

Workshop Report

Machine-actionable data management plans (maDMPs)

Stephanie Simms[‡], Sarah Jones[§], Daniel Mietchen[|], Tomasz Miksa[¶]

[‡] California Digital Library, Oakland, United States of America

[§] Digital Curation Centre, Glasgow, United Kingdom

[|] National Institutes of Health, Bethesda, United States of America

[¶] Technical University Vienna, Vienna, Austria

Corresponding author: Stephanie Simms (stephanie.simms@ucop.edu)

Reviewable v1

Received: 05 Apr 2017 | Published: 05 Apr 2017

Citation: Simms S, Jones S, Mietchen D, Miksa T (2017) Machine-actionable data management plans (maDMPs). Research Ideas and Outcomes 3: e13086. <https://doi.org/10.3897/rio.3.e13086>

Abstract

This report presents outputs of the International Digital Curation Conference 2017 workshop on machine-actionable data management plans. It contains community-generated use cases covering eight broad topics that reflect the needs of various stakeholders. It also articulates a consensus about the need for a common standard for machine-actionable data management plans to enable future work in this area.

Keywords

interoperability; persistent identifiers; research data management; research workflows; data discovery; research infrastructure; research policy; research funding; research publishing

Date and place

The workshop took place on February 20 in Edinburgh, UK, as part of the 12th International Digital Curation Conference (February 20-23, 2017). The workshop materials

are available via <http://www.dcc.ac.uk/events/workshops/postcard-future-tools-and-services-perfect-dmp-world>.

Aims of the workshop

With growing interest in active, actionable data management plans, the Digital Curation Centre (DCC) and University of California Curation Center (UC3) at the California Digital Library took the opportunity of the IDCC17 conference to convene a workshop on the topic. The aim of the workshop was to understand different stakeholder requirements and bring together a diverse international group to develop specific use cases for machine-actionable data management plans. The workshop participants included 47 people from 16 different countries, representing funders, developers, librarians, service providers, and the research community. In practical, brainstorming exercises, the groups discussed use cases focused on interoperability with research systems, leveraging persistent identifiers, evaluation and monitoring, and repository and institutional perspectives, and prioritized future work. The DCC and UC3 will use the workshop outputs to implement and pilot the use cases in the DMPRoadmap platform.

Background

Data management plans (DMPs) are becoming commonplace across the globe as a result of funders requiring them with grant proposals, but they are not being employed in ways that truly support the research enterprise. The current manifestation of a DMP—a static document often created before a project begins—only contributes to the perception that they are an annoying administrative exercise. What they really are—or at least should be—is an integral part of research practice, since today most research across all disciplines involves data, code, and other digital components.

Conversations about the need for machine-actionable^{*1} DMPs (also referenced as “active,” “dynamic,” or “machine-readable” DMPs) have been brewing for a few years. We still need a human-readable narrative, but there is now widespread recognition that, underneath, the DMP could have more thematic, machine-actionable richness with added value for all stakeholders: researchers, funders, repository managers, research administrators, data librarians, etc. As purveyors of DMP services, the DCC and UC3, amongst others, are taking action to reimagine DMPs in this context. One goal is to enhance their own service offerings, but in order to be successful, this must be a collaborative and community-driven effort with global applications, as research itself is global. The larger goal is to improve the experience for all involved by exchanging information across research tools and systems and embedding DMPs in existing workflows. We know that better data management is possible and think that better DMP infrastructure that serves as an educational platform and hooks into other systems is part of the solution.

To advance the idea of machine-actionable DMPs (maDMPs), we conducted a landscape survey of existing tools and standards and began [presenting](#) on the topic at international events in 2016. We also identified the Research Data Alliance (RDA) [Active DMPs Interest Group](#) and [FORCE11 FAIR DMP group](#) as ready-made fora for bringing everyone together to determine future directions. In addition to participating in these groups, we began hosting maDMP events, with an initial workshop at the International Digital Curation Conference (IDCC) in February 2017.

