



DCC Content Checklist for a Data Management Plan

Post-Consultation (v2.0)

Martin Donnelly (University of Edinburgh, martin.donnelly@ed.ac.uk)

Sarah Jones (University of Glasgow, s.jones@hatii.arts.gla.ac.uk)



© Digital Curation Centre, 2009. Licensed under Creative Commons BY-NC-SA 2.5 Scotland: <http://creativecommons.org/licenses/by-nc-sa/2.5/scotland/>

Introduction

Funding bodies increasingly require their grant-holders to produce data management plans, both during the bid-preparation stage and after funding has been secured.¹

The *DCC Data Management Plan* work is intended to help research teams respond to a recommendation in Lyon (2007): that “[e]ach funded research project should submit a structured Data Management Plan for peer-review as an integral part of the application for funding”².

It draws upon the DCC’s analysis of funders’ requirements to help project teams in creating two iterations of a data management plan; the first (‘preliminary’ version) for use at the grant application stage, and a second (‘full’ version) which is developed at the early-project stage.

The preliminary (or ‘core’) version (covers the issues that most research funders will expect researchers to address *at the application stage*. Core sections are indicated by a ‘Y’ in the rightmost column. The full version augments the core sections with additional information sections that are required by one or two major funders, as well as some contextual details that could usefully be included as best practice.

We use the *DCC Curation Lifecycle Model*³ (below) as an additional framework to bolster its comprehensiveness; this model will be helpful to researchers in defining roles and responsibilities pertaining to their data, identifying risks which arise at points of transition, and ensuring an appropriate and safe chain of custody for digital data.

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 (2009) *A report on the range of policies required for and related to digital curation*, version 1.2, (DCC, Glasgow)

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 <http://www.dcc.ac.uk/docs/publications/DCCLifecycle.pdf>

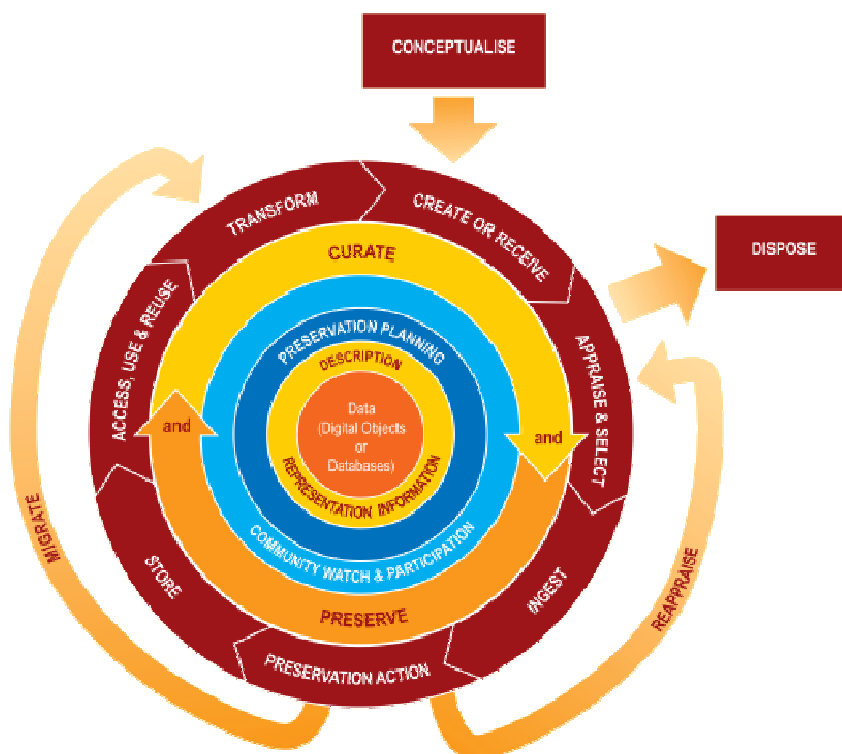


Figure 1 - The DCC Curation Lifecycle Model, S. Higgins/C. Blackall/S. Fairhurst

As a further benchmarking exercise, existing real-world data management plans were sought and studied in order to check the checklist’s completeness.⁴ Having sought the appropriate permissions from the originators, we hope in subsequent drafts to provide ‘gold-standard’ examples for each section which users will be able to consult and modify for their own use.

Document history and next steps

This is the second version of the *DCC Content Checklist for a Data Management Plan*, and takes into account feedback received from a variety of parties via a public consultation process. The most obvious change in this document is that each themed paragraph has been split into a series of atomic sections, employing ‘closed’ questions where possible. The phrasing has also been adjusted throughout to use a more active voice.

