official authority
- Legitimate interest pursued by controller

Contact

UK Data Service
University of Essex
Ukdataservice.ac.uk

corti@essex.ac.uk
@ukdataservice
@ukds-rdm
@LouiseCorti