This report represents the outputs of the [IDCC workshop](#) and synthesizes our information and idea-gathering work to date. For the workshop, we convened 47 participants from 16 countries representing funders, educational institutions, data service providers, and the research community. Before the workshop, we asked participants to reflect on three questions to seed discussion:

- Why are you motivated / excited / required to work on data management / DMPs?
- What are your pain points?
- What do you hope to get out of this workshop?

The responses revealed several areas of common interest, including the perception of the DMP as a hub or connector for different services, the potential to use the DMP as an advocacy and training tool to support researchers, and the desire to share the information in DMPs dynamically with a variety of research stakeholders and information systems. Full responses are available on Zenodo (Simms and Jones 2017).

Key outcomes and discussions: maDMP use cases

We introduced eight broadly defined topics for the IDCC workshop and asked participants to vote on them. The topics are ranked below according to the interests of workshop participants who organized themselves into groups to develop and prioritize use cases for the top four, while another small group undertook evaluation and monitoring. The topics are all interconnected, and the use cases reflect this; e.g., use cases involving ORCID IDs crosscut multiple groups. We summarized the use cases for each topic, categorizing some of them under the topics that were not explicitly covered during the workshop but where they naturally fall.

1. Interoperability with other research systems
2. Leveraging persistent identifiers (PIDs)
3. Institutional use cases
4. Repository use cases
5. Data discovery and reuse
6. Evaluation and monitoring
7. Disciplinary tailoring and recommender systems
8. Publishing DMPs

Interoperability use cases

The central theme of the workshop was interoperability and exchange of information across research systems. Groups considered various systems ranging from Current Research Information Systems (CRIS) to manage project details, funder systems, electronic lab notebooks (ELNs), active storage and repositories, and publisher systems. The need for common standards emerged as a top priority and is the main use case below. Another priority area is some form of integration with funder systems, since funders drive many of the requirements. The ultimate goal is achieving interoperability with a range of systems used by different stakeholders throughout the research lifecycle.

Common standards and protocols: All stakeholders expressed a need for common standards and protocols as a foundation for maDMP use cases to enable information flow between plans and systems in a standardized manner. This can be achieved using a common data model with a core set of elements. The model can be based on a template structure and/or use the [DMPRoadmap themes](#). It can also be extended with existing standards and vocabularies to follow best practices developed in various research communities. The resulting DMPs should be highly customizable, but a common core model would facilitate broad adoption across communities and enable interoperability of information contained in DMPs.

A nice-to-have, but not necessary, next step involves developing a common interface and default implementation in a variety of programming languages to enable a common way of accessing information in maDMPs. As a consequence, all tools and systems involved in processing research data can be extended easily to be able to provide and access information to/from a DMP. For example, a workflow engine can add provenance information to the maDMP, a file format characterization tool can supplement it with identified file formats, and a repository system can automatically pick suitable content types for submission and later automatically identify applicable preservation strategies.

Creating a common interface will increase the interoperability between systems and enable continuous testing of availability of systems and referenced resources. It will also enable validation of information provided, for example, by checking whether a provided DOI links to an existing dataset, if hashes of files match to their provenance traces, or whether a license was specified.

Furthermore, it will improve interoperability between repositories, especially in cases where a project generates different kinds of data that may be deposited in multiple repositories. maDMPs would maintain the links to all these individual datasets, thereby preserving the context in which the results were produced. Currently, such datasets are connected through associations with a dedicated publication, but this approach does not work well across multiple publications from a project, or if there are no dedicated publications, nor if outputs other than data are shared (e.g., software), or if some of the data ends up in supplementary files to publications.

Additional requirements in this area include that maDMPs:

- Must make use of existing vocabularies (e.g., Crossref Funder Registry) and ontologies whenever possible to enable machine actionability;
- Must employ common exchange protocols (e.g., JSON), including lightweight protocols for big data operations;
- Must be open to support new data types, models, and descriptions;
- Must link to data and identifiable entities such as people, repositories, and licenses, thus enabling validation and scalability;
- Should be available in a format that can be rendered for human use;
- Should accommodate versioning to support actively updated DMPs.

