Checklist for a Data Management Plan

Post-Consultation (v2.2: 6th January 2010)

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

Introduction

Funding bodies increasingly require their grant-holders to produce data management plans.[1] The *DCC Checklist for a Data Management Plan* has been produced 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 (comprising those sections given **in bold type**) covers the issues that most research funders will expect researchers to address **at the application stage**. 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.

[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

The checklist uses 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.

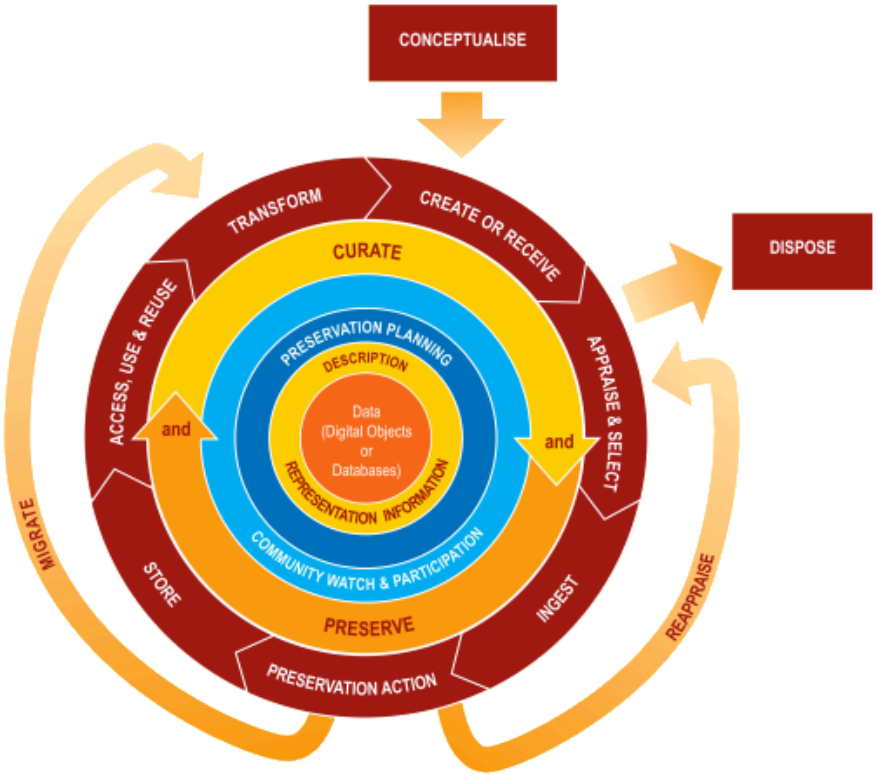

Figure 1 - The DCC Curation Lifecycle Model, S. Higgins/C. Blackall/S. Fairhurst

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.

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

Note on usage

The checklist acts as an aide for researchers (and/or other research-support 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.

[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).

Data Management Plan

ID	Section/ sub-section name and description	Core?	Guidance notes or links
	1. Introduction and context		
	1.1 Basic project information		
	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 What are the aims and purpose of the research?		
	1.3 Related policies		
	1.3.1. Funding body requirements relating to the creation of a data management plan	Y	Link to DCC page which summarises these policies
	1.3.2. Institutional or research group guidelines	Y	
	1.3.3. Other dependencies	Y	

ID	Section/ sub-section name and description	Core?	Guidance notes or links
	1.4 Basic Data Management Plan information		
	1.4.1. Date of creation	Y	
	1.4.2. Aims and purpose	Y	
	1.4.3. Target audience for this plan	Y	
	1.4.4. Statement on plan revision schedule		
	1.4.5. Does this version supersede an earlier plan?		
	1.5 Glossary of terms		
	2. Legal and ethical issues		Most people won't know the answers to this – point to sources of advice.
	2.1 Ethical and privacy issues		Where applicable users may wish to link to approvals they have gained from Ethics Committees and detail any specific conditions applied.
	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	
	2.1.3. Is the data 'personal data' in terms of the Data Protection Act 1998 (the DPA)?		

ID	Section/ sub-section name and description	Core?	Guidance notes or links
	2.1.4. What have you done to comply with your obligations under the DPA?		Reference to standard BS10012 o Data Protection — Specification for a Personal Information Management System.
	2.2 Intellectual property rights		Need to strike appropriate balance between concern for legal implications and getting research done. Inactivity due to legal overwhelm is to be avoided
	2.2.1. Is the dataset covered by copyright or the Database Right? If so, who owns the copyright and other intellectual property? (Ideally, this should address the risk of movement of staff between institutions mid-project.)	Y	http://www.dcc.ac.uk/resource/legal-watch/ipr-in-databases/
	2.2.2. How will the dataset be licensed if rights exist? (e.g. any restrictions or delays on data sharing needed to protect intellectual property, copyright or patentable data.)	Y	
	2.2.3. What is the dispute resolution process and/or mechanism for mediation?		
	3. Access, data sharing and re-use		
	3.1 Data sharing and re-use		
	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	

ID	Section/ sub-section name and description	Core?	Guidance notes or links
	3.1.3. What and who are the intended or foreseeable uses / users of the data?	Y	
	3.1.4. Are there any reasons not to share or re-use data? (Suggestions: ethical, non-disclosure, etc.)	Y	
	3.2 Access		
	3.2.1. Do you have an obligation to make the data available? (e.g. due to research funder policy or Freedom of Information legislation)	Y	Note that FoI differs slightly in Scotland from England and Wales.
	3.2.2. How and when 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 to the data?	Y	
	3.2.5. Will access be chargeable?	Y	
	3.2.6. Do you plan to publish findings which rely on the data?	Y	
	3.2.7. If so, do your prospective publishers place any restrictions on other avenues of publication?	Y	

