Write a Data Management and Sharing Plan

Veerle Van den Eynden
UK Data Service

Cancer Research UK
Data Sharing and Management
24 May 2017

UK Data Service

Let’s start

- Have you written a DMP before?
- What do you think is the purpose of a DMP?

Overview

- Data policies and funder requirements
- Tools and templates
- Topics
- Exercise

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

Why data management planning

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

Research benefits

- think what to do with research data, how collect, how look after
- keep track of research data (e.g. staff leaving)
- identify support, resources, services needed
- plan storage, short & long-term
- plan security, ethical aspects
- plan longterm sharing
- be prepared for data requests (FoI, funder)

Why 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.

Research funder data policies (RCUK)

- *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)

Research funder data policies (RCUK)

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)

Research funder data policies (CRUK)

We regard it good research practice for all researchers to consider at the research proposal stage how they will manage and share the data they will generate. Any applicants who consider that the data arising from their proposals will not be suitable for sharing must provide clear reasons for not making it available.

- data management and sharing plan as part of application

Cancer Research UK data sharing policy

Cancer Research UK Data sharing guidelines

Overview of UK research funders with data sharing policies and their requirements for data management and sharing plans›

Funder	Plan required?	Required at application	Data topics in DMP
Arts and Humanities Research Council (AHRC)	Yes	Technical plan	Standards, preservation, continued access and use
Biotechnology and Biological Sciences Research Council (BBSRC)	Yes	Data management and sharing plan	Type, format, standards, sharing methods, restrictions, sharing timeframe
Cancer Research UK (CRUK)	Yes	Data sharing plan	Volume, format, standards, metadata, documentation, sharing method, timescale, preservation, restrictions
Department for International Development (DFID)	Yes	Access and data management plan	Repositories, limits, timescale, responsibilities, resources, access strategy
Engineering and Physical Sciences Research Council (EPSRC)	No	Policy framework at institutional level (from 2015)	
Economic and Social Research Council (ESRC)	Yes	Data management plan	Volume, type, quality, archiving plans, difficulties sharing, consent sharing, IPR, responsibilities
Medical Research Council (MRC)	Yes	Data management plan	Collection methods, documentation, standards, preservation, curation, security, confidentiality, sharing and access, timescale, responsibilities
Natural Environment Research Council (NERC)	Yes	Outline data management plan	Data management procedures, created data
Science and Technology Facilities Council (STFC)	Yes	Data management plan	Type, preservation, metadata, value, sharing, timescale, resources needed
Wellcome Trust	Yes	Data management and sharing plan	What data, when share, where share, how access, limits, how preserve, what resources

Research funder data policies (EU)

European open access policies: Horizon 2020, European Research Council (ERC)

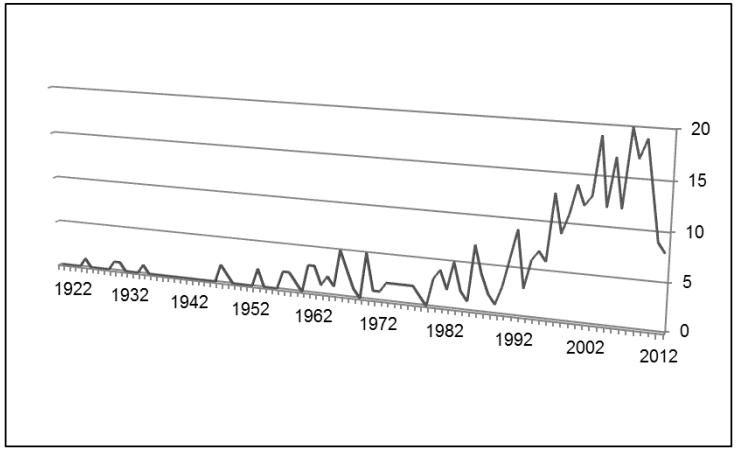
- communication & recommendation on access to / preservation of scientific information (July 2012) (publications & research data)
- pilot on open access to research data, primarily data underlying (open access) scientific publications for H2020
- FAQ open access to publications & data in Horizon 2020
- data management guidelines for Horizon 2020 (~ policies)
- DMP is WP deliverable after 6 months of project start

Service

Journal / publisher data policies

- data underpinning publication accessible
  - upon request from author
  - as supplement with publication
  - in public repository
  - in mandated repository (e.g. PANGAEA – Elsevier)
- citation via unique DOIs
- e.g. BioMed Central open data statement
- global registry of data repositories:
  - re3data

Growth in number of data repositories over time



Source: Databib 2012

CRUK data management plan

- Volume, type, content and format
- Standards for data collection and management
- Metadata, documentation, supporting material
- Method and timescale data sharing
- Long-term preservation plan, with data sharing agreement
- Restrictions on data sharing (IP protection, commercialisation)
- Confidentiality, ethical or consent issues

