

Exposing DMPs: Use cases, workflows and guidelines

Case statement for community review

July 2017

Contributors

David Carr - The Wellcome Trust
John Chodacki, California Digital Library
John Faundeen, Earth Resources Observation Center, USGS
Bev Jones University of Lincoln
Natalie Meyers, Centre for Open Science/ University of Notre Dame
Paul Millar, DESY
Fiona Murphy, MMC Ltd
Kathryn Unsworth, Australian National Data Service
Angus Whyte, Digital Curation Centre (Editor)
Elena Zudilova-Seinstra, Elsevier

Working Group Charter

A variety of stakeholders are showing growing interest in exposing data management plans (*) to other actors (human/machine) in the research lifecycle, beyond their creator and the funder or institution that mandates their production. Interested stakeholders include researchers themselves, funders, institutions, and a variety of service providers and community organisations including repositories, institutions, journals, publishers, and providers of tools for writing and maintaining plans.

Implementation and adoption is currently hampered by two problems:

- A lack of standards for expression and interchange of DMPs
- Insufficient understanding of the needs of users and the benefits and risks of different modes of action

This proposed working group will address both of these issues; the issue of a standardised form of expression for DMPs is the concern of the proposed DMP Common Standards Working Group. The group's output will include a reference model and alternative strategies for exposing plans, to best serve community interests in meeting FAIR principles,¹ based on shared experience of 'early adopters' in test implementations. It will be supported by work to gauge user needs and motivations for exposing DMPs as well as perceived risks and disbenefits. Note * our main focus is on Data Management Plans (DMPs) but we will seek examples of Software Management Plans (SMPs) where relevant to exposure use cases of interest to the Active DMP Interest Group.

Value Proposition

The key beneficiaries of the WG outcomes will be stakeholders with a common interest in using Data or Software Management Plans as instruments for demonstrating that research products have been

¹ Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., ... others. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4792175/>

managed according to research community standards and generic principles (e.g. that the research products should be FAIR), and that recognition is given for doing so.

There is potential value in exposing plans for a variety of stakeholders involved in their production and consumption. These include researchers themselves, funders, institutions, and a variety of service providers and community organisations including repositories, institutions, journals, publishers, and providers of tools to help write and maintain plans. The WG will provide a *Use Cases Catalogue* to describe implementation scenarios and articulate their benefits to researchers and other stakeholders, with case studies of how those benefits have been realised. Through consultation with users of well-established planning tools (DMPTool, DMPonline), the Use Cases Catalogue will also identify the degree of acceptance among researchers for the levels of exposure/publication each use case entails, barriers to realising the benefits, and any concerns about undesirable impacts.

Generalising from the scenarios and examples contained in the Use Cases Catalogue, the WG will produce a Reference Model to document generic components and workflows for exposing plans (and metadata about them), and offer recommendations for further action by each of the relevant stakeholder groups. By gaining endorsements for the Reference Model from relevant stakeholders for each use case we will provide a community endorsed approach to using plans to share demonstrable advancement in data sharing practice.

As a starting point we identify below some of the potential benefits for each stakeholder group:

For *researchers* a public DMP may be seen as a new way to make their research more visible and facilitate information exchange about best practices in data sharing. It may also evolve to become a research output in its own right; a creditworthy claim that the research team has a novel or interesting approach to data management (e.g. RIOjournal). This is analogous to publishing methods (e.g. [Methods](#)) and other research outputs (e.g. [Wellcome Open Research](#)).

For *researchers, data managers, research software engineers and others involved in delivering research outputs* the plan informs their work, reminding them where, when, and with whom to share those outputs. Exposing plans more widely to project participants when active data management is underway can better inform decision making (e.g. in labs) regarding how to manage the research project users & groups, which storage to write to and when, which data elements to make public, with which licenses, etc. Keeping the research project participants informed about the plan throughout the project's lifecycle will increase compliance, and allow project participants to note variance from any pre-award plan.

As a record of such decision making, an actively updated DMP or SMP can improve the coordination and in turn reproducibility of research by others. Participants who may not have authored the plan, or only contribute to part of the project will especially benefit from easier access to plans, as will those who may have joined the project part-way to completion. Finally, replication studies will be better informed if the original plan is shared, and this plan has been actively updated to include a record of any variance in context, along with the data or code underlying other published outputs. Integration of DMP/SMP authoring tools with active data management platforms would help researchers to keep plans in step with the actions they refer to, with the potential for automation. For example Los Alamos National Laboratory is prototyping integration between DMPTool and Open Science Framework.²

² Finnell, J., Klein, M., & Cain, B. J. (2017). Nucleus: A Pilot Project. *ArXiv:1705.07862 [Cs]*. Retrieved from <http://arxiv.org/abs/1705.07862> .

