The International Journal of Digital Curation
Volume 7, Issue 2 | 2012

# DMP Online and DMPTool: Different Strategies Towards a Shared Goal

Andrew Sallans,
University of Virginia

Martin Donnelly,
University of Edinburgh

## Abstract

This paper provides a comparative discussion of the strategies employed in the UK's DMP Online tool and the US's DMPTool, both designed to provide a structured environment for research data management planning (DMP) with explicit links to funder requirements. Following the Sixth International Digital Curation Conference, held in Chicago in December 2010, a number of US institutions partnered with the Digital Curation Centre's DMP Online team to learn from their experiences while developing a US counterpart. DMPTool arrived in beta in August 2011 and released a production version in November 2011. This joint paper will compare and contrast use cases, organizational and national/cultural characteristics that have influenced the development decisions, outcomes achieved so far, and planned future developments.

*International Journal of Digital Curation* (2012), *7*(2), 123–129. http://dx.doi.org/10.2218/ijdc.v7i2.235

 doi:10.2218/ijdc.v7i2.235

## Background

In December 2010, a group of institutions entered into discussion around creation of a single tool to address the data management planning needs of numerous funding agencies within the United States. This new collaboration brought together librarians, IT experts and researchers, and spanned international activities from the very beginning. Recognizing the extensive thought, broad consultation and shared effort that went into the development of the first version of DMP Online[1], the US team immediately sought international collaboration to leverage this existing work and mirror the international nature of collaboration that flourishes within many research communities.

The DCC's engagement with data management planning began with a recommendation in the seminal 'Dealing with Data' report (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 DCC produced an analysis of the major UK funders' increasingly prescriptive data management requirements (Jones, 2009), and from this and our own internal knowledge and experience with digital preservation, produced the generic and comprehensive 'Checklist for a Data Management Plan' (Donnelly & Jones, 2009). Triage found that the Checklist was, by its very nature, too lengthy and too fine-grained for researchers only starting to get to grips with data management issues, so in order to overcome the inflexibility and potentially off-putting length of the Checklist, an online tool was developed.

Users are first presented with only a subset of the questions, which are selected and organized according the particular funder to whom the research proposal is being submitted, as well as the stage (pre-funding, in-project, late- or post-project) that the research has reached.

## Use Cases

DMP Online assists researchers, data custodians and other stakeholders in creating, maintaining and exporting data management plans. The tool is designed to help researchers by defining the roles and responsibilities pertaining to their data, by identifying risks that might arise at points of transition, and by ensuring an appropriate and safe chain of custody for digital data as it passes from originator to subsequent stewards.

Where a researcher is applying to a funder that has expressed explicit data-related demands at the funding stage, DMP Online presents a template made up of the DCC clauses that correspond most closely to that funder's specific instructions, the idea being that by answering the DCC clauses the researcher will at the same time meet the funder's requirements. Where a researcher is applying to a funding council or other source that does not make explicit data-related demands at the application stage, DMP Online can present a generic template that users may modify as desired.

[1] DMP Online: http://www.dcc.ac.uk/dmponline

Users of DMP Online are required to register, although use is free and open to anyone with an email address. The website and user interface were designed to enable the requirements of different funders to be mapped straightforwardly to the equivalent DCC clauses, with onscreen guidance and links provided to assist in the completion of plans.[2]

Users also have the ability to export their plans (or sub-sections thereof, in order to satisfy different stakeholder groups) in a variety of formats, including PDF and DOCX, which assist in formatting and presenting their material in the right way for different audiences, and XML and Comma Separated Values (CSV), each of which supports linkage between DMP Online and other preservation tools and research information/administrative systems.

Throughout the early months of 2011, the DMPTool[3] team focused on the initial vision of the US tool, concentrating on refinement and prioritization of system requirements. In response to immediate pressure from the research community for support in development of US National Science Foundation data management plans,[4] the team decided to focus on this as a priority use case. At a macro level, bringing all of these distinct perspectives together in a coherent way offers the advantage of understanding how research data is managed broadly, and identifying opportunities for improving practices and sharing from one domain area to another. At a micro level, for the researcher, it can serve as a single portfolio to go to identify funder data management plan requirements, institutional or community resources to support data management practices, and even as a means of connecting researchers with experts in the area(s).

