

The Future of Data Management Planning: Tools, Policies, and Players

Stephanie Simms & Marisa Strong
California Digital Library

Sarah Jones & Marta Ribeiro
Digital Curation Centre

Abstract

DMPonline and the DMPTool are well-established tools for data management planning. As the software of each matures and the user communities grow, we turn our attention to issues of sustainability, culture change, and international collaboration. Here we outline strategies for addressing these issues. We propose to build a new, global framework for data management planning that links plans to researchers, funders, publications, data, and other components of the research lifecycle. By refocusing our efforts from promoting the creation of data management plans (DMPs) to comply with funder requirements to supporting the creation of *good* DMPs that can be implemented, we seek to further enable the open scholarship revolution, advancing science and society.

Introduction

DMPonline and the DMPTool were developed to meet an emerging need arising from the advent of research data policies. Each is now well established as *the* resource for researchers seeking guidance in creating data management plans (DMPs). Both services, and their sponsoring organisations, the Digital Curation Centre (DCC) and the California Digital Library (CDL), have succeeded in enabling researchers in the UK and US to comply with funder requirements in producing DMPs, but this is just one step along the road to advancing open scholarship.

We see an opportunity to leverage DMPs to support open scholarship by integrating them into the broader ecosystem of data management infrastructure. In order to achieve this goal, we must redefine success to include not just adoption of our services by libraries and institutions, but also widespread adoption by individual researchers, disciplinary communities, and funders. Working together with all stakeholders to make DMPs an essential, open part of the research lifecycle, and not just a matter of compliance, is the next step toward effectively managing and sharing research data. Here we review the current landscape and explore ideas for combining distributed expertise and resources to invest in the future of DMPs.

Correspondence should be addressed to Stephanie Simms, California Digital Library, 415 20th Street, Oakland, CA 94612, USA. Email: Stephanie.Simms@ucop.edu

The 11th International Digital Curation Conference takes place on 22–25 February 2016 in Amsterdam. URL: <http://www.dcc.ac.uk/events/idcc16/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution 4.0 International Licence. For details please see <http://creativecommons.org/licenses/by/4.0/>



A Brief History of Data Management Planning and Tools

DMPonline

In 2009, the DCC undertook an analysis of curation policies (Jones, 2009) and found that an increasing number of UK research funders expected grant applicants to submit a data management and sharing plan as part of their proposal. This development was in line with a growing movement towards open access to publicly funded research outputs. Researchers were being asked to reflect on the data they were creating and how that could be managed and shared to benefit a larger community. The DCC developed DMPonline in order to help research teams meet these funder requirements, as many were unaware of what constituted good practice or how to address data management concerns in proposals. The first version of DMPonline was launched at the Jisc conference in April 2010 and it quickly received significant uptake and international recognition. DMPonline was shortlisted along with the DMPTool for the Digital Preservation Coalition's Digital Preservation Awards at the end of 2012.

Over the years, the DCC has redesigned and improved DMPonline significantly. A second iteration with increased functionality—including versioning, additional admin reports, and export options—was released in March 2011. Further changes were made as a result of developments with the DMPTool team and feedback from projects in the Jisc Managing Research Data programmes that had piloted the tool. Moreover, the DCC conducted a detailed evaluation in 2012, engaging closely with end users to understand their experience of using the tool (Getler et al., 2014). This exercise produced a great deal of positive feedback, but there were also some clear messages for change. DMPonline Version 4, which included a major redesign of the user interface and changes to the underlying data model, was released in December 2013.

The recent changes to the data model of DMPonline bring the technical platform more in line with the DMPTool. Initially, DMPonline had employed the DCC *Checklist for a Data Management Plan* (2013) as an overarching framework to ensure consistency and comparability between plans written for different contexts. The aim was to enable the creation of a single DMP that satisfied the requirements of multiple stakeholders. However, in discussion with end users the DCC discovered that this approach caused confusion and meant that the templates were often overly detailed. The revised version focuses on presenting the exact requirements from different funders or institutions and has a much cleaner interface to ensure a streamlined presentation of all the information needed.