ID	Section/ sub-section name and description	Core?	Guidance notes or links
	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 What does the term 'data' comprise for the research? (Data description, including volume, type, content to be created etc.)	Y	Tool should give megabyte/ gigabyte/ terabyte/ petabyte options from a dropdown box.
	4.2 What data types will you be creating or capturing? (e.g. experimental measures, qualitative, raw, processed)	Y	RIN data types: 1. scientific experiments, which may in principle be reproduced, although it may in practice prove difficult, or not cost-effective, to do so 2. models or simulations, where it may be more important to preserve the model and associated metadata than the computational data arising from the model 3. observations – from the astronomical to the zoological – of specific phenomena at a specific time or location, where the data will usually constitute a unique and irreplaceable record; 4. derived data, resulting from processing or combining "raw" or other data (where care may be required to respect the rights of the owners of the raw data); 5. canonical or reference data relating, for example, to gene sequences, chemical structures, or literary texts
	4.3 Existing and new data		

ID	Section/ sub-section name and description	Core?	Guidance notes or links
	4.3.1. Have you surveyed existing data, in your own institution and from third parties?	Y	
	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 'added value' will the new data you create or capture provide to existing datasets?	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 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.)	Y	

ID	Section/ sub-section name and description	Core?	Guidance notes or links
	4.5 Which file formats will you use, and why? (e.g. recourse to staff expertise, Open Source, accepted standards, widespread usage.)	Y	Guidance provided by AHRC. Also, link to DIFFUSE standards work (and other DCC tools and resources such as the Curation Manual)
	4.6 Metadata		
	4.6.1. What contextual details are needed to make the data you capture or collect meaningful?	Y	Give examples of categories of metadata here: descriptive/ structural/ administrative/ preservation/ rights metadata/ representation information.
	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. To what extent will metadata creation be automated?	Y	
	4.6.5. Which metadata standards will you use?	Y	
	4.7 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)		

ID	Section/ sub-section name and description	Core?	Guidance notes or links
	4.8 What criteria will you use for Quality Assurance/Management (e.g. documentation, calibration, validation, monitoring, transcription metadata, peer-review.)		
	5. Short-term storage and data management		
	5.1 Anticipated data volumes. (Ballpark figures, orders of magnitude.)	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? (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	

ID	Section/ sub-section name and description	Core?	Guidance notes or links
	5.3.3. Whose responsibility will this be?	Y	
	5.4 Security		
	5.4.1. How will you manage access arrangements and data security?	Y	Note on sensitive data, off-network storage, storage on mobile devices (laptops, smartphones, flash drives, etc), policy on making copies of data, etc.
	5.5 How will you enforce permissions, restrictions and embargoes?		
	5.6 Other security issues		
	6. Deposit and long-term preservation		
	6.1 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.)	Y	
	6.2 Specifics		
	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).		

ID	Section/ sub-section name and description	Core?	Guidance notes or links
	6.2.3. How will you dispose of/transfer sensitive data? (Include justification of decisions.		
	6.3 Which archive/repository/central database/ data centre have you identified as a place to deposit data?		
	6.4 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
	6.5 What transformations will be necessary to prepare data for preservation / data sharing? (e.g. data cleaning/anonymisation where appropriate.)		
	6.6 What related (representation) information will be deposited? (e.g. references, reports, research papers, fonts, the original bid proposal, etc.)		
	6.7 Metadata and documentation		
	6.7.1. What metadata/ documentation will be created at each stage of deposit/ transformation?	Y	Metadata types: descriptive, structural, administrative, preservation. Give definitions and links to further resources in the right-hand column.
	6.7.2. How will this be created and by whom?	Y	
	6.7.3. Will you include links to published materials and/or outcomes?	Y	

ID	Section/ sub-section name and description	Core?	Guidance notes or links
	6.7.4. How will you address the issue of persistent citation?	Y	
	6.8 What procedures does your intended long-term data storage facility have in place for preservation and backup?		This should cover: - how regular? - by whom? - methods used (e.g. format normalisation, migration)
	7. Resourcing		
	7.1 Staff/organisational roles and responsibilities for implementing this plan		This should include time allocations, project management of technical aspects, training requirements, contributions of non-project staff etc. Individuals should be named where possible.
	7.2 Financial issues		This should cover (e.g.) payments to service providers within institutions, payments to external data centres for hosting data, income derived from licensing data, etc). It is also important to remember to build costs of in-project data management into the project budget.
	8. Adherence, review and long-term management		
	8.1 Adherence		
	8.1.1. When will adherence to this data management plan be checked or demonstrated?		
	8.1.2. Who will do this?		

ID	Section/ sub-section name and description	Core?	Guidance notes or links
	8.2 How and when will this data management plan be reviewed?		
	8.3 Longer-term responsibilities		
	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?		

	9. Agreement/ratification by stakeholders (if useful)		
	9.1 Statement of agreement (with signatures if required)		
	10. Annexes		
	10.1 Contact details and expertise of nominated data managers / named individuals		
	10.2 Other annexes as required/ desired		