Data Management Plan Content Checklist

Draft Template for Consultation

Martin Donnelly (University of Edinburgh, martin.donnelly@ed.ac.uk)
Sarah Jones (University of Glasgow, s.jones@hatii.arts.gla.ac.uk)

Introduction

Funders increasingly require their grant-holders to produce data management plans.[1] This content template draws upon their requirements to suggest 'preliminary' and 'full' versions of a data management plan. The preliminary version summarises the main issues most research funders expect researchers to address at the application stage. The full version extends the core sections with additional details that are required by one or two funders, as well as contextual details that could be included as best practice. Core sections required in the preliminary version are marked in **bold**.

Note on usage

This content checklist acts as an aide to producing data management plans for submission to funding bodies, rather than an internal action plan to operationalise the stages of data management. Ideally, such a document will also be developed in conjunction with this checklist.

Next steps

Following consultation, this content checklist will be developed into a working template, into which researchers can insert their own information. An interactive, Web-based tool is also planned, which will allow researchers to customise the template and view and adapt examples of best practice via a library of data management statements/options for each section.

Comments or suggestions on the desired functionality of this tool will be welcomed at the e-mail addresses given above.

---

[1] The DCC has provided a comparison of the curation requirements of the main UK research funders, see: http://www.dcc.ac.uk/resource/curation-policies/ and Sarah Jones (March 2009) *A report on the range of policies required for and related to digital curation,* version 1.2, (DCC, Glasgow)

Note on composition
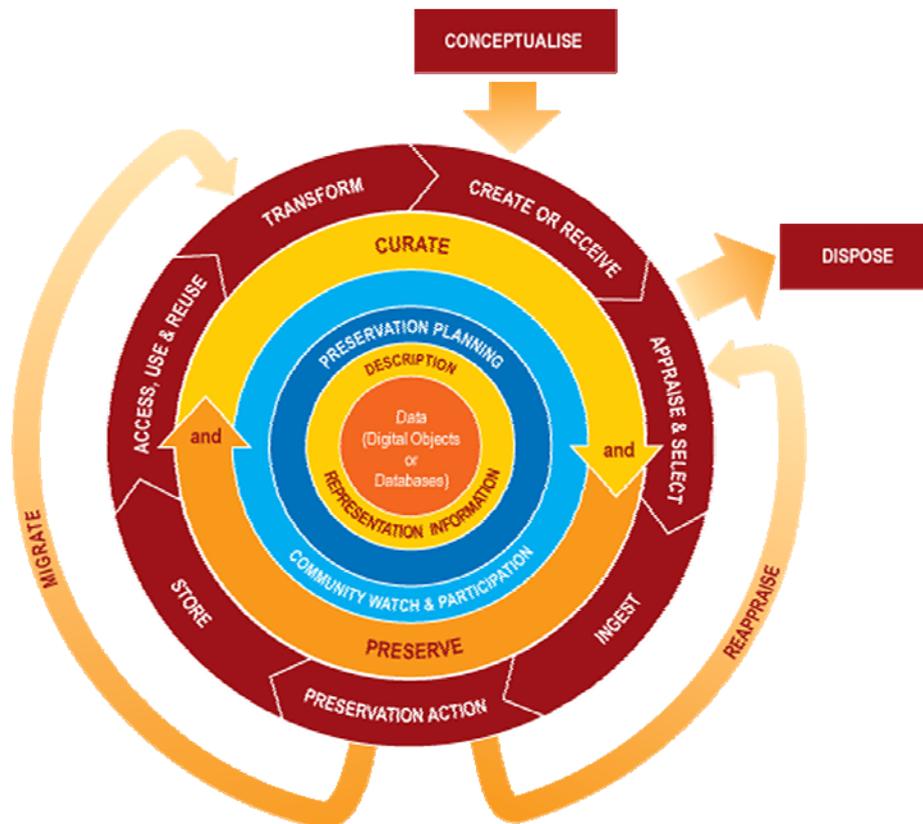
This checklist responds to a recommendation made in Lyon (2007) that "Each funded research project should submit a structured Data Management Plan for peer-review as an integral part of the application for funding."[2] Moreover, the UKRDS final report states that "central to the co-operative service model is the development of data management plans for the data life cycle as described by the Digital Curation Centre."[3] The DCC *Curation Lifecycle Model*[4] (above) was used as a framework to ensure the comprehensiveness of the checklist; this may be useful to researchers in defining roles and responsibilities, identifying risks in points of transition, and ensuring an appropriate chain of custody. As a further benchmarking exercise, existing data management plans were sought and studied to confirm its completeness.[5] We hope in subsequent drafts to provide real-world 'gold-standard' examples for each section, having sought the appropriate permissions. Comments on this draft are welcomed at the e-mail addresses above.

