

The International Journal of Digital Curation

Issue 1, Volume 5 | 2010

DMP Online: The Digital Curation Centre's Web-based Tool for Creating, Maintaining and Exporting Data Management Plans

Martin Donnelly, Sarah Jones, John W. Pattenden-Fail
Digital Curation Centre, UK

Abstract

Funding bodies increasingly require researchers to produce Data Management Plans (DMPs). The Digital Curation Centre (DCC) has created DMP Online, a web-based tool which draws upon an analysis of funders' requirements to enable researchers to create and export customisable DMPs, both at the grant application stage and during the project's lifetime.

Introduction and Context

The Digital Curation Centre (DCC) defines digital curation as “maintaining, preserving and adding value to digital research data throughout its lifecycle.”¹ The active management of research data reduces threats to their long-term research value, and mitigates the risk of digital obsolescence.

In 2009, a DCC analysis (Jones, 2009) of research funder policies and requirements for data management found that many funders “expect applicants to consider creation and management of their research outputs at the proposal stage in order to submit a data managements and sharing plan.” DMP Online is a web-based tool for creating, maintaining and exporting DMPs, and has been developed in order to help research teams meet funder requirements, and respond to the 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.”

The tool uses the DCC Curation Lifecycle Model (Figure 1) (Higgins, 2008) as an underpinning framework to bolster its comprehensiveness; this model is designed to help 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.

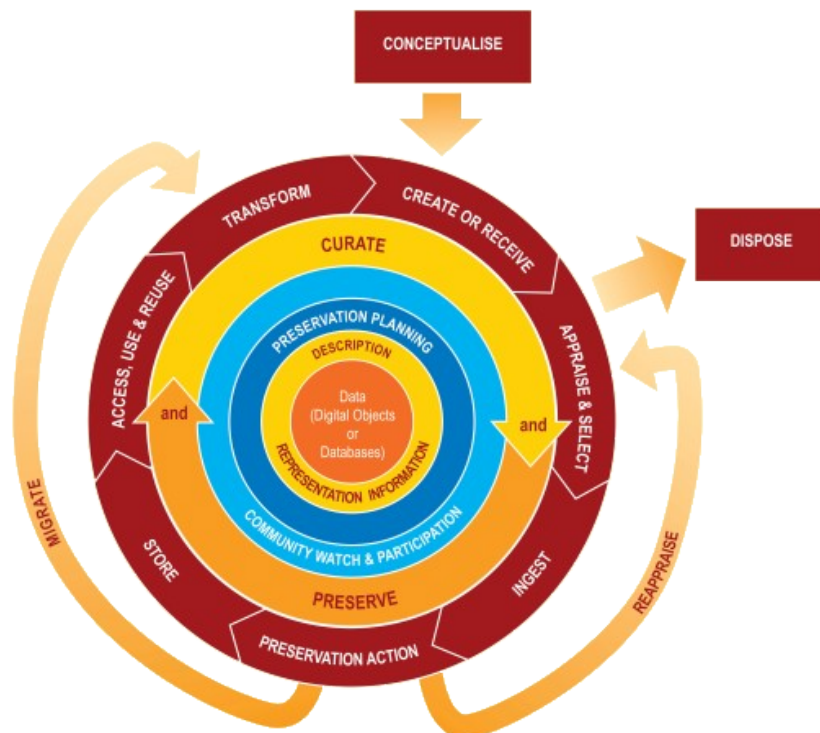


Figure 1. The DCC Curation Lifecycle Model, S. Higgins/C. Blackall/S. Fairhurst.

¹ DCC: What is Digital Curation? <http://www.dcc.ac.uk/digital-curation/what-digital-curation>

Developing a Comprehensive DMP Checklist

Analysing Research Funders' Requirements and Exemplar DMPs

DMP Online is a follow-on from an earlier piece of work - the DCC Content Checklist for a Data Management Plan (Donnelly & Jones, 2010) - which was in turn based upon the DCC's analysis of funders' requirements and a set of exemplar DMPs.

We began by comparing what the main UK research funders ask of their applicants with regard to explicit data-related statements.² There has been a long-standing expectation within some research councils (notably the Arts and Humanities Research Council (AHRC)³ and the Economic and Social Research Council (ESRC)⁴) that researchers should consider the sustainability and future use of digital outputs from the outset. As such, both Councils provide specific questions to be answered in a dedicated section of the Joint electronic Submission (Je-S) system. More recently, the Biotechnology and Biological Sciences Research Council (BBSRC)⁵, Medical Research Council (MRC)⁶ and Wellcome Trust⁷ have introduced requirements to produce a data management and sharing plan. In contrast to the AHRC and ESRC, these funders ask for a broad statement to be submitted alongside the grant proposal. Suggestions are provided for topics that could be addressed in the statement, however applicants can define the content based on the themes most relevant to their own research proposal.