DMPTool

Shortly after the initial release of DMPonline, it became apparent that a similar tool was needed in the US. By 2011, two major federal agencies had implemented data sharing policies in response to increasing demands for open access to the results of publicly funded research. The National Institutes of Health (NIH) was the first to require a DMP with large grant proposals. When the National Science Foundation (NSF) announced a similar requirement for all grant applicants to submit a two-page DMP, this tipped the scales and other agencies followed suit (Dietrich et al., 2012). The Interagency Working Group on Digital Data, composed of representatives from 20 federal agencies, provided general information about what DMPs should include (Interagency Working Group on Digital Data, 2009), but no clear standards or guidance for researchers to follow when

creating a DMP. Instead, they left it up to individual research communities, peer reviewers, and programme officers to define appropriate standards for themselves and refine them over time. Many of these stakeholders had never considered best practices for data management previously and lack fundamental data management skills.

Confronted with this challenge, a team of founding partners—from the CDL, DCC, DataONE, Smithsonian Institution, University of Illinois at Urbana-Champaign, University of Virginia Libraries, and University of California at Los Angeles and San Diego Libraries—came together to develop a tool that would address the specific requirements of US funders and provide a platform for information about data management planning. The team determined that a separate tool was necessary because the original data model for DMPonline was designed for the more centralised funding and policy situation in the UK. In October 2011, the team released the first version of the DMPTool, a free service that allowed users to create and download DMPs. It was an instant success, with steady uptake by individual researchers and the library community. The DMPTool also featured in the Library of Congress’ Top Ten Digital Preservation Developments of 2011 (LeFurgy, 2011).

The original partners reconvened in 2012 to enhance the functionality of the DMPTool based on user feedback and develop a strategy for supporting the needs of the entire community of stakeholders in an open source manner. The team secured funding from the Alfred P. Sloan Foundation for Version 2 and it was released in May 2014. The goals for the second version included emphasizing the importance of best practices beyond meeting the minimum requirements; enabling institutions to implement single sign-on and customise the tool with local resources; supporting collaborative plan authorship by researchers, and review and reporting by institutional admins; increasing the depth and breadth of coverage for funder and institutional requirements; creating a public library of sample DMPs; and building the diverse user communities (Strasser et al., 2014). A concurrent grant from the Institute for Museum and Library Services (IMLS) supported the development of education and outreach materials for the DMPTool2.

Usage Statistics

Total usage for both tools is presented below (see Table 1). Institutional partners for DMPonline comprise organisations with administrator rights. The majority of these are UK universities. In addition, the tool is being trialled by organisations in several other countries, including Australia, New Zealand, Belgium, Germany, France, the Netherlands, Finland, and Denmark. DMPTool institutional partners are primarily based in the US, with a handful of international partners representing universities and organisations in Australia and Singapore. There is steady growth in new institutional partners and new user registrations, with an average of 272 new users per month for DMPonline and 390 per month for the DMPTool in 2015. These numbers suggest a

Correspondence should be addressed to Stephanie Simms, California Digital Library, 415 20th Street, Oakland, CA 94612, USA. Email: Stephanie.Simms@ucop.edu

The 11th International Digital Curation Conference takes place on 22–25 February 2016 in Amsterdam. URL: <http://www.dcc.ac.uk/events/idcc16/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution 4.0 International Licence. For details please see <http://creativecommons.org/licenses/by/4.0/>



high level of interest in the tools, at least as starting points to learn about DMPs and specific funder requirements.

The number of plans per user suggests low return rates by researchers. The typical grant cycle and length of funded projects may account for some of this, but we hope to incentivise greater use of the tools through enhancements to integrate DMPs with other data management systems and align them with research workflows. Both teams anticipate that the total numbers of plans are further biased by test plans and are working on solutions to differentiate these in statistics (applying a coarse method, we identified $n = 3383$ plans with “test” in the title for the DMPTool).

The DCC maintains 13 funder templates and supports 59 templates created and managed by others. The CDL maintains 30 funder templates; more will be added in the future as agencies continue to release public access plans in response to the 2013 OSTP memo (discussed below). An additional 83 templates created by institutional partners are supported in the DMPTool.