---

[2] Liz Lyon (2007) *Dealing with Data: Roles, Rights, Responsibilities and Relationships*, available at: http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf

[3] *UK Research Data Service Feasibility Study: Report and Recommendations to HEFCE,* December 2008, available for download at: http://www.ukrds.ac.uk/

[4] http://www.dcc.ac.uk/docs/publications/DCCLifecycle.pdf

[5] We referenced a number of data plans produced by organisations such as the British Geological Survey (BGS) and British Atmospheric Data Centre (BADC). We also considered guidance produced for the UK Rural Economy and Land Use (RELU) programme, and by the Australian National University (ANU).

1. Introduction and context

1.1 Basic information (title, summary, funding, duration, partners etc)

1.2 What are the aims and purpose of research?

1.3 List of related policies (e.g. funding body requirements re. creation of a data management plan, institutional or research group guidelines, other dependencies) (Full version should demonstrate how these policies will be adhered to.)

1.4 What are the aims and purpose of data management plan? Who is the target audience? (Full version to include statement on plan revision schedule)

1.5 Glossary of terms

2. Legal, rights and ethical issues

2.1 Who owns the Intellectual Property and copyright?

2.2 How will the data be licensed? e.g. Creative Commons, attribution

2.3 What are the ethical and privacy issues? How will these be resolved? (e.g. anonymisation of data, consent agreements)

2.4 What is the dispute resolution process and/or mechanism for mediation?

3. Access, data sharing and reuse

3.1 How data will be made available? What is the process for gaining access? Are any permissions & restrictions placed on data?

3.2 How can / will the data be shared and re-use? Which bodies/groups are likely to be interested? What are the foreseeable uses?

3.3 Are there embargo periods? How is the release timeframe justified? Include note on right-of-first-use for original data collector/ creator/ investigator

4. Data collection/ development methods

4.1 What does 'data' comprise for the research? (Data description inc. volume, type, content to be created etc.)

4.2 Have you surveyed existing data? (inc. 3rd party data) What can be used/ extended? Are there any access issues? What is the 'added value' to reuse? Why does new data need to be created? What is the relationship between new dataset(s) and existing data? How will you manage interoperability i.e. what methods will be used to integrate the data being gathered in the project with pre-existing data sources?

4.3 How will you capture/create the data? inc. content selection, instrumentation, technologies and approaches chosen, methods for naming, versioning etc

4.4 How will metadata and documentation be captured? What form will it take? What standards will be used? What contextual details are needed to make data meaningful?

4.5 Why have you chosen particular standards and approaches? (e.g. recourse to staff expertise, Open Source, accepted domain-local standards, widespread usage)

4.6 What criteria will be used for Quality Assurance/Management (e.g. documentation, calibration, validation, monitoring, transcription metadata)

5. Data standards

5.1 Data types e.g. experimental measures, qualitative, raw, processed

5.2 Which file formats and platforms will be used and why? (e.g. recourse to staff expertise, Open Source, accepted standards, widespread usage)

5.3 How do data creation decisions take account of end user needs?

6. Short-term storage and data management

6.1 Anticipated data volumes?

6.2 Where will the data be stored? On what media? Who will be responsible? How will it be transmitted? (encryption if appropriate)

6.3 How will access arrangements and data security be managed? How permissions, restrictions and embargoes are enforced? Note on sensitive data, storage on off-network mobile devices etc

6.4 How regularly, by whom, and how will data be backed up?

6.5 Appraisal and retention timeframes (ideally with definite figures) (N.B. this may simply point to relevant institutional or funding body requirements/ policies: political, temporal, commercial, legal)

7. Deposit and long-term preservation

7.1 What is the long-term strategy for maintaining, curating and archiving the data?

7.2 On what basis will data be selected for preservation? How long will data be kept? (ideally with definite figures) (N.B. this may simply point to relevant institutional or funding body requirements/ policies: political, temporal, commercial, legal). How will you dispose/transfer sensitive data? Justification of decisions.

7.3 How will data be prepared for preservation / data sharing? (inc. anonymisation if appropriate)

7.4 Where and how data will be archived e.g. deposit in public repository or existing community database? Transmission of data (encryption if appropriate)

7.5 What related information will be deposited? - references, reports, research papers, fonts, original bid proposal, etc

7.6 What metadata/ documentation will be created at each stage of ingest/ transformation? – descriptive, structural, administrative, preservation etc. How will this be created and by whom?

7.7 What procedures are in place for preservation and backup? How regular, by whom, methods used? e.g. format normalisation, migration

8. Resourcing

8.1   Staff/organisational roles and responsibilities for implementing this plan, inc. time allocations, project management of technical aspects, contributions of non-project staff etc

8.2   Financial issues (e.g. payments to service providers within institutions, payments to external data centres for hosting data, income derived from licensing data, etc)

9. Compliance and review

9.1   How adherence to this data management plan will be checked or demonstrated?

9.2   How and when this data management plan will be reviewed?

10. Agreement/ratification by stakeholders (if useful)

10.1   Statement of agreement (with signatures if required)

11. Annexes

11.1   Contact details and expertise of nominated data managers

11.2   Other annexes as required