Intro, FAIR and DMPs

Veerle Van den Eynden
UK Data Service
University of Essex

Managing and sharing research data for
transparency and FAIRness
24 January 2020
Universit of Essex

UK Data Service

Overview

- What is the UK Data Service?
- Why is data management important
- FAIR data principles
- Data management planning

UK Data Service

UK Data Service

- Curate, preserve, provide access to social science data for reuse
- Funded by ESRC
- Data management advice for data creators
- Support for users of the service
- Information about the use to which data are put

ukdataservice.ac.uk

Some statistics about our UK Service

- **7,785** datasets in the collection

- **1190** qualitative and mixed methods collections

- **400** new datasets added each year

- **> 200** case studies of data reuse

- **25,000** registered users

- **60,000** downloads worldwide per year

- **4000+** user support queries per year

UK Data Service

Research data services team

- Supporting researchers to make research data shareable
- UK Data Service helps materialise Data Policy for the Economic and Social Research Council (ESRC)
  - Data management planning advice & guidance
  - Data management guidance & training, esp. on confidentiality, security, ethics
  - Research data published to make them available for re-use via:
    - [ReShare repository](#)

UK Data Service

Why is data management important?

- Data creation in research is often expensive
- Data are fundamental in research
- Good quality data lead to good quality research
- Data underpin published findings
- Helps compliance with ethical codes, data protection laws, journal requirements and funder policies
- Helps to protect data from loss, destruction and potential exposure

UK Data Service

What is data management?

- Data management refers to all aspects of handling, housing, maintaining and preserving data, that together ensure that data are of high quality, well organised, clearly documented, preserved and accessible and their validity controlled at all times

- Making research data available for future reuse (sharing them) certainly requires a good level of data management

- A data management and sharing plan helps researchers consider: when research is being designed and planned, how data will be managed during the research process and shared afterwards with the wider research community

UK Data Service

What makes data good for sharing and reuse ?

Other researchers can understand and reuse the data:

- high quality
- accurate
- well organised
- easily accessible
- well documented
- long-term validity

or **FAIR** data:

Findable, **A**ccessible, **I**nteroperable, **R**eusable

[Guidelines on FAIR Data Management for the H2020 programme](#)

UK Data Service

FAIR data principles

- Optimal reuse of data, by making data FAIR for humans and machines

making your data

FAIR

Findable

To aid automatic discovery of relevant datasets, (meta)data should be easy to find by both humans and machines and be assigned a persistent identifier.

Accesible

Limitations on the use of data, and protocols for querying or copying data are made explicit for both humans and machines.

Interoperable

(Meta)data should use standardised terms (controlled vocabularies), have references to other (meta)data and be machine actionable.

Reusable

(Meta)data are sufficiently well described for both humans and computers to be able to understand them and have a clear and accessible data usage license.

UK Data Service

FAIR data principles

To be Findable:
- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
- A1.1 the protocol is open, free, and universally implementable
- A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:
- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
- R1.1. (meta)data are released with a clear and accessible data usage license
- R1.2. (meta)data are associated with detailed provenance
- R1.3. (meta)data meet domain-relevant community standards

UK Data Service

FAIR key elements

- **A persistent identifier (PID) for the data object as a whole**
  Persistent identifiers [like DOIs](#) prevent link rot. Link rot is the process by which hyperlinks stop referring to the original source through time because they are moved or deleted. Without a PID, the data object simply will not be findable let alone reusable in the long-run.

- **A sufficient set of metadata**
  A sufficient and standardised set of metadata (elements which describe the data) will enhance findability, interoperability, and reusability. The quality of the descriptive information regarding the data has a profound impact on their reusability. So the more documentation of the data's context, the better. As a minimum, there should be sufficient amount of metadata to make the data findable but also understandable and reusable by other researchers.

- **A clear licence**
  Researchers (and computers) who find a dataset should immediately know what they are allowed to do with it. Stating clear re-use rights is like having a warm 'Welcome' on the doormat of your dataset. The motto is: 'open if possible, restricted if necessary'.