Table 1. Global usage statistics for DMPonline and the DMPTool (as of 31 Dec 2015). All reported values are total frequencies (n).

	Plans	Users	Plans/user	Institutional partners	Funder templates	Other templates
DMPonline	8441	5300	1.59	64	13	59
DMPTool	14,068	16,360	0.86	179	30	83

Evolving Policies and Requirements

The landscape with regards to DMPs is evolving rapidly. Funders continue to release new requirements. Pilot projects to support researchers with writing and implementing DMPs within institutions and at a national level are now common. Monitoring these changes and ensuring that the templates within our tools are up-to-date are major challenges.

In the UK, the DCC checks funder requirements periodically and fosters relationships with policy officers within research funders to encourage direct notifications of changes. Analyses performed by the research data management community (e.g., Knight, 2012) and notifications to listservs also help the DCC to keep track. Some funders, such as the Medical Research Council (MRC), point researchers to DMPonline but the DCC seeks more formal endorsements and collaborations with research funders. The DCC has been in discussion with the Natural Environment Research Council (NERC) to explore whether its designated data centres could adopt DMPonline to manage the process of writing DMPs with researchers post-award. In addition, The Netherlands Organisation for Health Research and Development (ZonMw) has been piloting DMPonline in two of its programmes. In the future, the DCC hopes to work more closely with funders to help them deliver their DMP requirements and exploit the functionality of the tools for internal review and compliance monitoring.

In contrast to the more centralised organisation in the UK, US funding agencies develop and refine requirements for specific research programmes individually. The pace of these activities has skyrocketed since the US Office of Science and Technology

Policy (OSTP) memorandum of February 2013, which mandated public access to the results of federally funded research, but without providing funds to comply with the mandate. In response, all stakeholders—libraries, institutions, researchers, and funders—have been pooling scarce resources to build infrastructure and expertise for data management (Berman & Cerf, 2013). Another complicating factor is that funding agencies rarely publicise new or updated requirements. The DMPTool relies heavily on a group of intrepid data librarians who created a Google Spreadsheet¹ (Whitmire, et al. 2016) and Twitter hashtag (#OSTPResp) to crowd-source the effort of tracking agency responses to the OSTP memo. The DMPTool team is also attempting to foster personal contacts with policy officers but this approach is difficult to scale owing to the decentralised system. Moreover, agencies have avoided endorsing anything outside of federal organisations and repositories, although there is hope that this might change as data management policies mature.

Community Engagement

It is widely acknowledged that community engagement on a global scale is critical for translating investments in research and research data into openly available, well-managed, reusable information (e.g., Bourne et al., 2015). Community engagement for DMPs encompasses a range of user groups, international policy bodies, the open source development community, and of course funders. Expanding our capacity for outreach, training, and engagement with these diverse communities is important for mutual sustainability as well as fostering the cultural changes taking place in higher education and the global research community.

The primary users of our tools are researchers, many of whom regard DMPs as a bothersome box to tick during the grant application process. It is therefore important for the guidance provided within the tools convey to researchers why they should care about a DMP and how DMPs can be beneficial to their research. At many institutions, library staff are addressing the need for education about research data management and providing consultation services for researchers (Rolando et al., 2015; Tenopir et al., 2011; Tenopir et al., 2015). These librarians, data managers, and IT professionals represent another crucial user group as they administer the tools for their institutions and customise the templates with local resources. The tools must also incorporate emerging international standards for data management and scholarly communications issued by the Research Data Alliance (RDA), DataCite, the Consortia Advancing Standards in Research Administration Information (CASRAI), and others.

