Data management plans

Veerle Van den Eynden

UK Data Service

University of Essex

NewComers

12 January 2018

UK Data Service

UK • DATA ARCHIVE

Let's start

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

UK Data Service

Overview

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

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

DM topics

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

UK Data Service

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

- Plan ahead what to do with research data, how collect, how look after
- keep track of research data (e.g. staff leaving)
- identify support, resources, budget, services needed
- plan storage, short & long-term
- plan security, ethical aspects
- plan long-term sharing
- Shows accountability (funder, institution, partners)
- Makes data FAIRer: Findable, Accessible, Interoperable, Reusable
  Guidelines on FAIR Data Management for the H2020 programme

UK Data Service

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.

UK Data Service

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)

UK Data Service

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)

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

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
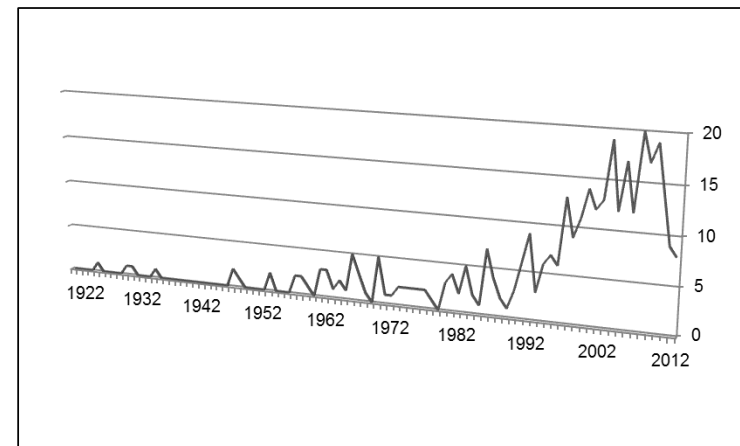
Source: Databib 2012

ESRC research data policy

*Publicly-funded research data are a public good, produced in the public interest, which shall be made openly available and accessible with as few restrictions as possible in a timely and responsible manner that meets a high ethical standard and does not violate privacy or harm intellectual property (*ESRC Research Data Policy, 2015)

- ESRC grant include a data management plan with their application, as an attachment to the Je-S form
- ESRC award holders deposit their research data in the ReShare repository (managed by UK Data Service) within three months of the end of their grant, to preserve them and to make them available for new research.

*Researchers who collect the data initially should be aware that ESRC expects that others will also use it, so consent should be obtained on this basis and the original researcher must take into account the long-term use and preservation of data. (*ESRC Framework for Research Ethics, 2012)

ESRC data management plan

Assessment of existing data

Information on new data

Quality assurance of data

Backup and security of data

Difficulties in data sharing and measures to overcome these

Consent, anonymisation, re-use strategies

Copyright / Intellectual Property Ownership

Responsibilities

Management and curation

ESRC DMP guidance

Data management planning for ESRC researchers

| Data management plan | ESRC data policy | ESRC research centres |

SHARE

All ESRC grant applicants generating data during their research have to include a data management plan with their Je-S grant application.

A data management plan helps to decide how research data will be managed throughout the research cycle and will be available for sharing afterwards. Most research data can be successfully archived and shared.

ESRC expects grant holders to consider all issues related to confidentiality, ethics, security and copyright before initiating the research. Any challenges to data sharing (e.g. copyright or data confidentiality) should be critically considered in a plan, with possible solutions discussed to optimise data sharing.

A data management plan includes the following topics.

ASSESSMENT OF EXISTING DATA

- an explanation of the existing data sources that will be used by the research project, with references
- an analysis of the gaps identified between the currently available and required data for the research

Where research grant applicants plan to create new data as part of their ESRC-funded proposal, they must demonstrate that no suitable data are available for re-use. ESRC encourages the re-use of existing data and therefore encourages applicants and grant holders to consider the breadth of data available from various sources before committing to primary data collection.

When using existing data sources, consider copyright and IPR of those data and the conditions of their use, to decide whether resulting derived data can be shared.

