
Licensing and governance for research data

Louise Corti and John Sanderson
UK Data Service
UK Data Archive, University of Essex

Webinar
18 April 2018

UK Data Service



Introductions



Louise Corti
Director, Collections
Development and Data
Publishing



John Sanderson
Research Support
Manager, Administrative
Data Service

Acknowledgements

Thanks to our colleagues at the UK Data Service for providing input into today's webinar content:

- Laura Beauchamp, Data Access Officer
- Susan Cadogan, Senior Collections Development Officer
- Scott Summers, Senior Research Data Services Officer
- Matthew Woollard, Director



What we will cover today

1. Core principles of data publishing
2. Legal and ethical issues surrounding sharing data
3. Pathways to access for data: The 5 Safes framework
4. Legal documents: Data licences and access agreements
5. Practicalities of operating governance for data access



1. Core principles of data publishing





R3data by subject

UK Data Service



Basic data publishing requisites

- ✓ Usable format
- ✓ Long-term preserved/backed-up
- ✓ Self-explanatory for users
- ✓ Non disclosive where promised
- ✓ Rights are in place to redistribute

Meeting the FAIR Principles





Findable

Accessible

Interoperable

Re-usable

Persistent identification of collections

Data Collection: selection, appraisal and management

- Scope supported by a **Collections Development Policy**
- Selection managed through a robust and auditable **appraisal and selection process**
- Data acquisition managed through a **rights framework**
- Ability to **manage access conditions** through a robust legal and technical framework
- Ability to store, curate and host data through a **trust framework**



2. Legal and ethical issues surrounding sharing data



Research: ethical and legal obligations

- Research relating to legal persons requires **ethical review** - subjects or organisations
- Comply with **relevant laws**
- Uphold standards of **research integrity**
- Avoid social and personal **harm**



Ethical, legal and research integrity challenges

1. Concerns about **appropriate re-use** of data
2. 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 data

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



Personal data

- The **General Data Protection Regulation (GDPR)** comes into force on 25 May 2018
- Applies to:
 - **Researchers** based in the EU who process (collect/use/delete) personal data about **a citizen of any country, anywhere** in the **world**
 - Researchers outside who collect data about **individuals within the EU**
- Data Protection legislation is not intended to, and does not, inhibit ethical research
- Legal gateways are required to process and handle personal data



Legal gateways: processing grounds for research

- UK **Statistics and Registration Service Act** (SRSA), 2007
 - allows information sharing between public authorities and the Statistics Authority for statistical purposes
 - confidential information should not be disclosed by anyone; unlawful disclosure is a criminal offence punishable by a fine or imprisonment
- Various 'Processing Grounds' under the GDPR, e.g. consent, public interest task and legitimate interest
 - Processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, shall be subject to appropriate safeguards



Data Protection

Highlights for data sharing:

- Collect data for specified, explicit and legitimate purposes
- Processing must be lawful, fair and transparent
- Right to erasure - right to be forgotten
- Minimize size of data - relevant and limited
- Accuracy – keep data up to date
- Transfer of data – only when protections are equivalent
- Retention of data – only as needed

Key messages

- Accountability, transparency and documentation
- ‘Anonymised’ or de-identified data - don’t fall under this legislation



Informed consent and data sharing

- Identify and explain the possible future uses of their data
- Offer the participant the 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
- Future governance for data access should recognise these consents

Copyright in data

- Individual **intellectual property rights**
- UK **Crown copyright**: material created by civil servants, ministers and government departments and agencies.
- **Database rights** in UK/EU law – class of literary works
 - Arrangement/selection and structure/organisation
- Who owns a ‘collection of data’ and which parts?