For *institutions* when research teams share DMP content through Research Information Management or CRIS systems, the information may be used by Research Offices and other institutional support services to assess demand for RDM services, such as storage, preparation for deposit, quality control mechanisms, or training. Some institutions have explored internal exchange of DMPs and related metadata between DMP support tools (DMPTool, DMPonline) and other institutional systems such as Current Research Information Systems and grants databases.³ More commonly, institutions have encouraged sharing of example DMPs, to promote good practice in writing DMPs across the institution, and help to meet demand for support in meeting funders' expectations. Publishing DMPs from successful bids may also demonstrate the institution's role in meeting funder expectations. We can also use the plan, at data storage or archiving, to determine if the metadata provided, and storage obtained or requested, meet the requirements originally set out.

For *data repositories* DMP details shared in advance of data deposition help the repository's own planning, and offer context and provenance information that may otherwise be difficult to retrace. For example the UK Natural and Environmental Research Council requires researchers to submit an outline DMP and, if their proposal is funded, provide a 'full' DMP in collaboration with the datacentre they intend to submit data to.⁴ This illustrates that DMP publication may help the repository cross-link between the research outputs it holds, and provide information that increases the potential for reuse.

For *funders* public DMPs may aid community scrutiny and knowledge-sharing, supplementing the funders' internal review and monitoring. A recent Open Review pilot by the European Commission illustrates this.⁵ Longer-term impacts could include more efficient use of funding towards FAIR outputs. For most funding bodies DMP are a relatively recent policy development, and so far there is little published evidence of such impacts. Recently however OpenAire has carried out a survey to gather feedback on the European Commission's approach to DMPs and the FAIR DMP template in particular.⁶

For *journals*, a public DMP offers context information to aid peer review of any data underlying a submitted article and therefore strengthen claims of transparency/reproducibility. In addition, DMPs may be publishable as a description of data management aspects of interest to journal audiences. For example DMPs are published by RIOjournal and BMC Research Notes, and International Journal of Digital Curation is planning to do so.

For *publishers and funders*, a published plan may also offer a similar role to a data article, providing a project-level view of data or software produced in research, and a record of how production has been managed in accordance with policy expectations. A plan that records relationships between data, code and other outputs can provide the funder and publisher with a means to more efficiently and effectively link up these research products, helping funders to track their investments and publishers to maintain the research record. For example the US Interdisciplinary Earth Data Alliance (IEDA) provides tools to link DMPs with data submissions, via information on the relevant National Science Foundation award.⁷
⁸ There is also strong interest from the Wellcome Trust in the potential to link plans to the resulting outputs, and in exposing plans to help enhance discoverability, as well as for the basis of compliance

³ Jones S. (2017, June 8) 'On the right track(s) - DCC release draws nigh' blog post:

<http://www.dcc.ac.uk/blog/right-tracks-dcc-release-draws-nigh>

⁴ NERC (n.d.) 'Data Management Planning': <http://www.nerc.ac.uk/research/sites/data/dmp/>

⁵ European Commission: <https://ec.europa.eu/futurium/en/content/d61-data-management-plan>

⁶ OpenAire: <https://www.openaire.eu/openaire-survey-dmp>

⁷ IEDA <http://www.iedadata.org/compliance/dmp/FAQ>

⁸ Reilly, M. & Dryden, A.R., (2013). Building an Online Data Management Plan Tool. Journal of Librarianship and Scholarly Communication. 1(3), p.eP1066. DOI: <http://doi.org/10.7710/2162-3309.1066>

checking. Wellcome has also recently moved from requesting DMPs to asking for more holistic outputs management plans (covering data, software, materials and IP).⁹

For *Research Data Management* service providers, public machine-readable plans can be used as a basis for added-value services that link research to outputs throughout the research lifecycle, and help maintain the integrity of the research record. Such services could include, for example, accumulating information about the actual data location, its characteristics and provenance for further automatic tracking, including data reuse or disposal. There are also opportunities to make plans interactive by linking DMP support tools to other tools and platforms offering guidance ‘wizards’¹⁰ and training layers to ensure that relevant advice and services are brought to the attention of researchers from the beginning of their project.

