IJDC | *General Article*

DMPTool 2: Expanding Functionality for Better Data Management Planning

Carly Strasser, Stephen Abrams,
and Patricia Cruse
California Digital Library

Abstract

Scholarly researchers today are increasingly required to engage in a range of data management planning activities to comply with institutional policies, or as a precondition for publication or grant funding. The latter is especially true in the U.S. in light of the recent White House Office of Science and Technology Policy (OSTP) mandate aimed at maximizing the availability of all outputs – data as well as the publications that summarize them – resulting from federally-funded research projects.

To aid researchers in creating effective data management plans (DMPs), a group of organizations – California Digital Library, DataONE, Digital Curation Centre, Smithsonian Institution, University of Illinois Urbana-Champaign, and University of Virginia Library – collaborated on the development of the DMPTool, an online application that helps researchers create data management plans. The DMPTool provides detailed guidance, links to general and institutional resources, and walks a researcher through the process of generating a comprehensive plan tailored to specific DMP requirements. The uptake of the DMPTool has been positive: to date, it has been used by over 6,000 researchers from 800 institutions, making use of more than 20 requirements templates customized for funding bodies.

With support from the Alfred P. Sloan Foundation, project partners are now engaged in enhancing the features of the DMPTool. The second version of the tool has enhanced functionality for plan creators and institutional administrators, as well as a redesigned user interface and an open RESTful application programming interface (API).

New administrative functions provide the means for institutions to better support local research activities. New capabilities include support for plan co-ownership; workflow provisions for internal plan review; simplified maintenance and addition of DMP requirements templates; extensive capabilities for the customization of guidance and resources by local institutional administrators; options for plan visibility; and UI refinements based on user feedback and focus group testing. The technical work undertaken for the DMPTool Version 2 has been accompanied by a new governance structure and the growth of a community of engaged stakeholders who will form the basis for a sustainable path forward for the DMPTool as it continues to play an important role in research data management activities.

Received 27 October 2013 | *Accepted* 26 February 2014

Correspondence should be addressed to Carly Strasser, California Digital Library, 415 20th Street, Oakland CA 94612 USA. Email: carly.strasser@ucop.edu

An earlier version of this paper was presented at the 9th International Digital Curation Conference.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Introduction

The Data Deluge

We are at an unprecedented turning point in scholarly communication, which has been brought on by the increasingly important role of data-intensive research. Datasets and their accompanying metadata are the currency of scientific and intellectual advancement, and deserve the same amount of attention, planning and scrutiny that publications receive. The move towards digital data has created a new set of issues: how does one handle the huge volume of available information effectively and efficiently to solve important problems? Knowledge of good data management techniques and software development lags behind the progression digital data. Consequently, researchers often do not have the skills to handle their datasets. This challenge is amplified by the fact that research data are seldom shared, re-used, or preserved (Tenopir et al., 2011). There is a growing awareness among practitioners and funders that this situation represents inefficient use of research dollars, missed opportunities to exploit prior investment, and a general loss for the scholarly community. Michener et al. (1997) described the loss of valuable data and insight about those datasets as "information entropy". This loss of information is becoming increasingly worrisome, as data management practices improve very slowly while the volume of data grows exponentially.

Data Management Plans

In a measure aimed at improving data stewardship for projects funded by the public, the Federal Funding Accountability Transparency Act was passed in 2006, ensuring that the public can access information on all entities and organizations receiving Federal funds[1]. This resulted in procedural changes at the National Institutes of Health, which began requiring data management plans for large grant proposals. Several government funders have since followed suit, including the National Science Foundation (NSF), National Oceanic and Atmospheric Administration (NOAA), U.S. Department of Agriculture (USDA), the Environmental Protection Agency (EPA) (Dietrich et al., 2012). These funding agencies provide general information about what data management plans (DMPs) should include, but are often vague about requirements and provide few resources for researchers to consult when creating their DMPs. Without the proper training or background in data management and digital curation, researchers are apt to continue their current, uninformed and incomplete data stewardship practices.

The DMPTool

In response to this need, eight institutions and organizations jointed together to create the DMPTool. The multi-institutional partnership was comprised of staff from the University of California Curation Center, California Digital Library, University of Virginia Library, University of Illinois at Urbana-Champaign, Smithsonian Institution, DataONE[2], University of California San Diego and Los Angeles Libraries, and the Digital Curation Centre. Creating the tool was a potentially risky venture that had no

1 Federal Funding Accountability and Transparency Act of 2006: Public Law 109-282, enacted 26th September 2006. Available at: http://www.gpo.gov/fdsys/pkg/PLAW-109publ282/pdf/PLAW-109publ282.pdf

2 DataONE: http://www.dataone.org

 doi:10.2218/ijdc.v9i1.319

dedicated financial support, relying fully on in-kind contributions and collaboration; the partners were rewarded for this entrepreneurial spirit by an impressive uptake of the tool by the community.