Funder integration: Another high-priority, general consideration is the need to integrate (on some level) with funder systems. All acknowledged the barriers to direct integrations and inability of most funders to mandate the use of specific tools. At the same time, it is important to note that in most contexts, funders drive the demand for DMPs and shape their content. Some means of interoperation between funders and other stakeholders (e.g., Offices of Research, data repositories) would facilitate grant submission, monitoring, and reporting. It would also help institutions and other service providers to stay up to date with funder requirements for DMPs in order to maintain templates and offer appropriate support for researchers. This could help funders demonstrate that DMP quality and compliance have an impact on funding success, which would in turn contribute to improving the quality of DMPs and data management practices. Funder-managed templates that could be shared across systems would be (incredibly) nice to have but not necessary. APIs are one potential mechanism for achieving funder integration.

Leveraging PIDs

All participants view PIDs as a key ingredient in the transition to maDMPs because they would enable information to be passed across existing workflows and systems to plan resources, connect outputs, and automate reporting and monitoring. However, they noted that PID education, for researchers and other stakeholders, is a prerequisite for maximizing their potential value. This is not an insurmountable hurdle, as many researchers already recognize some PIDs, e.g., DOIs in the context of article citations, or field-specific ones like Genbank accession numbers. However, basic data literacy training should include a primer on PIDs: what they are, advice on how to use them appropriately, and why they are important/useful.

Assertions: Employing PIDs in DMPs would allow stakeholders to track assertions about people, organizations/institutions, funders, repositories, and the grants, research resources used, and research outputs attributed to a person (e.g., using ORCID IDs, Funder IDs, Grant IDs, Org IDs, Repository IDs, DOIs for articles and datasets, etc.). Selected PIDs need to interoperate with internal identifiers (e.g., ORCID IDs and staff IDs) and be associated with openly available (CC0) metadata. These assertions would also enable DMPs to remain useful throughout the research process, converting them into a dynamic

inventory of research activities that can trigger actions at the appropriate moment(s) and help automate administrative processes such as reporting (see below).

PIDs enable automated associations that support reproducibility, data discovery and reuse, tracking usage and impact of research outputs for professional advancement, infrastructure funding, etc. Specific examples of PIDs that contribute to reproducibility include [Research Resource Identifiers](#) (RRIDs, used in biomedical research), and identifiers for scientific protocols, biological species, galaxies, and works in catalogues of prolific artists.

Notifications and reporting: Participants identified an extensive list of use cases that involve using PIDs to trigger notifications and automate reporting activities. These actions would alleviate administrative burdens on researchers, funders, and others. They could also improve data management practices by addressing needs and issues during the active phases of a research project. Examples of notifications and/or actions include, in no particular order:

1. Use PIDs to prepopulate sections of a DMP for which information is already available elsewhere (e.g., identifiers about the institutions, funders, people, infrastructure, and resources involved). The respective resources could then be notified about reuse;
2. Notify a repository or other infrastructure provider (e.g., supercomputing or sequencing center, ethical review board) when named in a DMP. Include key information, e.g., volume of data, file formats, licensing, and expected timeline;
3. Derive a description/identity of objects that do not yet exist; machine-sourced/generated metadata (e.g., about a dataset that will be generated and deposited in the future);
4. Notify a funder and/or institution when a dataset is deposited in a named repository and relay metadata and any associated IDs;
5. Notify authorities (institutional or governmental) when legal or policy requirements for data management (e.g., cases of reportable diseases) have been met;
6. Notify a research project when legal or policy requirements referred to in their DMP have been updated;
7. Notify a researcher when a repository accepts preservation responsibility for a dataset;
8. Pass information about grants, projects, and/or research outputs across profile systems to alleviate the need for manual entry. Automatically generate a CV or Biosketch;
9. Notify a research project when a new release of a software library they are using is available;
10. Identify publication of data (e.g., associated with a journal article or stand-alone data publication);
11. DMP of project B listing project A as a dependency allows project A to track reuse;
12. DMP of project A notifying DMP of project B that project A now has additional datasets available of the kind that project B has started to reuse;
13. Aggregation of DMPs for active, past, or upcoming research projects at the level of institutions, funders, repositories, authorities, instructors, etc. or even topics (e.g.,

public health emergencies like the Zika outbreak) for reporting, mining, teaching, and planning of future research and infrastructure.