UK Data Service

FAIR assessment tool

FAIR self-assessment tool

Findable ⓘ

Does the dataset have any identifiers assigned? — Globally Unique, citable and persistent (e.g. DOI, PURL, ARK c ⌄

Is the dataset identifier included in all metadata records/files describing the data? — No ⌄

How is the data described with metadata? — Comprehensively, but in a text-based, non-standard format. ⌄

What type of repository or registry is the metadata record in? — Data is in one place but discoverable through several registries ⌄

Accessible ⓘ

Interoperable ⓘ

Reusable ⓘ

Total across F.A.I.R

ARDC (2018) FAIR self-assessment tool, Australian Research Data Commons:
https://www.ands-nectar-rds.org.au/fair-tool

UK Data Service

Data management planning

- Many research funders require planning for data management and data sharing in research applications

- Expect to cost sustainable data management and sharing into research

- Overview of requirements:

  - Digital Curation Centre, Funders' data plan requirements

  - Knight, G. (2012) Funder Requirements for Data Management and Sharing. London School of Hygiene and Tropical Medicine, London.

  - CESSDA Training Working Group (2017) Data management requirements in Europe. *CESSDA Data Management Expert Guide*, Bergen, Norway: CESSDA ERIC.

UK Data Service

Funder	DMP required ●, recommended ● or not required ○
United Kingdom	
Arts and Humanities Research Council (AHRC)	●
Biotechnology and Biological Sciences Research Council (BBSRC)	●
Cancer Research UK (CRUK)	●
Department for International Development (DFID)	●
Engineering and Physical Sciences Research Council (EPSRC)	●
Economic and Social Research Council (ESRC)	●
Medical Research Council (MRC)	●
Natural Environment Research Council (NERC)	●
Science and Technology Facilities Council (STFC)	●
Wellcome Trust	●
Europe	
European Commission, Horizon 2020	●
European Science Foundation	○
European Research Council	●
Academy of Finland	●
Deutsche Forschungsgemeinschaft (Germany)	●
NOW (Netherlands)	●
KNAW (Netherlands)	●

Funder	DMP required ●, recommended ● or not required ○
ZonMw (Netherlands)	●
Research council of Norway	●
Swiss National Science Foundation	●
FWO (Belgium)	○
Czech Science Foundation	○
Vetenskapsrådet (Sweden)	○
Slovenian Research Agency	○
USA	
National Science Foundation (NSF)	●
National Institutes of Health (NIH)	●
Gordon and Betty Moore Foundation (GBMF)	●
Institute of Museum and Library Services (IMLS)	●
National Endowment for the Humanities (NEH)	●
National Oceanic and Atmospheric Administration (NOAA)	●
Bill and Melinda Gates Foundation	●
National Institute of Justice	●
Institute of Education Sciences	●
Australia	●
Australian Research Council (ARC)	●

Research funder data policies (UKRI)

- *Publicly funded research data are a public good, produced in the public interest, that should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property.*

- in accordance with relevant standards and community best practice

- metadata to make research data discoverable

- legal, ethical, commercial constraints on release of research data

- recognition for collecting & analysing data; limited privileged use

- acknowledge sources of data, intellectual contributions, terms & conditions

- use public funds to support the management and sharing of publicly-funded research data

  Research Councils UK Common Principles on Data Policy (2011)

  Guidance on best practice in the management of research data (2015)

  Concordat on Open Research Data (2016)

UK Data Service

Research funder data policies (UKRI)

Research Councils:

- Data sharing policy mandating or encouraging data sharing
- Data management / sharing planning required
- Award holders responsible for managing & sharing data, except EPSRC
- Fund data sharing support services and infrastructure

   e.g. UK Data Service (ESRC)

   NERC data centres (NERC)

   MRC Data Support Service (MRC)

   Atlas Petabyte Storage (STFC)

   Archaeology Data Service (AHRC)

