

“DMPTool: supporting the data lifecycle”

Submitted as a position paper for the 2011 NSF Research Data Lifecycle Management Workshop (<http://rds.columbia.edu/rdlm>) at Princeton University.

By Andrew Sallans (University of Virginia, als9q@virginia.edu)

Abstract:

This position paper is intended to raise awareness about the current “DMPTool” project, which aims to provide a single environment for development of research data management plans. The project has two principal goals: 1) simple and straightforward production of compliant data management plans, and 2) connection of researchers with available resources to support components of their data management plans. Feedback from the HPC community at the workshop will prove highly-valuable in continued development of the project.

Statement:

Starting in December 2010, a group of institutions began discussion around creation of a single tool to address the data management planning needs of numerous 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 and effort that went into the development of the first version of the UK’s Digital Curation Centre’s (DCC) “DMP Online” tool, the U.S. version immediately sought international collaboration to leverage existing work and complement the nature of international collaborations within many research communities.

Throughout the early months of 2011, the project team focused on the initial vision of the tool and aimed at refinement and prioritization of tool requirements. In response to immediate pressure from the research community for support in development of US National Science Foundation data management plans (DMPs), the team decided to focus on this as a priority use case. Despite that priority, the team still holds a broader vision of a tool that serves as a coordinating hub between the management of data across many disciplines, many funding agencies, many institutions, and many countries. 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.

The founding partners of the 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. In addition, the project has begun to include contributing partners, starting with the Inter-University Consortium for Political and Social Research (ICPSR). Beyond these initial partners, 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. We plan to continue adding contributing partners and growing the community to ensure broader sustainability of the service and more expansive integration of the system data and functionality with other research platforms and resources.

We are now fully-engaged in the first software development phase of the project, with significant effort going into translation of abstract models and requirements into specific features and functionality. As with many community-supported software development efforts, the work is being done in a distributed way, with some institutions contributing programming resources and some focusing on content and workflow. Following initial requirements gathering and modeling of the tool, we conducted a first round of conceptual usability tests with a small group of researchers and support providers at the DataONE Best Practices Workshop in May 2011. Our target release for a full phase one (“beta”) release of the project is now early August 2011, to be debuted at the annual meeting of the Ecological Society of America. This venue will offer access to over 7,000 researchers, within a domain area that is similar to that of our original testing. At the conference, we plan to release the tool through several DataONE data management education workshops, as well as through direct user interaction at the information booth. Following this release, we will continue to expand upon the content in the system, addressing the needs of other communities and requirements of other funding programs, as well as the addition of more advanced administrative and management functionality within the system itself.

We believe 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 data itself.

It is our hope that the participants of the Research Data Lifecycle Management Workshop will consider the possibilities of such a tool and how it might relate to the unique demands of the high-performance computing community. We look forward to the opportunity to discuss the project and new ideas.

Project websites:

- <http://www.cdlib.org/uc3/datamanagement/dmpo.html>
- <https://bitbucket.org/dmptool/main/wiki/Home>

Primary contacts:

- Patricia Cruse (Patricia.Cruse@ucop.edu)
- Andrew Sallans (als9q@virginia.edu)

Project Team

Digital Curation Centre (UK)

- Kevin Ashley, Director
- Martin Donnelly, Curation Research Officer

DataONE

- Amber Budden, Director for Community Outreach

Smithsonian Institution

- Günter Waibel, Director, Digitization Program Office, Office of the CIO

UC Curation Center, California Digital Library

- Trisha Cruse, Director
- Perry Willett, Service Manager

UCLA Library

- Todd Grappone, AUL for Digital Initiatives and Information Technology

UC Merced Library

- Donald Barclay, Deputy University Librarian

UC San Diego Libraries

- Bradley D. Westbrook, Metadata Librarian and Digital Archivist

University of Illinois at Urbana-Champaign

- Michael Grady, Executive Program Officer for CI, Office of the CIO
- Sarah Shreeves, IDEALS Coordinator and Scholarly Commons Coordinator, UL

University of Virginia Libraries

- Andrew Sallans, Head, Strategic Data Initiatives / Sci. Data Consulting Group
- Sherry Lake, Scientific Data Consultant, Scientific Data Consulting Group



"DMPTool: supporting the data lifecycle" by Andrew Sallans is licensed under a [Creative Commons Attribution-ShareAlike 3.0 Unported License](https://creativecommons.org/licenses/by-sa/3.0/).