Institutional use cases

Institutions (especially universities) are significant stakeholders in the RDM landscape and often have data management policies and/or DMP requirements of their own. Many participants brought this perspective in their roles as university administrators, data librarians, and technologists. They noted myriad challenges related to connecting people, resources, systems, and policies within an institution, as well as providing training and outreach services. Capacity planning was another high-priority application of maDMPs. Next steps for DCC and UC3 include modeling the flow of information within some pilot institutions to understand what can be passed between DMPs and existing systems (i.e., Offices of Research, library, IR, faculty profile systems, etc.) and test the use cases below.

Connect researchers with institutional resources/Capacity planning: Participants identified multiple possibilities for this dual-purpose application of DMPs. At most institutions, especially large research institutions, resources and responsibilities for digital research and policy tend to be distributed widely across IT and computing centers, libraries, Offices of Research, human resources, academic departments, etc. Recent efforts to do DMP consultations and, in some cases, build RDM programs have laid the foundation for connecting these dots, but most rely on building personal relationships within an institution and are only just beginning to connect with researchers. maDMPs present an opportunity to share information about resources within an institution more efficiently as well as to link researchers to these resources. They could be used to achieve two complementary aims: first, researchers could connect with available data management tools and services throughout the research lifecycle. Second, maDMPs would enable institutions to identify current and plan for future resource needs. Specific use cases within this area include:

- Help researchers choose appropriate tools: e.g., high performance computing (HPC), ELNs, secure storage, file transfer services;
- Connect researchers with ethics review, auditing, reporting, and other institutional workflows pertinent to the research described in a DMP;
- Connect researchers with training opportunities and consultation services; flag incomplete answers and offer training/advice;
- Map DMPs to domain-specific workflows to offer tailored templates and guidance;
- Connect researchers with IT services: get updates on data needs as a project progresses; budgeting for RDM; forecast storage and preservation needs, other infrastructure planning;
- Integrate DMPs with active data storage (both local and cloud), computation, and ELNs to inform storage allocation/purchase and enable immediate access to data;
- Integrate DMPs with access management for data and associated research outputs.

Institutional policies and governance: A subset of the above involves using maDMPs to inform researchers about institutional data and intellectual property (IP) policies, which can overlap with funder policies and present various challenges. It is imperative that researchers and institutions understand the policy landscape during the planning stages of a project in order to avoid problems down the line. For instance, sensitive data needs can be complex and arise in the context of all kinds of biomedical, environmental, and social science research. Data security and access control, compliance, and reporting could all be monitored by the appropriate stakeholders, within and beyond the institution, using maDMPs. Institutions could track compliance with local policies related to data retention periods and open access publications. Offices of Research could check compliance with funder policies via up-to-date, post-award DMPs (precedents are already in place for Horizon2020 in the EU, and NERC and EPSRC in the UK).

An integral component of this and other use cases is the need to make DMPs an open resource. Ideally, they should be publicly available in line with open, transparent, reproducible research objectives—open DMPs would demonstrate good research practice and facilitate data discovery and reuse. We acknowledge that the culture change toward greater openness is slow and uneven across the academy, and so at the very least DMPs should be shared within an institutional setting². This would provide institutions with a high-level picture of data needs and compliance with various policies.

Training, networking, and publishing: Yet another area of overlap with open maDMPs involves use cases associated with training, networking, and sharing/publishing. Sharing within or beyond individual institutions would facilitate outreach efforts between service providers and researchers, especially in the realm of RDM consulting and training. Service providers could evaluate RDM maturity within a department/school/faculty and tailor training for specific needs. Open maDMPs might have networking effects by alerting researchers to similar projects within their institution and/or promoting interdisciplinary work (e.g., environmental research). DMPs should also be made open to enhance the visibility of “good” examples for others to follow and acknowledge these efforts; these activities fall under sharing. The next step in promoting greater openness, which involves incentivizing the creation and maintenance of good DMPs, is publication (see Publishing DMPs below).