CRUK guidelines

Tools and templates

- CRUK template for DMP Population Research
- DCC's DMPonline tool

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

DM checklist

- Are you using standardised and consistent procedures to collect, process, check, validate and verify data?
- Are your structured data self-explanatory in terms of variable names, codes and abbreviations used?
- Which descriptions and contextual documentation can explain what your data mean, how they were collected and the methods used to create them?
- How will you label and organise data, records and files?
- Will you apply consistency in how data are catalogued, transcribed and organised, e.g. standard templates or input forms?
- Which data formats will you use? Do formats and software enable sharing and long-term validity of data, such as non-proprietary software and software based on open standards?
- When converting data across formats, do you check that no data or internal metadata have been lost or changed?
- Are your digital and non-digital data, and any copies, held in a safe and secure location?
- Do you need to securely store personal or sensitive data?
- If data are collected with mobile devices, how will you transfer and store the data?
- If data are held in various places, how will you keep track of versions?
- Are your files backed up sufficiently and regularly and are back-ups stored safely?
- Do you know what the master version of your data files is?
- Do your data contain confidential or sensitive information? If so, have you discussed data sharing with the respondents from whom you collected the data?
- Are you gaining (written) consent from respondents to share data beyond your research?
- Do you need to anonymise data, e.g. to remove identifying information or personal data, during research or in preparation for sharing?
- Have you established who owns the copyright of your data? Might there be joint copyright?
- Who has access to which data during and after research? Are various access regulations needed?
- Who is responsible for which part of data management?
- Do you need extra resources to manage data, such as people, time or hardware?

https://www.ukdataservice.ac.uk/manage-data/plan/checklist

Key planning issues

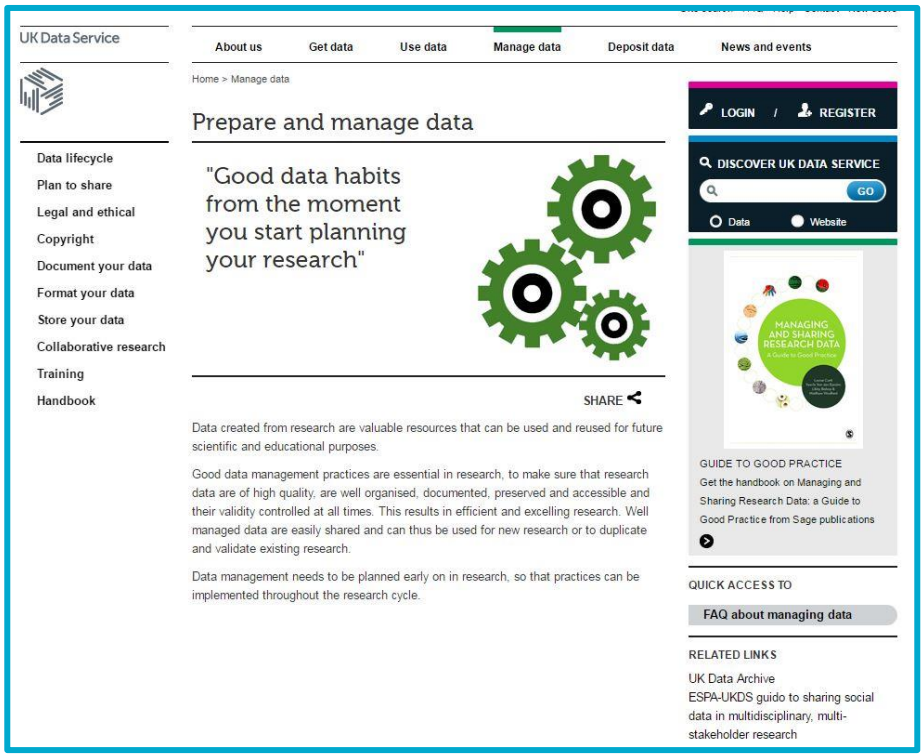
- Know your legal, ethical and other obligations towards research participants, colleagues, research funders and institutions
- Know your institution’s policies and services: storage and backup strategy, research integrity framework, IPR policy, institutional data repository
- Plan where to deposit data
- Assign roles and responsibilities to relevant parties
- Incorporate data management into research cycle
- Implement and review management of data during project meetings and review

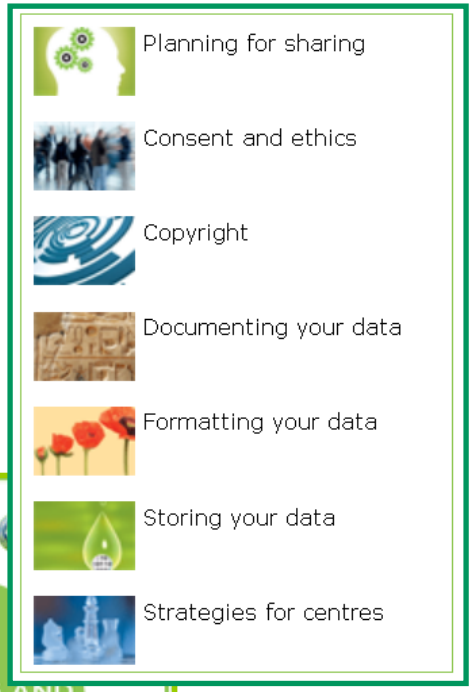
DM topics

- File formats: proprietary versus open/standard
- Data documentation: to understand data
- Quality control
- Storage, backup and security
- Ethical and legal
  - Consent for data sharing reuse
  - Anonymisation
  - Access control
- Copyright and IP of data
- Responsibilities

Our managing and sharing data resources

- UKDS Prepare and manage data web guidance
- Sharing social data in multidisciplinary, multi-stakeholder research Best practice guide for researchers
- Training programme





UK Data Service

Example DMPs

- http://www.dcc.ac.uk/resources/data-management-plans/guidance-examples