As part of the DMP analysis process, we also compared guidance produced for the UK Rural Economy and Land Use (RELU) programme⁸ and the data management guidance and manual conceived by the Australian National University (ANU).⁹ The DMP templates offered by these groups are more comprehensive than the expectations of any individual funder analysed in the first phase, and so brought to light elements that could be useful for inclusion in more detailed, operational plans. We also referenced a number of existing real-world data management plans in order to check the template's completeness. Several of these came from NERC-funded centres such as the British Geological Survey (BGS) and British Atmospheric Data Centre (BADC) which write data plans for thematic programmes; so again the coverage and details were to a higher level than would be expected of DMPs at the grant proposal stage.

Developing the Content Checklist for a Data Management Plan

Having analysed and synthesised the expected coverage of DMPs - and bolstered this with our own internal expertise - we suggested two iterations of such a plan; a first ('preliminary' version) for use at the grant application stage, and a second ('extended version') to be developed at the early-project stage, and updated in conjunction with the operational plan throughout the project's lifecycle.

² Summary of UK research funders' expectations for the content of data management and sharing plans: <http://tinyurl.com/DCC-Funder-Analysis>

³ The Arts and Humanities Research Council (AHRC): <http://www.ahrc.ac.uk>

⁴ The Economic and Social Research Council (ESRC): <http://www.esrc.ac.uk>

⁵ Biotechnology and Biological Sciences Research Council (BBSRC): <http://www.bbsrc.ac.uk>

⁶ Medical Research Council (MRC): <http://www.mrc.ac.uk>

⁷ The Wellcome Trust: <http://www.wellcome.ac.uk>

⁸ Relu Data Support Service: Data management plan: <http://www.data-archive.ac.uk/relu/plan.asp>

⁹ The Australian National University (ANU) - Data Management: <http://ilp.anu.edu.au/dm/>

The preliminary version (comprising those sections given in bold type in the DCC Data Management Plan Content Checklist) covers the issues that most research funders will expect researchers to address at the application stage. These issues typically fall into five key areas:

- What data will be created (type, format) and how;
- Plans for associated metadata and documentation, noting standards to be used;
- How data will be accessed and shared, justifying any restrictions (e.g., embargoes);
- Management of Intellectual Property and ethics;
- The long-term archiving and data sharing strategy.

The extended version augments the core sections with additional information required by one or two major funders, as well as some contextual details that could usefully be included as best practice.

Public Consultation

After consulting internally among DCC colleagues, we opened the DMP Content Checklist to a public consultation via the DCC website. The clauses that populate DMP Online follow on from the post-consultation Checklist for a Data Management Plan (v2.2) (Donnelly & Jones, 2010), and take into account feedback received from a variety of stakeholders via a public consultation process. The major change between the consultation document and v2.2 is that each themed paragraph has been split into a series of atomic sections, employing closed questions where possible. The phrasing was also adjusted throughout to make greater use of the Active Voice.

Development of the Tool

The screenshot shows the DMP Online interface. At the top, there's a navigation bar with links: About DMP Online, Instructions for use, News, Documents, Privacy Statement / Terms of Use, Help. Below this, it says 'Logged in as martin' with links for My home, Admin, Logout, and Front page. The main content area features a 'Test BBSRC plan' with an '[edit details]' link. Below that is an 'Export!' button. The plan details include: Budget: £45,879.00; Running from: Apr 01 2010 to Apr 24 2010; Partners: NA. A section titled 'Funding requirements for Biotechnology and Biological Sciences Research Council:' contains a table with 8 rows of data.

Section	Total (funder clauses)	Total (DCC clauses)	Completed	Status
1 Data areas and data types	1	2	1	100% complete Go >>>>
2 Standards and metadata	2	5	0	not started Go >>>>
3 Relationship to other data available in public repositories	1	1	0	not started Go >>>>
4 Secondary use	1	1	0	not started Go >>>>
5 Methods for data sharing	2	3	0	not started Go >>>>
6 Proprietary data	1	2	0	not started Go >>>>
7 Timeframes	1	2	0	not started Go >>>>
8 Format of the final dataset	1	1	0	not started Go >>>>

Figure 2. A sample DMP homepage.

The website and user interface were designed to enable the requirements of different funders to be mapped straightforwardly to the equivalent DCC clauses, and for onscreen guidance and links to be provided to assist in the completion of DMPs. (Figure 2.)

The tool is built atop the Ruby on Rails framework, and runs on an Ubuntu GNU/Linux server via the Apache web server. Data are stored in a MySQL database, and all technologies used in its development are free or open-source. The site is hosted by the Humanities Advanced Technology and Information Institute (HATII) at the University of Glasgow, which is also responsible for the development and hosting of other digital preservation-related project sites, such as Planets¹⁰, DRAMBORA¹¹, the Data Audit Framework¹² and DigitalPreservationEurope¹³.