Repository use cases

Data repositories play a key role in the long-term management of data, ensuring that it is preserved and remains accessible. The majority of DMP requirements ask researchers to identify an intended data repository, recognizing that repositories are more suitable than other commonly used media (e.g., hard drives, project websites). It is rare, however, for repositories to play an active role in the data management planning process. One instructive exception is the Natural Environment Research Council (NERC) in the UK, which has designated data centers. NERC-funded researchers identify which one they will deposit in at the grant application stage and then [collaboratively develop](#) the DMP with the data center post award. Workshop participants focused on two-way information exchange

between repositories and DMPs and recommender systems to alert researchers to appropriate services.

Repository recommender: An essential component of every DMP is the plan for preserving data and other outputs, which in most cases involves selecting a repository. DataCite's [re3data](#) service is an excellent resource, however, the list of results can be overwhelming and difficult to navigate. Participants pitched ideas about filtering the list to recommend repositories based on the researcher's discipline, country, data type, or specific needs (e.g., generating PIDs to suit H2020 requirements). Additional approaches to a repository recommender service include:

- Connecting researchers with community-curated lists such as [biosharing.org](#) (e.g., via an API);
- Mining extant DMPs for a specific funder and/or discipline and making recommendations based on the top-cited repositories;
- Once the data type is identified in a DMP, information could be provided on the top 10 repositories where data of this kind has been deposited in the past. Additional filters could be offered to highlight trusted digital repositories and those that assign PIDs (as in the re3data catalogue);
- It could be useful to filter by repositories used by the researcher's own institution.

Any recommender services should have different functionality for data generators and reusers, instructors, tool developers, institutions, funders, and others.

Begin the archival/preservation process: When a researcher names a repository in a DMP, maDMPs could alert repositories to data in the pipeline. This would allow repository managers to initiate discussions with researchers early on. It would also facilitate capacity planning and help repositories monitor changing requirements from users. Key information such as data types, volumes, and standards could be extracted from the DMP and shared with relevant stakeholders. Information from the DMP could be used to facilitate the deposit process (e.g., by prepopulating user details and basic metadata in the repository upload form), and once data is deposited, the DOI or other identifier could be sent back to automatically update the DMP, assisting with evaluation and monitoring use cases. The Publishing DMPs use cases below include ideas about depositing/preserving DMPs along with other research outputs.

Data discovery and reuse

Of the nearly 40,000 DMPs that have been written so far with DMPonline and the DMPTool, very few are available in ways that would help people, machines, or institutions find out about the research and data they describe. In the following, we consider approaches to using maDMPs for discoverability and reuse; the approaches also rely on making DMPs public, versioned, and aware of PIDs. These four dimensions each confer benefits for data discovery in their own right, but each of the possible combinations—which can also be explored independently—increases the benefits substantially.

PIDs: Including PIDs in DMPs allows for some basic integration with other PID-aware environments, just like recognizing a specific PID of a particular format (e.g., a GenBank [accession number and version](#)) on a printed page can allow a human familiar with this kind of PID to manually access online information associated with it (e.g., a gene sequence) if they know where to look.

To find out about updates of the record associated with a PID mentioned in a DMP is hard for a DMP available only on paper or in some unstructured format like PDF, although the manual process outlined above could be repeated at additional points in time, as is the case with the [CrossMark](#) button.

Versioning: DMPs are planning tools, and like most plans, they need to be adapted on an ongoing basis. Keeping track of modifications throughout the lifecycle of a research project is essential in order to keep the goals of the original plan in sight, to identify new ones as necessary and to notify relevant stakeholders (e.g., the repositories named in a DMP) of changes. Therefore, DMPs should be properly versioned, as discussed below under Resource type and DMP publishing options.

Machine actionability: Automating lookup processes has clear benefits but requires that instructions on the recognition of the PID and its versioning as well as the where and the how of accessing the associated online information are available in ways that machines can act on. If that is the case, human users (as well as automated tools) can be notified of new versions of a PID-associated record and retrieve it automatically. Such automation also supports aggregating information for all PIDs of particular kinds within a DMP or across DMPs, aggregating the DMPs themselves, having them notify each other in case of updates, or mashing them up in other ways, e.g., to check for compliance with applicable policies. Once aggregated or mashed up, classical digital discovery mechanisms like “Related DMPs”—e.g., “DMPs referring to the same GenBank record” or “DMPs authored by the same author PID”—can be used to explore the DMP collection, to retrieve associated information in bulk or to prepopulate a new DMP upon its creation. On that basis, there could also be links to “Related papers”—e.g., papers citing similar software packages, datasets, grants, or publications as a given DMP—and other features.