To improve usability, add new features, and respond to new trends in data management practices, both the DCC and CDL are encouraging external developers to

¹ See the crowd-sourced spreadsheet at: <http://bit.ly/FedOASummary>

Correspondence should be addressed to Stephanie Simms, California Digital Library, 415 20th Street, Oakland, CA 94612, USA. Email: Stephanie.Simms@ucop.edu

The 11th International Digital Curation Conference takes place on 22–25 February 2016 in Amsterdam. URL: <http://www.dcc.ac.uk/events/idcc16/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution 4.0 International Licence. For details please see <http://creativecommons.org/licenses/by/4.0/>



make contributions to the open source code bases. The DMPonline code has been implemented by Queensland University of Technology² (QUT) in Australia, the University of Alberta³ on behalf of the Canadian Association of Research Libraries (CARL) Portage group, and a number of pilot installations are currently underway across Europe. The DCC is beginning to reap the benefits of community code development. Pull requests from QUT, the Canadian consortium, and various UK developers have helped to fix bugs, update gems, and add new features for the benefit of the wider community.

Integrations between systems that support different data management activities and across national borders are other promising frontiers. The CDL is working with a developer who received a grant from the non-profit Center for Open Science to connect their Open Science Framework data management platform with the DMPTool. A collaboration project from RDA Europe will enable the DCC to embed the Metadata Standards Directory into DMPonline. The DCC was also awarded a grant from the University of Edinburgh to internationalise DMPonline. As a result of the European Commission's Open Research Data Pilot under Horizon 2020, the DCC has seen a surge of interest in DMPonline across Europe. The tool has been translated into several foreign languages, and there is demand for national versions so that local organisations and funders can be presented depending on the country context. The award from Edinburgh will allow the DCC to extend the functionality of DMPonline and provide these options. As more initiatives, support projects, and tools for data management planning emerge worldwide, it is critical that we coordinate and build on existing resources.

Sustaining Plans, Sustaining Tools

Both DMPs and the services that support their creation have become prime movers in the global push toward open scholarship. As a result, both of our organizations are committed to sustaining and improving data management planning efforts. By collaborating and sharing the expertise gained through the DMPonline and DMPTool initiatives, we hope to expand our reach to create global consensus about policies and best practices. The ultimate aim is to promote tools not only to create a DMP, but to create a good DMP and to enable its implementation. This goal is tied into slow but steady culture changes in the academy, scholarly publishing, and a growing number of government policies. The extent to which a DMP is regarded as a living document, and not just a matter of compliance, will determine the success of open science/open data movements, help incentivise researchers to invest time in managing their data, and enable collaborative research.

In light of the growing interest in DMPs and the increasing number of requirements emerging worldwide, a lot of expectation is placed on the DCC and CDL as service providers. It is challenging to keep up with the queries and new feature requests with existing resources. In order to maintain the tools and ensure their relevance to the global research community, we are exploring co-development of a single technical platform, planning assessment strategies so that we can be more responsive to shifting user needs, and devising sustainable funding models. Specific approaches to furthering a global

² See the QUT Data Management Planning tool at: <https://dmp.qut.edu.au>

³ See DMP Assistant at: <https://portagenetwork.ca>

DMP agenda include mining DMPs to understand and report current practices, integrating the tools with other research data systems to facilitate institutional processes, and promoting interoperability through community standards for DMPs. Achieving sustainability requires synergy among all stakeholder communities; no group or organisation can solve data management challenges individually and each contributes diverse experiences, expertise, and resources.

Technical Strategies

In the beginning, the underlying data models were different for DMPonline and the DMPTool, but over the years these models have converged to the point that it makes sense to co-develop a single tool. Both tools only require registration with an email address. Both continue to experience high uptake by institutions and offer institutional branding, the ability to customise with local resources, and methods to generate usage statistics. The basic functionality of the tools is also the same: users can create, maintain, and export a DMP to submit with a grant application. The primary difference lies in the templates, which address the requirements of specific national funders, including guidance text and relevant links. A common platform would provide a one-stop-shop for learning about data management planning, and ideally templates could be managed by funders and/or disciplinary communities.

The DCC and CDL began preliminary discussions about converging on a single codebase in Autumn 2015. An initial comparison of the systems shows that the main functionality is shared across the two tools, but that the implementation differs. For example, DMPonline uses the concept of themes to enable administrators to manage their guidance resources more effectively: advice can be written once according to common themes and applied across all templates. This avoids redundancy and minimises the changes institutions need to make when funder templates are updated. We plan to review and extend these themes jointly to ensure they meet global requirements.