Users are required to register for the site. To protect against spam-generating scripts, the tool uses the reCaptcha service to verify that users are human. From a database design perspective, ‘administrator’ users have maximum flexibility in setting up the DMP forms. Funder requirements are likely to change in time, so the system enables non-programmers to edit the mappings between individual funders and the corresponding DCC clauses. This flexibility allows for: one-to-one mappings (where one funder's requirement maps directly to one DCC wording); one-to-many mappings (where a funder's requirement maps to multiple DCC questions); and one-to-none, for cases where there are no equivalent mappings to the DCC terms. The latter generally occur when the funder asks for non data-related elements to be included within a DMP (or equivalent, such as the AHRC's Technical Appendix).

It was decided not to hard-code questions into the database. Instead, an abstract system was set up whereby questions are stored in a “questions” database table. Each row of this table defines one DCC question or subject heading. The fields store the text of the question, the DCC number of the question, and a question type (text entry, true/false, or heading). Because it is important for users to be able to add and remove questions dynamically, database tables were set up to store these custom mappings.

Where a user is applying to a council which makes explicit data-related demands at the funding stage,¹⁴ the user is presented with the DCC clauses which correspond most closely; by answering the DCC clauses, the user de facto meets the funder's requirements. Where a user is applying to a funding council that does not make explicit data-related demands at the application stage, the user is presented with a superset of all of the clauses which the mapped funders require, from which the user can add or remove as desired.

At the application (pre-funded) stage, the user interface comprises four columns: the funder's requirements, the equivalent DCC clauses, user input boxes, and a fourth column giving guidance and helpful links (Figure 3). Post-funding, the first of these columns disappears to allow more room on the screen.

¹⁰ Preservation and Long-term Access through Networked Services (Planets): <http://planets-project.eu/>

¹¹ The Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) Toolkit: <http://repositoryaudit.eu/>

¹² The Data Audit Framework (DAF): <http://www.data-audit.eu/>

¹³ DigitalPreservationEurope (DPE): <http://www.digitalpreservationeurope.eu/>

¹⁴ At end-March 2010, these were: AHRC; BBSRC; ESRC; MRC; Wellcome.

Figure 3. Funder requirements are mapped to DCC clauses, and guidance is offered in the rightmost column.

An elegant interface using the jQuery Javascript/Ajax library allows the quick addition and removal of questions, and users also have the ability to export their plans as PDF files, which present information in a similar way to the onscreen interface.

Testing of DMP Online

The DCC is currently providing dedicated support for the Joint Information Systems Committee (JISC)’s Managing Research Data programme.¹⁵ Many of the projects within this programme intend to support researchers with Data Management Plan requirements. Several have already consulted the DCC’s policy and data management resources,¹⁶ and have volunteered to test DMP Online once the beta version is released in Spring 2010.

Future Developments

We previously mentioned the DMP exemplars which were used to develop the Checklist which underpins the online tool. Having sought the appropriate permissions from the originators, we may in the future wish to provide “gold-standard” examples for each section which users will be able to consult and modify for their own use.

That said, there is an acknowledged risk with this approach that people may lapse into a ‘box-ticking’ frame of mind, and thereby fail to engage adequately with the job at hand. It is therefore important to strike a balance between offering users appropriate levels of support and guidance without going so far as to render the exercise meaningless.

¹⁵ Managing research data: JISC: <http://www.jisc.ac.uk/whatwedo/programmes/mrd.aspx>

¹⁶ DCC: Policy and Legal: <http://www.dcc.ac.uk/resources/policy-and-legal/>

Conclusion

We have built a customisable online DMP template tool, into which researchers can enter their own information via an interactive Web interface, depending on their own needs and the requirements of their chosen funder. Users are able to include and exclude individual clauses according to their specific needs, and export their plans in PDF format. Onscreen guidance and suggestions for further help are 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.

References

- Donnelly, M., & Jones, S. (2010). *Checklist for a data management plan*. Version v.2.2, January 6, 2010. Edinburgh: Digital Curation Centre. Retrieved from DCC website:
http://www.dcc.ac.uk/sites/default/files/documents/tools/dmpOnline/DMP_checklist_v2.2_100106-publicVersion.doc
- Higgins, S. (2008). The DCC Curation Lifecycle Model. *International Journal of Digital Curation* 3(1). Retrieved from
<http://www.ijdc.net/index.php/ijdc/article/view/69>
- Jones, S. (2009). *A report on the range of policies required for and related to digital curation, v. 1.2*. Digital Curation Centre, UK. Retrieved March 30, 2010 from DCC website:
http://www.dcc.ac.uk/sites/default/files/documents/reports/DCC_Curation_Policies_Report.pdf
- Lyon, E.J. (2007). *Dealing with data: Roles, rights, responsibilities and relationships*. University of Bath, UK: UKOLN. Retrieved from UKOLN website:
<http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/publications.html#2007-06-19>