Publishing: Publishing a single DMP, even after an embargo period, in an unstructured format and without PIDs or versioning, can be valuable, as it allows others to find out about the research described in it (which may or may not be discoverable otherwise, especially if the research is still ongoing) and to engage on that basis. Publishing DMPs at larger scales would allow them to be aggregated within a specific DMP collection and/or across different public DMP collections or even across public collections more generally. Conversely, public DMPs could also become discoverable from outside DMP collections, e.g., by mechanisms like “DMPs referring to this item” in data or literature repositories, or through simple web searches.

Evaluation and monitoring

The current manifestation of DMPs is not well suited to automated compliance checks, but this is a critical need for review processes to scale. When DMPs are not evaluated for quality and a poor DMP is perceived to be of no consequence, policies are quickly undermined. Participants focused on determining checkpoints and making recommendations for more structured DMP content. They also noted that funders as well as reviewers need training in DMP evaluation, and that evaluation rubrics would help everyone assess plans (cf. Whitmire et al. 2016). And once again, open DMPs would support evaluation and monitoring use cases.

Automated compliance checks: Funders, institutions, and repository managers need an automated mechanism to determine whether researchers did what they said they would do in a DMP. This is a fundamental, high-priority use case for maDMPs although stakeholders were careful to point out that a narrow focus on compliance monitoring and enforcement risks increasing frustration levels among researchers. A thoughtful approach to compliance should therefore consider incentives and rewards, e.g., with recognition for tenure and promotion, as well as potential side effects of introducing measures in this space (Edwards and Roy 2017).

Quality/validation checks: Funders, institutions, repository managers, etc. also need an automated approach to validating whether stated plans regarding data management are appropriate. Wherever possible, maDMPs should offer closed questions (e.g., list repositories and metadata standards for a particular discipline, acceptable file formats, etc.). If stated plans are not appropriate, a program officer and/or service provider should receive a push notification at which point they can get in touch with the author of the DMP.

Disciplinary tailoring

This is among the most challenging issues, yet it ranks as a high priority. All stakeholders emphasized the need to offer relevant guidance at appropriate points throughout the research lifecycle rather than the current approach of asking broad, unstructured questions at the planning stage when few details are known for sure and presenting generic RDM best practices as guidance. Data management strategies can vary dramatically between and even within disciplines, and so the wisdom has been to leave it to researchers and/or research communities to determine their own standards and best practices. Only a handful of communities that benefit from standardization have cohered around common practices (e.g., genomics research) and some others are beginning to follow suit (e.g., fMRI brain imaging). Although the culture change is slow and precisely targeted guidance may never be available for all disciplines, there are opportunities to hook into some existing systems and databases. There was consensus during the workshop about the urgent need to experiment with serving up more helpful guidance and improve the DMP experience for everyone.

To this end, the DMPRoadmap project is developing some pilots. Repository recommender services (described above) are one obvious area for experimentation. RDA funding is

available to test an integration of the RDA metadata standards catalogue. Tagging or filtering by community/disciplinary affiliation might facilitate these efforts.

Biosharing.org is a disciplinary partner with an API that can be used to connect researchers to a curated database of resources for the biomedical and environmental research communities as they are writing a plan. The [Biological and Chemical Oceanographic Data Management Office](#) (BCO-DMO) repository that services researchers funded through a variety of NSF programs represents another pilot for exploring how to tailor guidance and structure plans in the DMPRoadmap platform.

Additional opportunities should be identified and drawn into the maDMP discussion. This is also an important consideration when developing common standards for DMPs; i.e., we need an expressive format with lots of optional fields to accommodate different disciplines. [ELIXIR](#), a European life sciences infrastructure initiative, is developing a [Data Stewardship Wizard](#) that provides researchers with a decision-tree style checklist. The answers chosen prompt different pathways through the questions or allow researchers to dig deeper on topics of interest. A working group of Science Europe is developing [DMP Protocols](#), which present opportunities for maDMPs. The protocols would present model responses for different domains and the range of viable options. Serving up the protocols via tools like DMPonline and DMPTool are another approach to offering more tailored guidance.