Starting in January 2011, the team began development of a freely-available web application that guides users through the process of creating a DMP for a range of funders. The resulting DMPTool (Version 1) allows users to edit, save, share, print and download their data management plans. Since its release in October 2011, the DMPTool has generated huge interest among the science and library community, including its appearance on the Library of Congress' Top Ten Digital Preservation Developments of 2011 (LeFurgy, 2011). As of December 2013, more than 6,000 users from over 800 institutions have used the DMPTool (Figure 1).

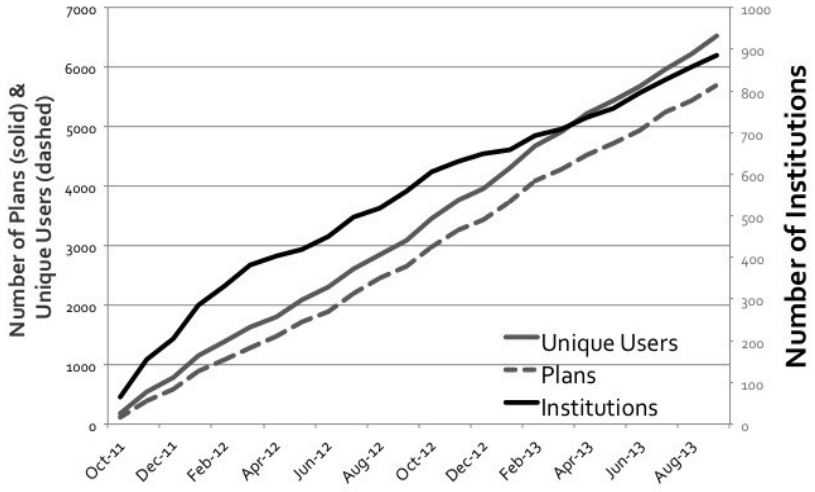

Figure 1. Uptake of the DMPTool (Version 1).

The success of DMPTool Version 1 prompted the DMPTool group to request funding from the Alfred P. Sloan Foundation[3] to develop Version 2 (hereafter DMPTool2); this funding was secured in January 2013. While effective data management and sharing is one of the most pressing issues facing scholarly communication, no single group owns the problem. It takes a diverse approach that leverages capacity and knowledge from a community of stakeholders, including libraries, researchers, funders and institutions. The intent of the DMPTool2 project was to help provide a platform around which this community of stakeholders can assemble and collaboratively work to meet the needs of the whole in an efficient manner.

DMPTool2 Project

The Approach

The DMPTool2 project focused on meeting the needs of the identified DMPTool user constituencies (researchers, libraries, funders and institutions) in a community-supported, open source manner. The DMPTool provides a centralized point for the

3 Alfred P. Sloan Foundation: http://sloan.org

consolidation, sharing and dissemination of expertise and advice in good data management practices and solutions. Institutions can avoid developing programs in isolation, with no knowledge of practices, tools, techniques and resources that have been tested or already exist elsewhere. While it is possible for individual idiosyncratic management regimens to be successful, the overall quality and success of DMPs are likely to be higher if they are created using commonly accepted standards and practices based on requirements and recommendations of funders, institutions, librarians and research communities. To address these challenges the DMPTool2 project focused on the following project goals:

- Promote the importance and practice of data management, sharing and preservation, emphasizing the utility of best practices beyond meeting funder requirements;
- Enable institutions to easily shape and exploit the DMPTool to meet local needs;
- Foster the emergence of an engaged open source community of DMPTool users and developers;
- Maintain transparency in all project activities to facilitate community involvement;
- Increase the depth and breadth of the tool's coverage for funder and institutional data management requirements;
- Enable collaboration and connect stakeholders to each other and institutional resources, including services, expertise and guidance; and
- Provide support for the full data management life cycle, including provision for collaborative plan authorship, and review and reporting by institutional administrators.

High-Level Requirements

The DMPTool project partners met in early 2012 to discuss the success of the tool, review feedback from users, and begin developing functional requirements for future development. Over the next nine months, these requirements were shared with the community via workshops, meetings and social media; refined; and fleshed out into technical requirements that were used to plan for the next phase of the tool's development. Our major focus for the development of requirements was that the tool should meet the needs of both types of end users: plan creators and administrators. These needs were assessed in an ad hoc fashion based on community feedback regarding the existing DMPTool, plus informal conversations, workshops, presentations and social media. The project partners also formed two advisory boards, the Researcher Advisory Board[4] and the Administrative User Advisory Board[5], to guide our efforts and decision-making. In addition to the users' needs, we paid careful attention to potential needs of institutions themselves. To that end, one of the original project partners, the Smithsonian Institution, provided a detailed institutional use case that served as the foundation for allowing template creation at the institutional level.

4 Researcher Advisory Board: https://bitbucket.org/dmptool/main/wiki/ResearcherAdvisoryBoard
5 Administrative User Advisory Board: https://bitbucket.org/dmptool/main/wiki/AdministrativeUserAdvisoryBoard

Development Process

The requirements were used as the guidelines for the development process. A full set of the requirements used to guide the DMPTool2 development is available online[6]. Use cases were used to frame the development (see below). Based on the thorough requirements created by the DMPTool2 project team, we developed wireframes that underwent extensive review. These wireframes were then shared with a design firm that developed the basic colour scheme and flow, focusing on the user's visual experience with the website. Finally, the designs were combined with the functional software developed in Bootstrap to create the final website.