Data sources that can be consulted are:

- Discover UK Data Service, with over 6,000 datasets of key economic, social and historical data spanning many disciplines and themes
- RCUK Gateway to Research of past and present research grants and their outputs

INFORMATION ON NEW DATA

Provide information on the data that will be produced or accessed by the research project, e.g.

- data volume
- data type
- data quality, formats, standards documentation and metadata
- methodologies for data collection and/or processing
- source and trustworthiness of third party data

Using standardised and interchangeable data formats ensures the long-term usability of data. Clear and detailed data descriptions and annotation, together with user-friendly accompanying documentation on methods and contextual information, makes data easy to understand and interpret and therefore shareable and with long-lasting usability.

Guidance on data formats

Guidance on documenting data

UK Data Service

NERC outline data management plan

Project information

Organisation

Roles and responsibilities

Data generation activities

Data management approach

Metadata and documentation

Data quality

Exceptions or additional services

NERC DMP guidance

UK Data Service

Horizon 2020 data management plan

Data set reference and name

Data set description

Standards and metadata

Backup and security of data

Data sharing

Archiving and preservations (incl storage and backup)

Horizon 2020 DM guidelines

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 technical plan requirements](#)
  - [NERC DMP guidance and template](#)

- 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

UK Data Service

DM plan
PUCCUK project

storage area set up files
on server for all files
check university backup policy
survey design: attention to format,
e.g. mandatory questions
set up SPSS database,
incl. labels, codes, question text
planned methods written out

Design research:
- design survey
- design interview questions
  & schedule

Sharing data

Prepare fieldwork

design consent form
incl. consent data sharing
contact interviewees
contact transcriber &
agree terms

check file names
and folder structure

clean up data files
delete redundant versions
keep single version of
each file
explore options for
long-term storage
e.g. repository/
archive

Analyse data:
- analyse survey
- code interviews & media
  info → analyse

Collect data:
- survey online
- do interviews
- gather media info
  into files

test audio recorder
attention file format
MP3/WAV
regular transfer audio
files to server
regular copy of survey
responses table to
server (daily/weekly)
whilst survey is online
attention to file naming

during interviews:
- discuss data sharing
- agree whether
  anonymising needed
- take relevant notes
  as context

Processing data:
- survey → SPSS & validate
- transcribe interviews
- transcripts & media files
  → NVIVO