3. Pathways to access for data: The 5 Safes framework

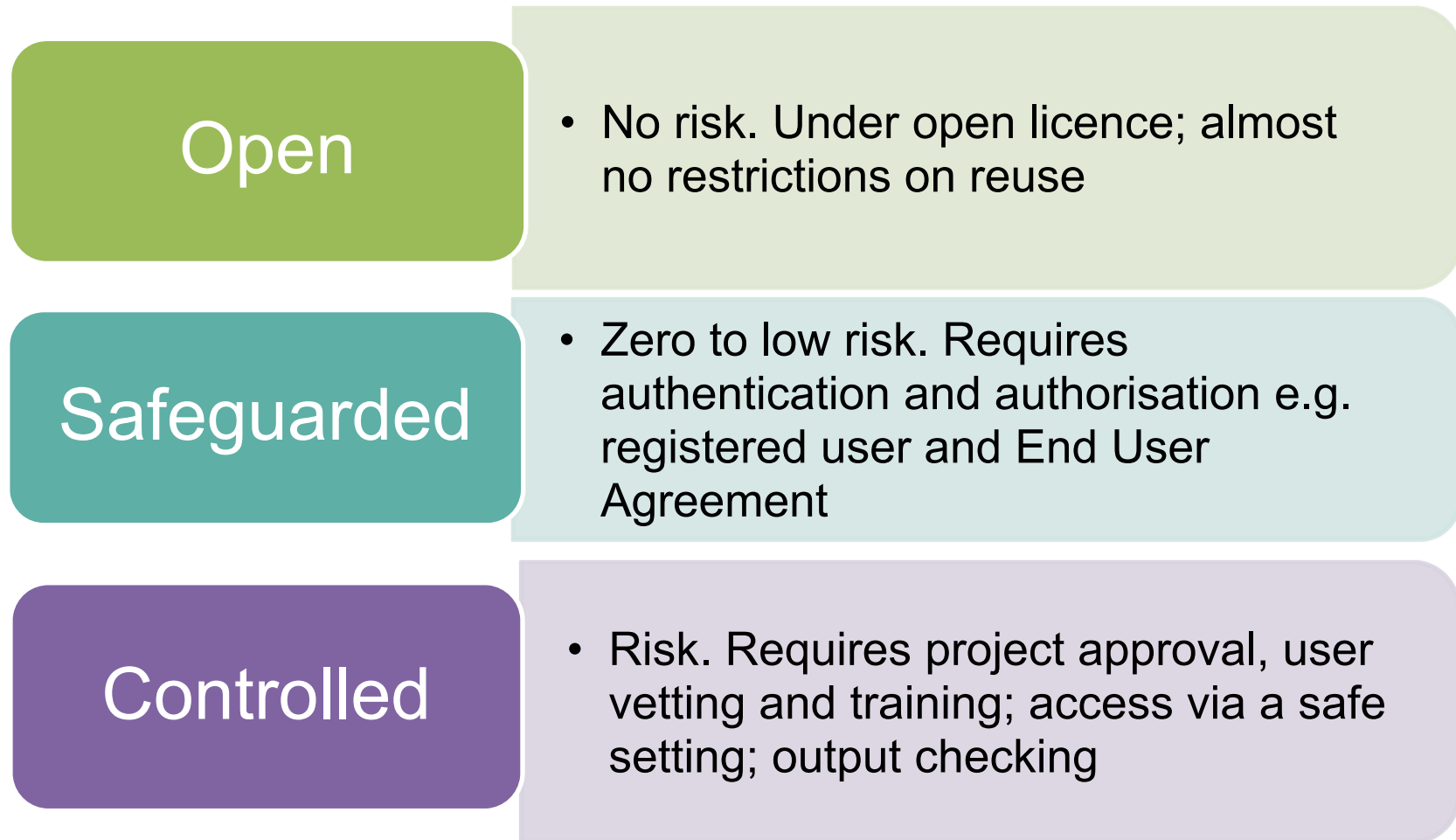


Strategies for enabling safe access to data

- ✓ 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



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



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



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



Controlled access: safe haven

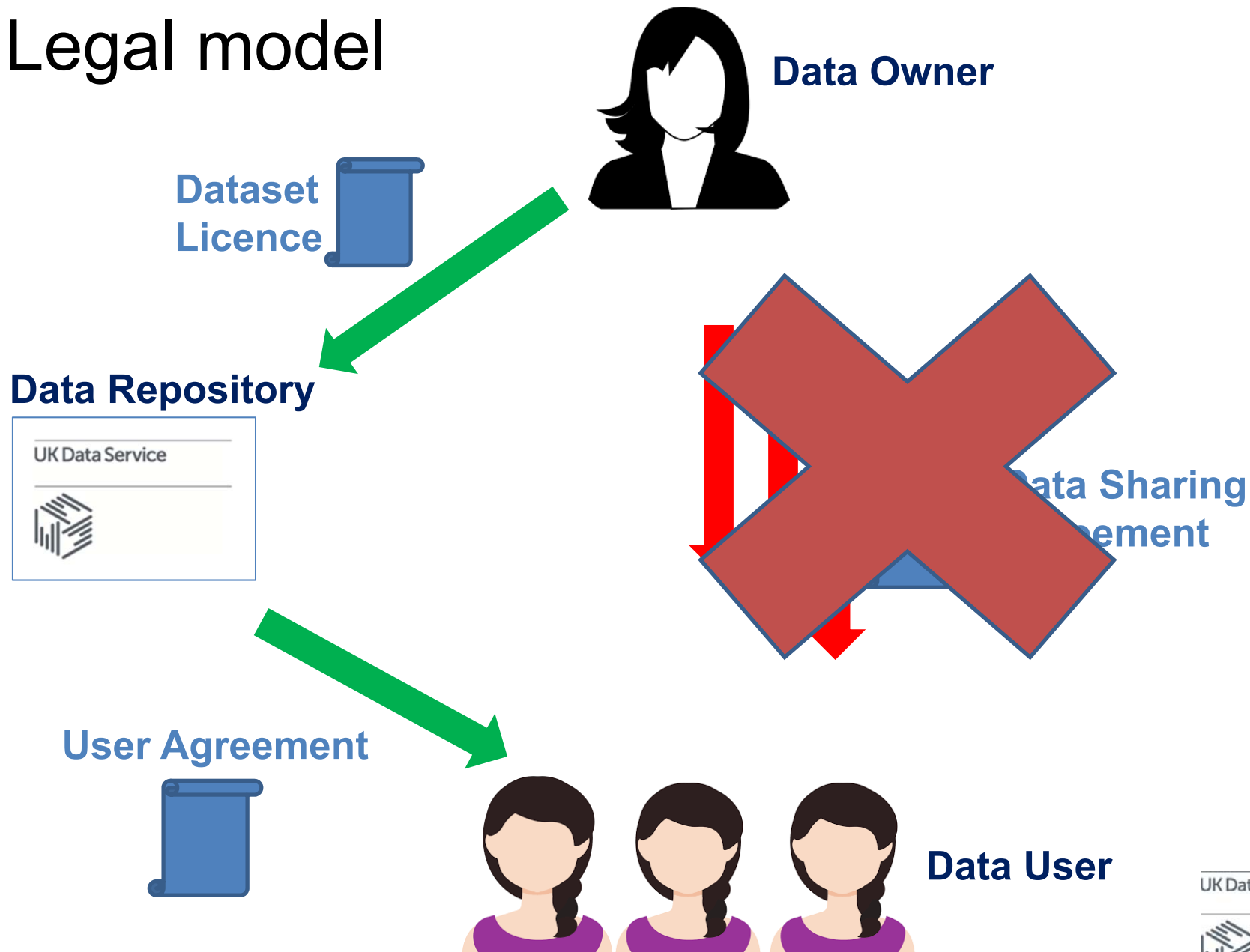
- Demonstrable research need for more detailed data
- Controlled access requires:
 - Approved/Accredited Researcher application SAFE PROJECT
 - User Agreement signed by researcher and their institution SAFE PEOPLE
 - Face-to-face training session SAFE PEOPLE
 - Secure access via: remote desktop or safe room SAFE SETTING
 - Outputs checked SAFE OUTPUTS



4. Legal documents: data licences and access agreements



Legal model



A Licence agreement

- A legal arrangement between the creator/depositor of the data collection and the host repository
- Clarifies who **owns the data** and whether they have **the right to allow** a repository to publish data; can sign on behalf of a rights holder
 - *May need to gain **copyright clearance** to redistribute material with existing copyright*
- Grants a repository a non-exclusive **right to preserve and disseminate** the data on their behalf
- Sets out what a user is **allowed to do** with the data
 - *The licence type should be **displayed** to users*



Formal citation

Enables a **formal citation** to be made for data, with copyright holders as 'authors'

CITATION

The citation for this study is:

NatCen Social Research, University College London. Department of Epidemiology and Public Health. (2018). *Health Survey for England, 2016*. [data collection]. UK Data Service. SN: 8334, <http://doi.org/10.5255/UKDA-SN-8334-1>

[Select the text above](#) to add data citation in your outputs.