New Features for DMPTool2

Dashboard

Once logged into the tool, both types of users (creators and administrators) are presented with a dashboard that provides a list of active DMPs, their current status in the tool, and templates for creating new plans.

Administrative interface

Institutional administrators will be able to provide customized help text, resources, and suggested answers for researchers from their institutions via the new administrative interface. Among the most important new features for administrators is the ability to create data management plan templates for use by researchers from their institutions. These templates can be generated for different groups on campus, with an eye towards discipline-specific questions, help text, or helpful resources. The new administrative interface is self-service, which means there is no need to contact personnel from the DMPTool team to update the institution-specific resources or information.

Plan co-creation

Grant writing does not happen in a vacuum. While one person may be responsible for the creation of the document itself, there are stakeholders within and outside of the project that will want to contribute to the creation of the data management plan. In addition, there may be multiple researchers working on a single project that requires a data management plan. In response, the DMPTool2 allows for co-ownership of plans. Plan creators can designate specific users, giving those users ability to edit and provide feedback on their data management plan.

Sample DMPs library

Plan creators can share their data management plans either publicly or with others in their institution. These sample DMPs are collected in a searchable webpage on the DMPTool2 website. Our hope is that the DMPTool2 will, over time, build up a significant number of sample DMPs that can be used as reference materials for plan creators as they generate their own plans.

Submitting for review

Administrators have the option to allow plan creators from their institution to submit their completed data management plans for review. Users that have been granted

6 DMPTool2 Requirements: https://bitbucket.org/dmptool/main/downloads/DMPT2-requirements-v14.pdf

'reviewer' permission (e.g., a librarian who has been trained in data management) will be able to comment on the plan before it is exported and added to the grant as a whole. This will facilitate communication between plan creators, institutional administrators, and other campus stakeholders in data management.

DMPTool Governance and Partnership

Steering Group

The DMPTool Steering Group is a small coordinating body composed of persons elected from the original DMPTool partner institutions. The DMPTool Steering Group directs the technical, content and community development of the DMPTool. They collect feedback and suggestions from the DMPTool partners.

DMPTool Partners

DMPTool partners are institutions, profit and nonprofit entities, or other groups that are committed to the DMPTool's utility and success as an effective and efficient way to create data management plans. Our goal in establishing and recognizing our partner institutions is to build a committed community of users that sustain and support the evolution of the DMPTool. Institutions can become partners by agreeing to the Partner Principles below (no signature required), plus completing at least one of the following:

1. Establishing institutional authentication with the DMPTool (e.g., Shibboleth);
2. Customizing the tool with resources, help text, suggested answers, or other information; and/or
3. Contributing to the maintenance and enhancement of the DMPTool codebase.

Partner Principles

Our work (past, current, and future) on the DMPTool is guided by the principles below. Institutions and organizations that partner with the DMPTool are expected to understand and abide by these principles:

- Continuous improvement of the DMPTool's utility and features;
- User-driven requirements and priorities for development;
- Community-driven forward progress on tool improvement;
- Enthusiasm for the DMPTool;
- Integrity of plans created and resources offered;
- Commitment to an open process for development, enhancement and improvement; and
- Quality of code and materials created.

Lessons Learned

We attribute the success of the DMPTool2 project to several things. First, the careful creation of the requirements documentation prior to the start of the DMPTool2 project was crucial to the project's success. The requirements were often consulted in the course of development to prevent scope creep, guide thinking, and to ensure that project goals were being met. Second, the project's many partners required that we clearly defined and distributed the different roles in the project to both ensure complete coverage and prevent overlap in work efforts. Third, the first version of the tool provided us with a venue for collecting invaluable feedback on user needs, ideal features and potential functionality for the DMPTool2. Finally, we were able to combine the efforts of the DMPTool2 project funded by the Alfred P. Sloan Foundation with a related grant from the Institute for Museum and Library Services, which focused on developing outreach and education materials for the DMPTool.

References

Dietrich, D., Adamus, T., Miner, A., & Steinhart, G. (2012). De-mystifying the data management requirements of research funders. *Issues in Science and Technology Librarianship, 70*. doi:10.5062/F44M92G2

LeFurgy, B. (2011). Top 10 digital preservation developments of 2011 [Web log post]. Retrieved from The Signal: Digital Preservation blog: http://blogs.loc.gov/digitalpreservation/2012/01/top-10-digital-preservation-developments-of-2011/

Michener, W., Brunt, J., Helly, J., Kirchner, T., & Stafford, S. (1997). Non-geospatial metadata for the ecological sciences. *Ecological Applications, 7*(1), 330-342. doi:10.1890/1051-0761(1997)007%5B0330:NMFTES%5D2.0.CO;2

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