UK Data Service

Common DMP topics

- A description of the data that will be created (or reused) during research
- Ethical and legal requirements
- Quality assurance of data, including any standards applied
- How data will be documented
- Data storage and backup measures and required equipment or infrastructure
- Plans for sharing and preservation of data, who will have access and whether there are any embargoes or restrictions
- Data management roles and responsibilities
- Costing or resources needed

UK Data Service

Tools and templates

- Funder template for DMP
  - [ESRC DMP requirements in data policy](#) and [DMP guidance](#)
  - [MRC DMP guidance and template](#)
  - [AHRC DMP requirements](#)

- [Horizon 2020 DM guidelines](#): FAIR as guiding principles for data management planning

- DCC's [DMPonline](#) tool

"We back up our data on sticky notes because sticky notes never crash."

UK Data Service

Costing data management

- [Data management costing guide](#)
  (Westerhof, Pronk, van der Kuil and Mordant 2016)

- [UKDS data management costing tool](#)

UK Data Service

Where things may go wrong

- Perfect DMP is written to ensure funding
- DMP seen as just admin
- Good intentions are not put into practice
- Typical discrepancies in qualitative research:
    - In DMP the researcher indicates that consent for data archiving and future reuse will be sought from interviewees, when in practice this is not discussed with participants
    - Ethical review did not consider data archiving / sharing
    - consent form used may not mention any plans for data archiving, does not ask for permission to make data available for future reuse and may state simply that all interview data will be held in confidence, may contain statements that limit future reuse

UK Data Service

ESRC data policy, principle 4

- There may be legal, ethical or commercial constraints to data sharing; these should be considered in detail before commencing research with the aim of maximising data sharing.

- The ESRC supports the position that most data can be curated and shared ethically provided researchers pay attention right from the planning stages of research to the following aspects:
  - when gaining informed consent, include consent for data sharing
  - where needed, protect participants' identities by anonymising data; and
  - address access restrictions to data in the data management and sharing plan, before commencing research.

- The ESRC regards non-deposit of research data as an exception and reserves the right to request deposit where there is insufficient evidence to prevent data sharing, or to explain why sharing was not considered prior to data collection. Only where researchers have demonstrated due diligence in all three areas will exceptions be granted.

UK Data Service

Key planning issues

- Know your legal, ethical and other obligations towards research participants, colleagues, research funders and institutions
- Know your institution's policies and services: storage and backup strategy, research integrity framework, IPR policy, institutional data repository
- Assign roles and responsibilities to relevant parties
- design data management according to the needs and purpose of the research;
- Incorporate data management into research cycle
- implement and review management of data as part of continued research progression and review.

UK Data Service

Useful resources

- [Research Data Mantra](#)
- [CESSDA Data Management Expert Guide](#)
- [ANDS 23 (research data) Things](#)
- [Top 10 FAIR Data & Software Things](#)
- [UKDS Manage data guidance](#)
- [Sharing social data in multidisciplinary, multi-stakeholder research: best practice guide for researchers](#)

UK Data Service

Tools and templates

- [Model consent form](#) (doc) that takes into account consent for data sharing and future data reuse
- [Sample survey consent statement](#) (doc) that considers consent for data sharing and future data reuse
- [Transcription template](#) (pdf) of our preferred uniform style and layout for transcripts deposited with us
- Example [transcription instructions](#) (pdf) to help transcribers use a consistent style, layout and editing
- [Transcription confidentiality agreement](#) (pdf) to ensure that audio recordings and transcripts that may contain personal and confidential information about interviewees is handled correctly by a transcriber
- [Data list template](#) (xls) to create an informative overview guide of a collection of interview transcripts or other qualitative data
- [Data management costing tool](#) (pdf) to help cost data management tasks in a research grant application
- [Text anonymisation tool](#) (zip) to help you find disclosive information to remove or pseudonymise in qualitative textual data files.
- [ReadMe file template](#) (doc) is a summary file used to inform future users about the content of all data/documentation files/folders available in each ReShare collection

UK Data Service

Questions

Veerle Van den Eynden

veerle@essex.ac.uk