survey to SPSS
check validity
of responses

store consent forms
safely

transcribed interviews
to interviewees
for approval

check transcription
accuracy

attention
to storage

audio to transcriber
with template
& guidelines
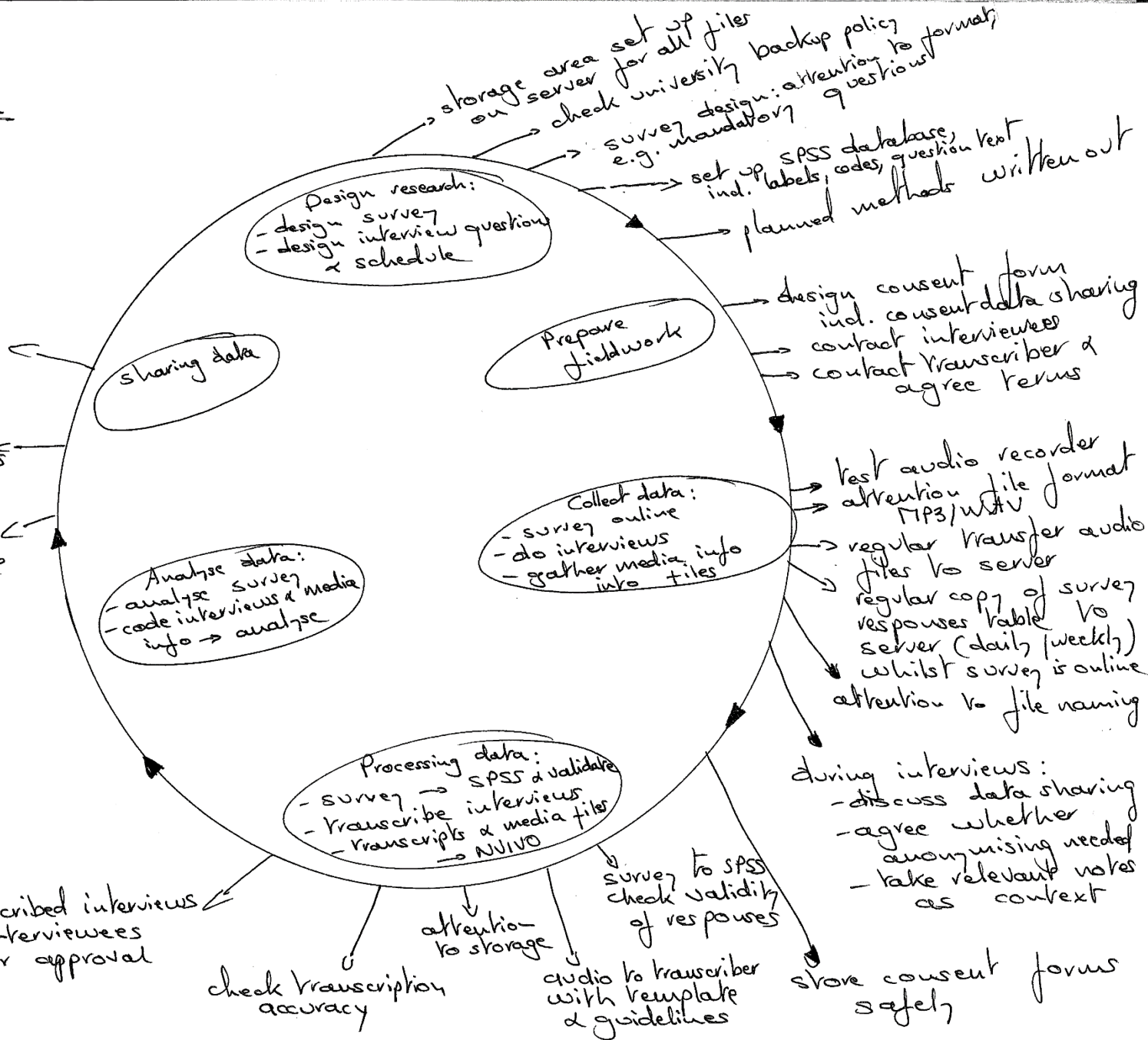
Costing data management

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

- [UKDS data management costing tool](#)

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
- 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

Roles & responsibilities

- Project director: design, oversee research
- Research staff: design research, collect, process and analyze data, where keep data, who has access
- Laboratory or technical staff: generate metadata and documentation
- Database designer
- External contractors: data collection, data entry, transcribe, process, analysis; agree standard protocols
- Support staff: manage and administer research and funding, ethical review and assess IPR
- Institutional IT services: data storage, security and backup services
- External data centres: facilitate data sharing.

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](#)

Planning for sharing

Consent and ethics

Copyright

Documenting your data

Formatting your data

Storing your data

Strategies for centres

UK Data Service

Example DMPs

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

UK Data Service

Exercise: DMP for your research

Reflect on a DMP for your research: which data are you creating, which DM aspects need your attention

Start a DMP online plan or use the ESRC DMP questions:

- existing data sources that will be used by the research project, with references
- analysis of the gaps identified between the currently available and required data for the research
- information on the data that will be produced or accessed by the research project:
    - data volume, data type
    - data quality, formats, standards documentation and metadata
    - methodologies for data collection and/or processing
    - source and trustworthiness of third party data
- planned quality assurance and back-up procedures [security/storage]
- plans for management and curation of primary or third party data
- expected difficulties in data sharing, along with measures to overcome these difficulties, explicitly stating which data may be difficult to share and why
- explicit mention of consent, confidentiality, anonymisation and other ethical considerations and, in particular, strategies taken to not preclude further reuse of data
- copyright and intellectual property ownership of the data
- responsibilities for data management and curation within research teams at all participating institutions

UK Data Service