This checklist will shortly be developed into a static document template which research teams can populate. The final product will be a working, customisable online DMP template tool, into which researchers can drop their own information via an interactive Web interface. Users will be able to customise and re-order sections according to their specific needs, and unobtrusive guidance and suggestions for further help will be provided. In time it is hoped that users will be able to view and adapt examples and expressions of good data management practice via an openly accessible library corresponding to each section.

⁴ 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).

Note on usage

This checklist acts as an aide for researchers (and/or other research-supporting staff) charged with producing data management plans for submission to funding bodies, and their subsequent development once funding has been awarded; it is not an internal action plan for operationalising the stages of data management. Ideally, the research team should also develop an internal action plan to be used in conjunction with this checklist.

Data Management Plan

ID	Section/ sub-section name and description	Core (Compulsory)
1.	Introduction and context	
1.1	<i>Basic project information</i>	
1.1.1	Name of project	Y
1.1.2	Funding body/bodies	Y
1.1.3	Budget	Y
1.1.4	Duration	Y
1.1.5	Partner organisations	Y
1.2	<i>What are the aims and purpose of the research?</i>	
1.3	<i>List of related policies</i>	
1.3.1	Funding body requirements relating to the creation of a data management plan	Y
1.3.2	Institutional or research group guidelines	Y
1.3.3	Other dependencies	Y
1.4	<i>Basic Data Management Plan information</i>	
1.4.1	Date of creation	Y
1.4.2	Aims and purpose	Y
1.4.3	Target audience	Y
1.4.4	Statement on plan revision schedule	
1.4.5	Does this version supersede an earlier plan?	

ID	Section/ sub-section name and description	Core (Compulsory)
1.5	<i>Glossary of terms</i>	
2	Rights and ethical issues	
2.1	<i>Ethical and privacy issues</i>	
2.1.1	Are there ethical and privacy issues?	Y
2.1.2	If so, how will these be resolved? (e.g. anonymisation of data, institutional ethical committees, formal consent agreements.)	Y
2.2	<i>Rights and intellectual property</i>	Y
2.3	<i>Who owns the Intellectual Property and copyright? (This should address the risk of movement of staff between institutions mid-project.)</i>	Y
2.4	<i>What is the dispute resolution process and/or mechanism for mediation?</i>	
2.5	<i>How will the data be licensed?</i>	Y
3	Access, data sharing and re-use	
3.1	<i>Data sharing and re-use</i>	
3.1.1	Will you share the data you capture or create?	Y
3.1.2	Which bodies/groups are likely to be interested in the data?	Y
3.1.3	What are the foreseeable contemporary or future uses for the data?	Y
3.1.4	Are there any reasons not to share or re-use data? (Suggestions: ethical, non-disclosure, etc.)	Y

ID	Section/ sub-section name and description	Core (Compulsory)
3.2	Access	
3.2.1	Do you have an obligation to make the data available? (e.g. due to funding council policy or Freedom of Information legislation. Note that Fol differs in Scotland from England and Wales.)	Y
3.2.2	How will you make the data available?	Y
3.2.3	Will any permission restrictions need to be placed on the data?	Y
3.2.4	What is the process for gaining access?	Y
3.2.5	Will access be chargeable?	Y
3.2.6	Do you plan on publishing data-related findings?	Y
3.2.7	If so, do your prospective publishers place any restrictions on other avenues of publication?	Y
3.3	Timing	
3.3.1	Is there a right-of-first-use agreement for the original data collector/ creator/ principal investigator?	Y
3.3.2	Details of any embargo periods for political/commercial/patent reasons	Y
4	Data standards and capture methods	
4.1	<i>What does the term 'data' comprise for the research? (Data description, including volume, type, content to be created etc.)</i>	Y
4.2	<i>What data types will you be creating or capturing? (e.g. experimental measures, qualitative, raw, processed)</i>	Y
4.3	Existing and new data	
4.3.1	Have you surveyed existing data, in your own institution and from third parties?	Y

ID	Section/ sub-section name and description	Core (Compulsory)
4.3.2	What existing datasets could you use or build upon?	Y
4.3.3	Are there any access issues?	Y
4.3.4	What is the 'added value' to re-use?	Y
4.3.5	Why do you need to capture or create new data?	Y
4.3.6	What is the relationship between new dataset(s) and existing data?	Y
4.3.7	How will you manage integration between the data being gathered in the project and pre-existing data sources? (This should cover provenance, trust and data quality.)	Y
4.4	<i>How will you capture or create the data? (This should cover content selection, instrumentation, technologies and approaches chosen, methods for naming, versioning, meeting user needs, etc, and should be sensitive to the location in which data capture is taking place.)</i>	Y
4.5	<i>Which file formats will you use, and why? (e.g. recourse to staff expertise, Open Source, accepted standards, widespread usage.)</i>	Y
4.6	Metadata	
4.6.1	What contextual details are needed to make the data you capture or collect meaningful?	Y
4.6.2	How will you create or capture these metadata?	Y
4.6.3	What form will the metadata take?	Y
4.6.4	Will metadata creation be automated?	Y
4.6.5	Which metadata standards will you use?	Y

