

DMPRoadmap: Making Data Management Plans Actionable

The California Digital Library (CDL), in partnership with the Digital Curation Centre (DCC), requests \$731,989, matched with \$755,660 in cost share, for a two-year Project Grant in the National Digital Platform category to enhance DMPRoadmap¹, a new, internationalized platform formed by the merger of the DMPTool and DMPonline into a single technical platform. This core data management planning infrastructure will reposition data management plans (DMPs) as living and machine-actionable documents useful for all stakeholders in the research enterprise.

At present, most researchers treat DMPs as an annoying box to tick during the grant application process. This is partly because the information contained in static DMPs cannot be leveraged for subsequent research activities (e.g. reporting). Federal funders—required by the 2013 OSTP Memo to ensure public access to research data—note that the DMPs they receive tend to be of poor quality and can only be monitored manually. Despite these challenges, DMP support services are in high demand and libraries have stepped in to meet the need. The cornerstone of library support in the US is the DMPTool, developed in 2011 by the CDL and a group of founding collaborators— including the DCC —to provide a central clearinghouse of information about US funder requirements, data management standards, and guidance. Since then, the CDL has continued to develop the DMPTool and track evolving funder requirements, and the user community has expanded to include 188 institutions, 20,000 users, and over 13,000 plans. The Public DMPs and educational resources also receive heavy traffic. The DCC has parallel experience with DMPonline, a similar tool that addresses UK funder requirements and is now being adopted across Europe, Canada, and Australia with the proliferation of open data policies. At this stage, we have two popular platforms that will soon become one (DMPRoadmap), an engaged and increasingly international community, recognition among funders, and we have learned a lot about DMPs. It is now time to redefine success and revisit the question: What was the point of DMPs in the first place?

Answer: to get researchers to think about how to produce and publish FAIR data, i.e., data that is Findable, Accessible, Interoperable, and Reusable.² In order to achieve this goal, we still need a human-readable narrative, but underneath the DMP should have more thematic richness with added value for all stakeholders. Moving toward machine-actionable³ DMPs has been a topic of conversation for some time now, and has spawned multiple international working groups (e.g., RDA, CASRAI, FORCE11), but with few practical results. There is consensus, however, about the need for more and better example DMPs, better training, and a common structure for DMPs across all funders and national boundaries. The DMPRoadmap project is ideally positioned to consolidate the interests of these stakeholder groups with our experience, connections, and proven ability to promote the creation of good DMPs. We seek funding to add support for machine-actionable DMPs into the

¹ <https://github.com/dmproadmap>

² www.force11.org/group/fairgroup/fairprinciples

³ The Data Documentation Initiative defines machine-actionable as “information that is structured in a consistent way so that machines, or computers, can be programmed against the structure” (www.ddialliance.org/taxonomy/term/198).

DMPRoadmap platform that will enable institutions to manage their data, funders to mine the DMPs they receive, infrastructure providers to plan their resources, and researchers to discover useful data.

Work plan

In order to transform DMPs into a dynamic inventory of digital research methods, protocols, software, articles, data, and other outputs, we must integrate them into existing research workflows and related systems. Thus, we will implement the following use cases:

1. Apply DMPonline “themes” to US funder templates: DCC defines 28 themes corresponding with UK funder questions (e.g., data volume, metadata). A refined list of themes can help standardize and structure DMPs to facilitate the mining of relevant text.⁴ By adding themes to DMPRoadmap, we can test the potential of basic tagging before exploring a more specific vocabulary.
2. Repository recommender service via re3data.org: This provides researchers with information about metadata standards, etc. at the beginning of a project; push notifications to selected repositories about data in the pipeline; and supports an automated function for data tracking.
3. Assign DOIs to DMPs of record: Use DOIs to get award details and other persistent identifiers (Researcher, Grant, Funder, etc.) back into the DMP over time; push/pull notifications across systems to plan resources, connect outputs, automate reporting and monitoring.
4. API to create a plan: Enables, e.g., Offices of Research to pre-populate a plan with the correct template, guidance, and basic details (e.g. PI, title) to reduce the burden on researchers.

In addition to technical solutions, we must expand our capacity to connect with key stakeholders. This requires dedicated outreach to funders in the decentralized US context, in particular to promote the adoption of themes to enable interoperability. We will also participate in data management fora (e.g., RDA plenaries, IDCC), as well as disciplinary meetings to gather requirements and build consensus. Additional project collaborators include US- and UK-based user groups. In the US, the founding members of the DMPTool maintain monthly calls to discuss user feedback and perform usability testing. The two-year work plan encompasses iterative phases of developing, implementing, and testing with various stakeholder groups using an agile methodology.

Outcomes

We see greater potential for DMPs to boost data discoverability, auditability, and reuse.

DMPRoadmap will achieve these goals by supporting:

- The creation of high quality, machine-actionable DMPs for monitoring compliance, etc.
- US researchers working across institutional, disciplinary, and national borders
- Campus-based data management training initiatives
- Funder adoption of common DMP vocabularies and structures

⁴ Used by CASRAI as a starting point for its DMP Ethics vocabulary.

By repositioning DMPs as key elements in a networked data management ecosystem, DMPRoadmap will extend our reach, keep costs down, and move best practices forward, allowing US researchers and institutions to participate in a truly global open science ecosystem.

Budget

The total cost of the DMP Roadmap project is \$1,487,649, including \$731,989 in IMLS funds and \$755,660 in cost share, allocated as \$914,563 in salaries and wages; \$321,926 in fringe benefits; \$8200 in travel; \$18,171 in supplies, materials, and equipment; and \$224,789 in indirect costs.