## Organizational/Cultural Characteristics Influencing Decisions

The two tools had very different development paths. The founding partners of the DMPTool project consist of the University of California (specifically UC Curation Center, UCLA, and UCSD), the University of Illinois – Urbana Champaign, the University of Virginia, the Smithsonian Institution, the DataONE community, and the Digital Curation Centre. The project has also began to include contributing partners beyond the initial group, starting with the Inter-University Consortium for Political and Social Research (ICPSR). More widely, the project team has also received enthusiastic interest from a number of other institutions and government agencies, including the US National Science Foundation, the US Forest Service, the US Geological Survey, and the US National Oceanic and Atmospheric Administration, as well as several for-profit organizations that specialize in support of the research management process. The plan is to continue adding contributing partners, conducting

[2] For more on the development of the first iteration of the tool, including technical detail, see Donnelly, Jones & Pattenden-Fail, (2010).

[3] DMPTool: http://dmp.cdlib.org/

[4] As an indication of the level of interest that had been generated in the US as a result of the National Science Foundation mandate, other institutions have begun work on planning-related tools. Notably, the Distributed Data Curation Center at Purdue Libraries has developed a documentary Data Management Plan Tool taking the form of a Self-Assessment Questionnaire (http://research.hub.purdue.edu/resources/7), and University of Minnesota Libraries have also signalled an intention to develop an online tool.

 doi:10.2218/ijdc.v7i2.235

broad structured user testing on new subjects, and growing the community to ensure the broader sustainability of the service and more expansive integration of the system data and functionality with other research platforms and resources.

By contrast, the three major releases of DMP Online have been developed solely by the Digital Curation Centre, in partnership between the Universities of Edinburgh and Glasgow. A lone programmer cut the code, and frequent public stakeholder consultation was engaged in to satisfy the team that the approach taken was the right one. Multiple consultation exercises were undertaken to ensure that the tool met the needs of the respective stakeholder groups, and to bolster the quality and comprehensiveness of the underlying Checklist and accompanying guidance.

## Key Differences

There are two main differences between the tools' modes of operation. The first is cultural. DMP Online is a shared, national service which is currently hosted and administered by the DCC on behalf of the entire UK higher education sector. This approach was not felt to be ideally suited to US institutions, which identified a need for localized guidance and branding. More broadly, these differences are in response to the ways in which funder requirements have evolved. In the UK, the requirements have been overarching and consistent, while in the US responsibility has been delegated to each funder, and often each community, to indicate important characteristics for their work area.

The other key difference is philosophical. In DMPTool, the users answer the questions and requirements of the funding body directly, and are provided with guidance from two sources: the funder itself, and the user's home institution. In DMP Online, the funder requirements are mapped to a generic DCC Checklist, and users answer the DCC questions that correspond to the funders' verbatim instructions. The accompanying guidance comes from a variety of expert origins, and was crowd sourced during repeated consultation exercises.

Each approach has its advantages and restrictions. Mapping enables comparison across domain (and institutional boundaries), comparing like with like and hopefully giving the opportunity to learn more about disciplinary cultures and preservation needs. It also enables a single DMP to be created which satisfies the requirements of multiple stakeholders, be these funders, institutions, publishers, or whoever. On the downside it introduces a degree of subjectivity which can only really be satisfactorily addressed via liaison with funders and institutions, and eventually endorsement and ownership of their 'templates' within the system. Conversely, the direct approach utilized in DMPTool means that the users can be sure that they are answering the questions that the funders want answered, but the uses to which the plans can be put becomes somewhat more constrained.

## Outcomes

The first version of DMP Online launched in April 2010 at the JISC Conference in London, and a second iteration with increased functionality (including versioning of plans and requirements, additional admin reports and export options) was released in

March 2011. As a result of developments with the DMPTool team and Strand C of the JISC 07/11, a full redesign and overhaul of DMP Online took place in late 2011 and early 2012.

DMP Online v3.0 launched in May 2012. New features include multiple owners for plans, which will enable the sharing and transferring custody of plans at the appropriate stages in the research lifecycle, and support for multiple templates, enabling users to create 'hybrid' data management plans satisfying, for example, both institutional *and* funder requirements. JISC has also funded two projects to investigate integrating DMP Online with institutional administrative systems, which should serve to situate the tool within the broader research ecosystem. This has facilitated better integration with a range of preservation and administrative tools, and implementation in a cloud environment. The software is now Open Source, and the tool's future development plans are exposed more plainly.

The beta version of the DMPTool launched in August 2011, and the full production version debuted in November 2011 with a formal announcement at the Fall Forum of the Digital Library Federation in Baltimore. As with many community-supported software development efforts, the work has been done in a distributed way, with some institutions contributing programming resources and others focusing on content and workflow. Following initial requirements gathering and modelling of the tool, a first round of conceptual usability tests was conducted with a small group of researchers and support providers at the DataONE Best Practices Workshop in Santa Fe, New Mexico in May 2011. Additional testing of the beta version was completed at the annual meeting of the Ecological Society of America, which offered access to over 7,000 researchers within a domain area similar to that of the original testing group. Further testing between the beta and production releases was completed by a number of groups and institutions, led by the University of Virginia Library's User Experience team.