ID	Section/ sub-section name and description	Core (Compulsory)
4.7	<i>Why have you chosen particular standards and approaches for metadata and contextual documentation? (e.g. recourse to staff expertise, Open Source, accepted domain-local standards, widespread usage)</i>	
4.8	<i>What criteria will you use for Quality Assurance/Management (e.g. documentation, calibration, validation, monitoring, transcription metadata, peer-review.)</i>	
5	Short-term storage and data management	
5.1	<i>Anticipated data volumes. (Ballpark figures, orders of magnitude.)</i>	Y
5.2	Storage	
5.2.1	Where (physically) will you store the data?	
5.2.2	On what media will you store the data?	
5.2.3	Whose responsibility is the storage of the data?	
5.2.4	How will you transmit the data, if required? (Should address encryption if appropriate.)	
5.3	Back-up	
5.3.1	How will you back-up the data? (Should address off-site storage.)	Y
5.3.2	How regularly will back-ups be made?	Y
5.3.3	Whose responsibility will this be?	Y

ID	Section/ sub-section name and description	Core (Compulsory)
5.4	<i>How will you manage access arrangements and data security? How will you enforce permissions, restrictions and embargoes ? Note on sensitive data, off-network storage, storage on mobile devices (laptops, smartphones, flash drives, etc), policy on making copies of data, etc.</i>	Y
6	Deposit and long-term preservation	
6.1	<i>What is the long-term strategy for maintaining, curating and archiving the data? (Reminder that project can consult institutional archivist(s) and/or records managers in long-term retention plans.)</i>	Y
6.2	<i>Specifics</i>	
6.2.1	On what basis will data be selected for preservation?	
6.2.2	How long will/should data be kept beyond the life of the project? (N.B. this may simply link to relevant institutional or funding body requirements/ policies: political, temporal, commercial, legal).	
6.2.3	How will you dispose/transfer sensitive data? (Include justification of decisions?)	

ID	Section/ sub-section name and description	Core (Compulsory)
6.3	<i>Which archive/repository/central database/ data centre have you identified as a place to deposit data?</i>	
6.4	<i>Appraisal and retention timeframes (ideally with definite figures) (N.B. this may simply link to relevant institutional or funding body requirements/ policies: political, temporal, commercial, legal) (Online tool: may want to allow users to upload policies.)</i>	
6.5	<i>What transformations will be necessary to prepare data for preservation / data sharing? (e.g. data cleaning/anonymisation where appropriate.)</i>	
6.6	<i>What related (representation) information will be deposited? (e.g. references, reports, research papers, fonts, the original bid proposal, etc.)</i>	
6.7	<i>What metadata/ documentation will be created at each stage of deposit/ transformation? (e.g. descriptive, structural, administrative, preservation etc.) How will this be created and by whom? Will you include links to published materials and/or outcomes? How will you address the issue of persistent citation?</i>	Y
6.8	<i>What procedures does your intended long-term data storage facility have in place for preservation and backup? (How regular, by whom, methods used? (e.g. format normalisation, migration)</i>	
7	Resourcing	
7.1	<i>Staff/organisational roles and responsibilities for implementing this plan, inc. time allocations, project management of technical aspects, training requirements, contributions of non-project staff etc. (Named individuals.)</i>	

ID	Section/ sub-section name and description	Core (Compulsory)
7.2	<i>Financial issues (e.g. payments to service providers within institutions, payments to external data centres for hosting data, income derived from licensing data, etc) [reminder to build costs of in-project data management into project budget.]</i>	
8	Adherence, review and long-term management	
8.1	<i>Adherence</i>	
8.1.1	When will adherence to this data management plan be checked or demonstrated?	
8.1.2	Who will do this?	
8.2	<i>How and when will this data management plan be reviewed?</i>	
8.3	<i>Longer-term responsibilities</i>	
8.3.1	Is there a formal process for transferring responsibility for the data?	
8.3.2	Who will have responsibility over time for decisions about the data once the original personnel have gone? (Likely to be custodians in data centres.)	
8.3.3	Who will meet the costs of long-term management and storage?	

ID	Section/ sub-section name and description	Core (Compulsory)
9	Agreement/ratification by stakeholders (if useful)	
9.1	<i>Statement of agreement (with signatures if required)</i>	
10	Annexes	
10.1	<i>Contact details and expertise of nominated data managers / named individuals</i>	
10.2	<i>Other annexes as required/ desired</i>	