Publishing DMPs

There are growing trends towards both informal sharing and formal publication of DMPs. Opening DMPs brings many benefits and is something we actively encourage. DMPs can be aggregated and mined to identify trends or aid discoverability and reuse of data. They also serve as a useful training resource; many institutions refer to “good” examples to help other researchers get started.

Resource type and DMP publishing options: Numerous stakeholders expressed a need to define a resource type for DMPs to distinguish them from datasets and other research outputs (e.g., when deposited together in repositories). This should be supported by DataCite and other common metadata schemas.

There is a strong desire to assign DOIs to DMPs in order to link DMPs with related outputs of a project such as publications, datasets, and software (see Leveraging PIDs). This would aid reproducibility, as the context of the research and all the outputs could be shared together. It is necessary to think through the implications of assigning DOIs to DMPs, however, especially if we aim to support a lifecycle approach with dynamic updating. At a minimum, there should be two versions of record for a DMP: one submitted with a grant proposal and a second one at the grant closeout/reporting stage. Another way to think about this could be in the context of software versioning: every commit has an ID on GitHub or elsewhere, but a DOI only gets assigned to the subset of versions submitted to a repository or publisher. This also dovetails with ongoing efforts in the RDA and elsewhere to define best practices for citing dynamic data.

Various entities are testing the idea of DMPs as a publishable unit to promote greater openness and enhance their value to researchers. For instance, DMPs form [part](#) of the European Commission's Open Public Review pilot, in which deliverables of a small set of Horizon 2020-funded projects are being posted for public review while the projects themselves are ongoing.

DMPRoadmap already supports sharing DMPs within an institution or openly in a [public list](#), but this and other platforms could introduce a more formalized concept of publishing as part of the DMP workflow. This could be done by adding options to publish DMPs in journals (e.g., [RIO Journal](#) and [BMC Research Notes](#)) or deposit them in repositories (e.g., Zenodo, Dataverse, Figshare), alongside the standard export feature.

Conclusions and next steps

This document presents a list of community-generated maDMP use cases. It also articulates a consensus about the need for a common standard for maDMPs to enable future work in this area. At the RDA 9th Plenary meeting in Barcelona during the [Active DMPs IG session](#) (6 April 2017), we propose establishing a working group to develop standards for DMPs.

Recommendations regarding DMPs will also be made via the European Commission's [FAIR data expert group](#), specifically with regard to the structure of the Horizon 2020 DMP template to automate monitoring of deposits in repositories and balance a generic approach to DMPs with the need for disciplinary tailoring.

The DCC and UC3 will continue to pursue international collaborations related to DMPRoadmap through pilot projects. As part of an iterative process for developing, implementing, testing, and refining these use cases, they will model domain-specific and institutional pilot projects to determine what information can realistically move between stakeholders, systems, and research workflows. There is some existing funding to support a subset of this work; the organizations are actively seeking additional sources of funding to carry the project forward.

Existing funding includes an RDA Europe collaboration award to support embedding the Metadata Standards Directory and biosharing.org resources into the DMPRoadmap platform. The biosharing integration will support the biomedical research community and taps into larger initiatives such as ELIXIR. OpenAIRE funding will support an export to Zenodo feature, and EUDAT will contribute to further API development.

Another disciplinary pilot project involves partnering with the NSF-funded BCO-DMO to use its [GEOTRACES](#) corpus, a long-term, international study of marine biogeochemistry. Purdue University and the University of California, San Diego will serve as institutional pilots to model the flow of information across Offices of Research, libraries, repositories, and faculty profile systems. In addition to technical solutions, these projects will expand our

capacity to connect with key stakeholders, with particular emphasis on addressing the needs and practices of researchers and funders.

In addition to the upcoming RDA meeting, we will circulate these use cases with the FORCE11 FAIR DMPs group and identify additional opportunities to connect with international maDMP initiatives as well as working groups in related pursuits (e.g., controlled vocabularies, decision trees for data management). We will continue to collect feedback on these use cases and facilitate discussions about how to prioritize our next steps as a community, ideally through use-driven experimentation in multiple directions.