As of the November 2011 release, the DMPTool included funder workflows for the National Science Foundation, the National Endowment for Humanities, and the Gordon and Betty Moore Foundation. The initial release gained positive reactions and enthusiasm, and will hopefully lead to broader funding and support for continued development. The team hopes to continue to expand upon the content in the system, addressing the needs of other communities and the requirements of other funding programs, as well as the addition of more advanced administrative and management functionality within the system itself.

Over the first seven months of activity, the DMPTool saw an overwhelming community response. During this period, the tool had over 2,000 unique users, enabled the creation of over 1,700 data management plans, enrolled over 50 institutions with single sign-on capacity via Shibboleth integration, and had 19 of those institutions take the extra step of providing localized institutional guidance to their constituents. In addition, the team added additional funding agency requirements from the National Institutes of Health, the Institute of Museum and Library Services, the National Oceanographic and Atmospheric Administration, and new directorates from the National Science Foundation. Lastly, the team also established a cross-referral partnership with the DataBib project, an IMLS-funded collaborative, annotated bibliography of primary research data repositories.

 doi:10.2218/ijdc.v7i2.235

As of June 2012, DMP Online has approximately 1,000 registered users and over 1,000 plans have been created. The tool has received endorsement from a growing number of funders, and additional templates have been created for disciplines (as outputs of JISC-funded projects), institutions (via the DCC's Universities Modernisation Fund work), and publishers.

## Future Directions

While the development paths of the tools have diverged, both groups retain a broader vision of a joined-up tool (or interoperating suite of tools) that serves as a coordinating hub for the management of data across many disciplines, many funding agencies, many institutions and many countries, with shared good practice as a common goal.

With regard to DMPTool, the team believes that the current design of the system makes it possible for the research community to gain insight into the methods and practices of research data management across the entire lifecycle at both a micro and a macro level. It offers value to the individual researcher through a focused data management plan development workflow and just-in-time resource associations, while also offering high-level functionality that makes meta-analysis of data management planning practices across many domains a possibility. It offers the broad research community the opportunity to understand and refine practices for better integration of research data management processes, and consequently more interoperable and reusable data. Future efforts will be focused on developing a sustainable governance and business structure which appropriately incorporates the interests of the broad community of users and supporters, and allows for on-going development of new functionality and integration with other systems.

## Conclusion: A Shared Goal and Future Convergence

As one of us has noted elsewhere,[5] data management is a flow, a chain of events with multiple actors representing a variety of stakeholder groups. It is neither necessary nor appropriate for all stakeholders to become experts in every facet of the endeavour, but the planning process provides an opportunity to clearly stake out the roles and responsibilities for each stage of the process and to keep them up to date as requirements change over time. In order for these two tools to be more successful in their common goal of improving data management practice – and continuing to foster a data management and sharing culture – it is crucial that research data initiatives continue to develop closer liaison relationships with the funding bodies and policy makers who hold the purse-strings that capture researchers' and institutions' attentions, as well as the young researchers and research support staff who carry out and support the work. This mixture of approaches is crucial, building the bridge from both ends and balancing top-down and bottom-up, carrot with stick. And, as ever, making the case to a wide-ranging group of stakeholders remains key.

[5] For example, Donnelly, M. (2011)

## Acknowledgements

The development of DMP Online is supported by finance from JISC. The development of DMPTool is funded internally by the project consortium.

## References

Donnelly, M. (2011). Data management plans and planning. In G. Pryor (ed.) *Research Data Management.* New York: Facet.

Donnelly, M. & Jones, S. (2009 – updated 2011). *Checklist for a data management plan.* Digital Curation Centre. Retrieved from http://dmponline.dcc.ac.uk/system/attachments/8/original/DCC_Checklist_DMP_v3_md_sj.pdf?1300724157

Donnelly, M., Jones, S. & Pattenden-Fail, J.W. (2010). DMP Online: A demonstration of the Digital Curation Centre's web-based tool for creating, maintaining and exporting data management plans. In M. Lalmas et al. (eds.), *Research and Advanced Technology for Digital Libraries (ECDL 10)*. Berlin, Springer.

Jones, S. (2009). *Comparison of funder requirements*. Report for the Digital Curation Centre. Retrieved from http://www.dcc.ac.uk/resources/data-management-plans/funders-requirements

Lyon, L. (2007). *Dealing with data: Roles, rights, responsibilities and relationships.* Consultancy Report for the JISC. Retrieved from http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dealing_with_data_report-final.pdf