Our intention is to develop a prototype of a single platform based on the DMPonline codebase, incorporating additional functionality such as plan review, sharing, and status indicators from the DMPTool. We will also extend on forthcoming enhancements to the tools such as internationalisation and integration projects, and issue a joint roadmap for future work. By pooling our development effort and co-managing a single product, we will gain efficiencies and be better placed to respond to changes. Together we can deliver an enhanced data management planning platform and drive an agenda that moves DMPs beyond a compliance checkbox.

The common platform should provide the necessary hooks and APIs for integration with the full range of RDM activities. This includes treating the DMP as a dynamic checklist for project activities and research outputs, a means for determining compliance with stated goals, providing a repository recommendation service, more robust

Correspondence should be addressed to Stephanie Simms, California Digital Library, 415 20th Street, Oakland, CA 94612, USA. Email: Stephanie.Simms@ucop.edu

The 11th International Digital Curation Conference takes place on 22–25 February 2016 in Amsterdam. URL: <http://www.dcc.ac.uk/events/idcc16/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution 4.0 International Licence. For details please see <http://creativecommons.org/licenses/by/4.0/>



intelligence for business reporting, interoperation with sponsored research administration systems and researcher profiles, and other priorities to be jointly determined.

Financial Strategies

To date, both tools have been developed and supported primarily through grant funding. Moving forward we must explore other options and avoid overreliance on a single funder. Jisc has funded the majority of the work for DMPonline with additional sources of revenue from individual universities and European Commission funded projects. The DMPTool has been funded by the Alfred P. Sloan Foundation, the IMLS, and the CDL, with additional support from project partners. We hope to diversify funding sources, particularly by pursuing collaborations with research funders for mutual benefit.

The DCC is currently exploring options to introduce charges for delivering DMPonline services outside of the UK. The tool will remain free for individuals wishing to create DMPs, however, models are being introduced to recover costs for delivering separate instances of DMPonline or supporting organisations to customise the tool. The award to internationalise DMPonline will allow the team to develop locale-aware support that could be provided as a chargeable service. This will enable the content presented in DMPonline to be tailored to a given context. For example, the available languages, organisation lists, and templates will vary for different countries or organisations. Local user authentication procedures, date/time, and data format conventions will also be supported, as well as custom interfaces to brand the tool as a local service. Based on initial support requests, the DCC sees potential to deliver DMPonline services on behalf of individual institutions, as well as agreeing contracts with national agencies to deliver data management planning tools and support in a similar vein to the Jisc subscription model for the UK.

Subscription models are also being considered, for example, for individual organisations that wish to add templates and guidance to DMPonline but do not require the additional locale-aware functionality or want a separate hosted instance of the tool. The templates maintained by the DCC may also be provided on a subscription basis for those who wish to deploy a local instance of the code but do not wish to maintain a policy watch function. Business models to deliver services outside of the UK are currently being developed and will be agreed in consultation with the user community. At present, the CDL has no plans to charge users of the DMPTool, which will retain its brand and continue to support templates for US funders.

Conclusion

Data management is now acknowledged as a global challenge: research is global, policies are becoming global, and thus the need is global. Successful strategies for meeting this need will require coordination of all efforts—infrastructure and education—at the global scale. The DCC and CDL have collaborated informally from the outset to reduce needless duplication of effort and share experiences (Sallans & Donnelly, 2012). But as new initiatives and requirements for data management planning and open scholarship multiply worldwide, we seek to establish a closer working partnership in order to maximise investments towards a common goal: leveraging DMPs to promote good practices that enable data sharing and reuse, and ultimately

advance open scholarship. Our organizations are well positioned to support the research community in achieving this goal with our deep knowledge of the technical as well as the community aspects of research data management in different national contexts. Given the level of interest in the tools, the range of groups with which we are engaging, and our ongoing collaboration, we see great potential in the future of data management planning.

Acknowledgements

DMPonline has been developed by the DCC and is maintained with financial support from Jisc and the University of Edinburgh. The DMPTool was developed with funding from the Alfred P. Sloan Foundation, the IMLS, and the CDL in collaboration with project partners.