Workshop participants

The list of participants is given in Table 1.

Table 1. List of workshop participants.		
Name	Affiliation	Country
Alex Ball	University of Bath	United Kingdom
Amber Leahey	Scholars Portal	Canada
Andy Riddick	British Geological Survey	United Kingdom
Benjamin Faure	INIST-CNRS	France
Bev Jones	University of Lincoln	United Kingdom
Brian Riley	California Digital Library	United States of America
Chuck Humphrey	Portage Network	Canada
Daniel Spichtinger	European Commission	Belgium
Daniella Lowenberg	California Digital Library	United States of America
David McElroy	Birkbeck- University of London	United Kingdom
Falco Hüser	Technical University of Denmark	Denmark
Fernando Aguilar Gómez	IFCA	Spain
Fernando Rios	Johns Hopkins University	United States of America
Gene Melzack	The University of Sydney	Australia
Heila Pienaar	University of Pretoria	South Africa
James Wilson	UCL	United Kingdom
Jari Friman	Helsinki University Library	Finland
Jez Cope	University of Sheffield	United Kingdom
Jimmy Angelakos	EDINA, University of Edinburgh	United Kingdom
John Chodacki	California Digital Library	United States of America
Joshua Finnell	Los Alamos National Laboratory	United States of America

Kevin Ashley	Digital Curation Centre	United Kingdom
Lisa Johnston	University of Minnesota	United States of America
Margreet Bloemers	ZonMw	Netherlands
Mari Elisa "Mek" Kuusniemi	University of Helsinki	Finland
Marie-Christine Jacquemot	INIST-CNRS	France
Marisa Perez	Universidad Autónoma de Madrid	Spain
Marisa Strong	California Digital Library	United States of America
Marta Teperek	University of Cambridge	United Kingdom
Michael Heeremans	University of Oslo	Norway
Michael Moosberger	Dalhousie University	Canada
Myriam Mertens	Ghent University	Belgium
Patrick McCann	University of St Andrews	United Kingdom
Pedro Principe	University of Minho	Portugal
Peter McQuilton	University of Oxford	United Kingdom
Peter Neish	The University of Melbourne	Australia
Poppy Townsend	Centre for Environmental Data Analysis	United Kingdom
Rachael Kotarski	British Library	United Kingdom
Ray Carrick	EDINA, University of Edinburgh	United Kingdom
Rob Hooft	DTL	Netherlands
Roman Ujbanyai	VUB	Slovakia
Sarah Jones	Digital Curation Centre	United Kingdom
Stephanie Simms	California Digital Library	United States of America
Thilo Paul-Stueve	Kiel University	Germany
Tomasz Miksa	TU Wien	Austria
Weiwei Shi	University of Alberta Libraries	Canada
William Michener	University of New Mexico	United States of America

Acknowledgements

We acknowledge and appreciate the contributions of those who participated in the workshop. Many additional individuals and organizations have helped shape the ideas and use cases herein. This project is supported by an RDA/US Data Share fellowship awarded to S. Simms, sponsored through a grant from the Alfred P. Sloan Foundation #G-2014-13746.

References

- Edwards M, Roy S (2017) Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition. *Environmental Engineering Science* 34 (1): 51-61. <https://doi.org/10.1089/ees.2016.0223>
- Simms S, Jones S (2017) Machine-actionable data management plans (maDMPs): IDCC17 workshop materials. Zenodo <https://doi.org/10.5281/zenodo.475054>
- Whitmire A, Carlson J, Westra B, Hswe P, Parham S (2016) The DART Project: using data management plans as a research tool. *Open Science Framework* <https://doi.org/10.17605/OSF.IO/6ZUDN>

Endnotes

- *1 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”, as per <http://www.ddialliance.org/taxonomy/term/198>.
- *2 Anecdotal evidence suggests that sharing within institutions is feasible: DMPTool administrators report a ready willingness among researchers to share DMPs within their institution, although not always with a wider public audience, and DMPonline users have requested an institutional sharing functionality.