Select citation format: XML citation formats: [CSL](#) [EndNote](#)











Example licences: Creative Commons

- Easy to use and flexible
- Widespread international adoption of the licences
- Availability in human-readable and machine-readable forms
- Appealing for data sharers, as rights are well clarified
- But caution: consider carefully which licence to select – Creative Commons licences are not revocable

Once material is received under a CC license, the user always has the right to use it under those license terms, even if the licensor changes his or her mind and stops distributing under the CC license terms.

Rights, licensing and usage conditions

 creative commons LICENSES		Copy & Publish	Attribution Required	Commercial Use	Modify & Adapt	Change License
	Public Domain	✓	✗	✓	✓	✓
	BY Attribution	✓	✓	✓	✓	✓
	BY-SA Attribution ShareAlike	✓	✓	✓	✓	✗
	BY-ND Attribution NoDerivs	✓	✓	✓	✗	✓
	BY-NC Attribution NonCommercial	✓	✓	✗	✓	✓
	BY-NC-SA Attrib NonComm ShareAlike	✓	✓	✗	✓	✗
	BY-NC-ND Attrib NonComm NoDerivs	✓	✓	✗	✗	✓

ata Service



Licence selector

Choose a License

Answer the questions or use the search to find the license you want

Start again

←

→

What do you want to deposit?

SoftwareData

Search for a license...

Creative Commons Attribution (CC-BY)

This is the standard creative commons license that gives others maximum freedom to do what they want with your work.

Publicly AvailableCC BY OPEN DATA

Creative Commons Attribution-ShareAlike (CC-BY-SA)

This creative commons license is very similar to the regular Attribution license, but requires you to release all derivative works under this same license.

Publicly AvailableCC BY SA OPEN DATA

Creative Commons Attribution-NoDerivs (CC-BY-ND)

<http://ufal.github.io/public-license-selector/>

UK Data Service



UKDA Depositor Licence types

Concordat

- Formal deposit/access relationship with national Statistics Offices datasets (e.g. ONS)

Open

- Open Government Licence (OGL)
 - UK aggregate census data
 - Teaching data from ONS surveys
- Creative Commons
 - Teaching data
 - Historical collections

Standard licence

- Safeguarded and controlled datasets
- Selected access route (Access Policy)
- Appendix sets up specific clauses

UK Data Service End User Agreements

- Standard **terms and conditions of use** of the data collection(s)
 - Legally binding
- **Additional** terms and conditions
 - Click use agreement
 - Formal application and approval
 - Secure Access User Agreement; application and training

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](#)



User Agreements: Key areas covered

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



Controlled data: Secure access agreement

Researchers **access data and undertake analyses** in a safe setting/haven

- Access to **personally identifiable data** under a legal gateway
- Agree to conditions for **handling personal data** and **breach penalties**
- Agree to be trained to become an **Approved Researcher**
- **Institution** must countersign on behalf of the user

Non-compliance

- Operate a **Licence Compliance Policy** supported by **Managing Licence Compliance** procedures
- Severe consequences for non-compliance
 - Removal of individual access to data; maybe permanent
 - Individual or institutional suspension from ESRC data services or funding
 - For government data that fall under the [Statistics Act 2007](#), penalties ([S39](#)) may include criminal sanctions
- Self-reported unintentional breaches are penalised with discretion – may need re approval or retraining e.g.
 - Copying information from a screen
 - Someone looking over a user's screen



5. Practicalities of operating governance for data access



Restrictions on data: managing the vetting process

Who?

- Individual permission
 - Difficult to manage long-term
 - Problem of retired or dead depositors...
- Access via a [Data Access Committee \(DAC\)](#)
 - Resilience
 - Range of expertise

How?