References

- [journal article] Berman, F. & Cerf, V. (2013). Who Will Pay for Public Access to Research Data? *Science*, 341, 6146, 616–617. doi:10.1126/science.1241625
- [journal article] Bourne, P. E., Lorsch, J. R. & Green, E. D. (2015). Perspective: Sustaining the big-data ecosystem. *Nature*, 527, S16–S17. doi:10.1038/527S16a
- [journal article] Dietrich, D., Adamus, T., Miner, A., & Steinhart, G. (2012). Demystifying the data management requirements of research funders. *Issues in Science and Technology Librarianship*, 70. doi:10.5062/F44M92G2
- Digital Curation Centre. (2013) *Checklist for a Data Management Plan*. Version 4.0 Retrieved from <http://www.dcc.ac.uk/resources/data-management-plans/checklist>
- [journal article] Getler, M., Sisu, D., Miller, K., & Jones, S. (2014). DMPonline Version 4.0: User-Led Innovation. *International Journal of Digital Curation*, 9, 1, 193–219. doi:10.2218/ijdc.v9i1.312
- [report] Interagency Working Group on Digital Data. (2009). Harnessing the Power of Digital Data for Science and Society: Report of the Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council. Retrieved from: http://www.nitrd.gov/About/Harnessing_Power_Web.pdf

Correspondence should be addressed to Stephanie Simms, California Digital Library, 415 20th Street, Oakland, CA 94612, USA. Email: Stephanie.Simms@ucop.edu

The 11th International Digital Curation Conference takes place on 22–25 February 2016 in Amsterdam. URL: <http://www.dcc.ac.uk/events/idcc16/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution 4.0 International Licence. For details please see <http://creativecommons.org/licenses/by/4.0/>



- [report] Jones, S. (2009). A report on the range of policies required for and related to digital curation, v. 1.2. Digital Curation Centre, Edinburgh. Retrieved from http://www.dcc.ac.uk/sites/default/files/documents/reports/DCC_Curation_Policies_Report.pdf
- [report] Knight, G. (2012). Funder requirements for data management and sharing. Project report. London School of Hygiene and Tropical Medicine, London. Retrieved from <http://researchonline.lshtm.ac.uk/208596>
- [web log post] LeFurgy, B. (2011). Top 10 digital preservation developments of 2011. Retrieved from The Signal: Digital Preservation blog: <http://blogs.loc.gov/digitalpreservation/2012/01/top-10-digital-preservation-developments-of-2011/>
- [journal article] Rolando, L., Hswe, P., Wells Parham, S., Westra, B., & Whitmire, A.L. (2015). Data Management Plans as a Research Tool. *Bulletin of the American Society for Information Science and Technology* 41(5), 43–45. doi:10.1002/bult.2015.1720410510
- [journal article] Sallans, A. & Donnelly, M. (2012). DMPonline and DMPTool: Different Strategies Towards a Shared Goal. *International Journal for Digital Curation*, 7, 123–129. doi:10.2218/ijdc.v7i2.235
- [journal article] Strasser, C., Abrams, S., & Cruse, P. (2014). DMPTool2: Expanding functionality for better data management planning. *International Journal of Digital Curation*, 9, 1, 324-330. doi:10.2218/ijdc.v9i1.319
- [journal article] Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLoS ONE*, 6(6), e21101. doi:10.1371/journal.pone.0021101
- [journal article] Tenopir, C., Hughes, D., Allard, S., Frame, M., Birch, B., Baird, L., Sandusky, R., Langseth, M., & Lundeen, A. (2015). Research Data Services in Academic Libraries: Data Intensive Roles for the Future? *Journal of eScience Librarianship* 4(2), e1085. doi:10.7191/jeslib.2015.1085
- [web log post] Whitmire, A., Briney, K., Nurnberger, A., Henderson, M., Atwood, T., Janz, M., Kozlowski, W., Lake, S., Vandegrift, M., & Zilinski, L. (2015). A table summarizing the Federal public access policies resulting from the US Office of Science and Technology Policy memorandum of February 2013. figshare. <http://dx.doi.org/10.6084/m9.figshare.1372041>. Retrieved 24 January 2016.