- Ad hoc vs. clear processes

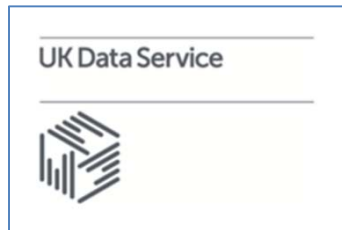
What?

- Requirement for applicant
- Information to support decision making

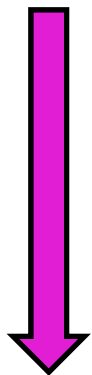
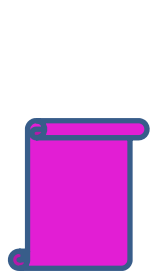
Arrangements can be revised over time

Data Access Committee

Data Repository



Project application



Researcher application



Approved Researcher

Positive Decision



Data access

UK Data Service



Data Access Committee

- A **robust governance process** where a group of individuals review the application and collectively make a decision
- Most **time and resource intensive way** to govern data access - reserved for circumstances that warrant this level of scrutiny
- Applicable for situations where there are **complex factors to consider**, a need for multiple perspectives, and hard decisions to make:
 - Does the data sensitivity warrant it? Is there a risk cost to consider?
 - Is it resource intensive to support the usage of the data? Is there a financial cost to consider?
 - Is there a finite amount of data resource actually available to exploit, such as bio samples?



Setting up a DAC: requirements

- Decision making framework
 - [Criteria for assessment](#) (ONS)
 - Thresholds for **approval** or **rejection**
- Defining relationship to and dependencies on **other review processes** (e.g. ethics review)
- **Panel membership**: chair, appropriate stakeholder representation, lay members, subject expertise, secretariat
- Materials for the panel
 - Application, accompanying reports, assessments



Supporting a DAC: pre-tasks

- Ensure application form is **clear and unambiguous**
 - Capture the right content for the DAC
 - Online forms are preferable
- Support for applicants
 - [Guidance to applicants](#) (NHS Digital)
 - [Pre-application checklist](#) (NHS Digital)
- Manual effort to **screen and triage** applications
 - Redirect applicants where appropriate
 - Resolve errors before they reach the DAC



Project application forms – key information

- The project description - purpose
- The data, variables required and any linkage requirement
- Justification of data requirement
- Funding source
- The research team
- Data access point/storage location/site
- Dates of access
- Outputs expected and how confidentiality will be protected



Operating a DAC: practicalities

- **Cycle of meetings**
 - Volume of applications
 - Timescales for other dependent processes
 - Your ability to resource the DAC
 - Commitment for members
- Method: **virtual vs face-to-face** - cost consideration vs effectiveness
- **Presentation approaches**
 - Tabling papers
 - Advocating from within the 'service'
 - Applicant presenting direct
- **Recording decisions**, communication of outcomes
 - Expectations of applicants & [transparency obligations](#)



After access: final governance steps

- Disclosure control and output checking
- Sign-off on outputs, preparedness for publication
- Commitments from the application - publications

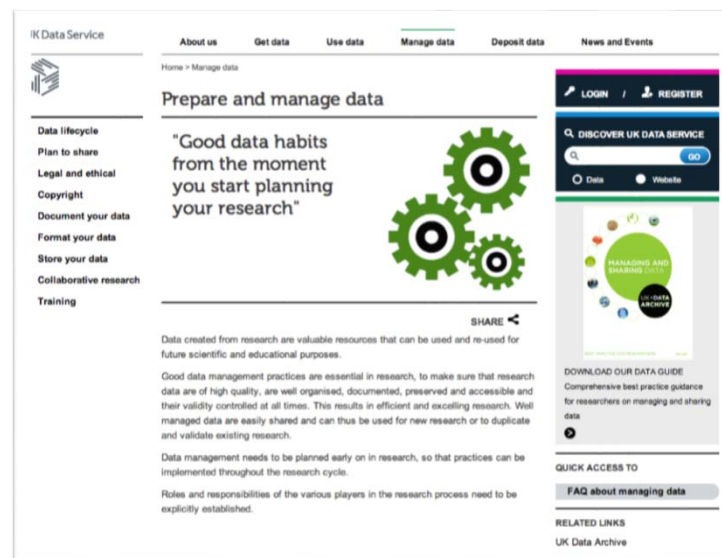
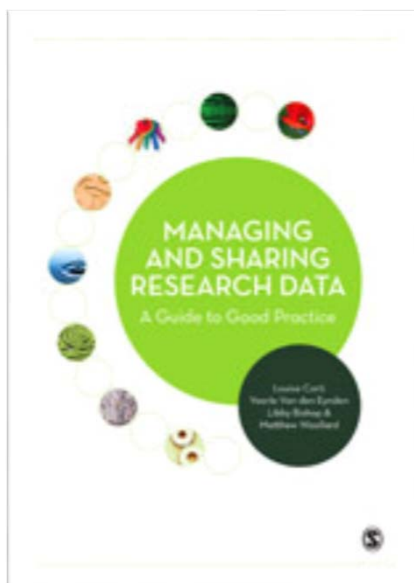
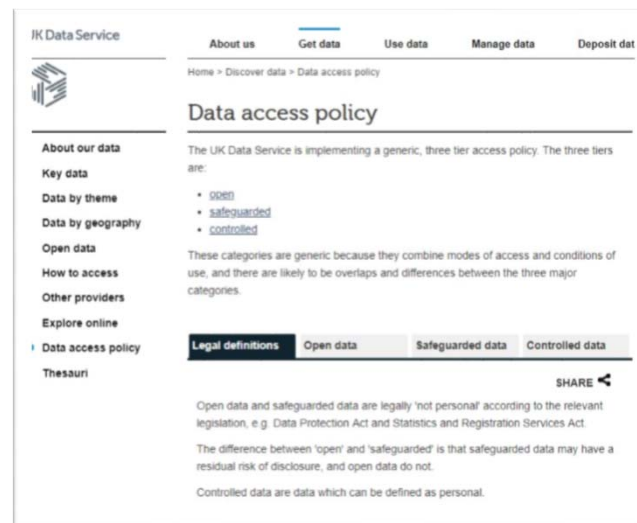
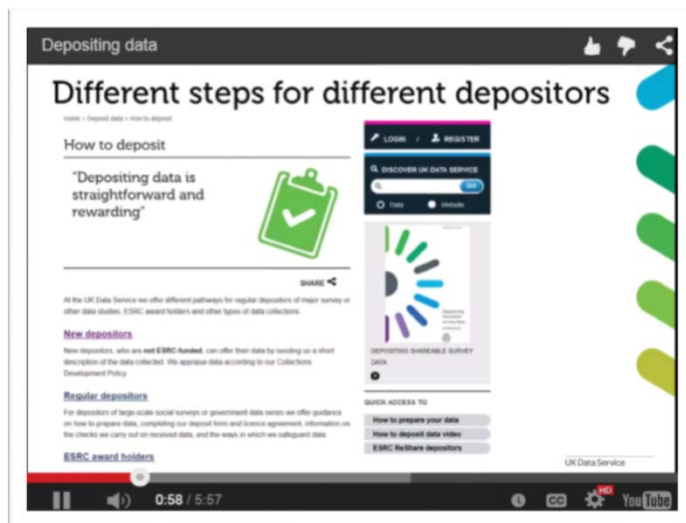


Summary: sharing data safely

- ‘Treat’ data to maximise opportunities for reuse
- Check data for any rights issues prior to publication
- 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 disclosive data through appropriate legal gateways
- Provide accountability across the access lifecycle



ukdataservice.ac.uk



UK Data Service



Watch this space

- University of Glasgow: Seminar on: Choose the Right Rights, Use the Data Right. 6 April 2018
- We have agreed to run a joint workshop with University of Glasgow on the practicalities of licensing data and governance in early Autumn



Keep connected

Collections@ukdataservice.ac.uk

UK Data Service

University of Essex, Colchester, UK

Subscribe to UK Data Service list:

www.jiscmail.ac.uk/cgi-bin/webadmin?A0=UKDATASERVICE

Follow UK Data Service on Twitter: @UKDataService, @UKDSRDM

Youtube: www.youtube.com/user/UKDATASERVICE

Thankyou

UK Data Service