Finally, as DMPs are “becoming the main tool to enforce policies on research data management”,¹¹ collections of publicly accessible plans may become a research resource in their own right. This should for example enable studies of the impact of policy changes on data management practice, and help track disciplinary variations in practice that can inform research communities to establish ‘RDM protocols’ that guide individual researchers in a given field.

Engagement with existing work in the area

The case statement builds on related work on machine-actionable DMPs, and on data publishing workflows, and complements parallel efforts in the RDA and elsewhere. The statement draws on use cases identified in the IDCC17 workshop report by Simms et al.¹² That report identifies 8 main use cases, each of which relies on some element of DMP exposure, and is based on priorities identified by participants in this workshop and previous RDA discussions. We will further define and expand on those use cases. Where they involve exposing DMP content about the planned use of storage, we will adopt the standard vocabulary that the Storage Service Definitions working group are developing, where appropriate, to facilitate meaningful comparison of storage requirements across different DMPs.

The value proposition set out in the previous section rests on our ability to more fully articulate community requirements, and to recommend actions that stakeholders can take to help ensure that exposed/published DMPs provide measurable benefits. The proposers are directly or indirectly involved in a number of initiatives that will support that work, including the following:

- Current joint work by the *Digital Curation Centre (DCC) and University of California Curation Center (UC3)* outlined in the workshop report, towards gathering and responding to community requirements in this area through their DMP support tools DMPtool and DMPonline.
- The proposed work by *DMP Common Standards WG* to develop a common data model and access mechanisms for machine-actionable DMPs, which will enable implementation of the use cases we model.
- The proposed work by *QoS-Data LC Definitions (Storage Service Definitions) WG*, which will inform the service recommendation use case, potentially enabling storage requirements documented in a DMP to be automatically matched to storage services.

⁹ Wellcome Trust: <https://wellcome.ac.uk/news/our-new-policy-sharing-research-data-what-it-means-you>

¹⁰ See for example ELIXIR Data Stewardship wizard: <http://dmp.fairdata.solutions/>

¹¹ IG Domain Repositories: <https://www.rd-alliance.org/ig-domain-repositories-rda-9th-plenary-meeting>

¹² 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>

- The Domain Repositories IG, to ensure that use cases reflect the interest in establishing domain-level RDM protocols.
- Previous work by the *WDS/RDA Data Publishing Workflows WG* to produce a Reference Model for Data Publishing¹³, and guidelines for service providers connecting repositories to workflows earlier in the research lifecycle.¹⁴
- Ongoing work of the *RDA/WDS Publishing Data IG*, to ensure workflows for DMP publication connect with work to ensure research outputs are cross-linked.
- Work proposed by the *Data policy standardisation and implementation IG* to harmonise data policy, to ensure any policy recommendation are complementary and avoid duplication of effort.

In addition, we have overlapping membership with:

- The FORCE11 working group on FAIR DMPs which is collating international activity in this area (<https://www.force11.org/group/fairdmp>) and can ensure our work packages are complementary.
- The recently formed Australian Data Management Planning IG and its sub-group on Machine-actionable DMPs.

The work will be informed by results from the recent survey by OpenAire about the Data Management Plan template for the European Commission's Horizon2020 programme, in particular the priorities identified for additional functionality in associated DMP tools.¹⁵

Work Plan

Encouraging early steps have been taken towards providing a spectrum of access modes to a DMP, from partial exposure of plans and metadata through to publication. This case statement has set out some of the potential reasons for exposing DMPs to wider access, interrogation and scrutiny. However, many researchers view DMPs as an administrative burden. While there may be widespread support for the view that DMPs offer researchers an opportunity to manage their data better, and offer various professional groups an opportunity to help them do that, there is little evidence of who gains or is adversely affected by DMP sharing and exposure along the way.

We believe there is widespread community interest in further reference information and guidance on when and how data and software plans should be shared, in order to best serve community interests in meeting FAIR data principles with least effort. At present the guidance is limited to that available from plan authoring tools such as DMPonline and DMPTool. This Working Group aims to fill that gap by delivering the following outputs:

¹³ *Key components of data publishing: using current best practices to develop a reference model for data publishing* <https://zenodo.org/record/56789>.

¹⁴ *Connecting data publication to the research workflow: a preliminary analysis* ([google doc](#)) IDCC17 Edinburgh, 22 February 2017

¹⁵ 'OpenAIRE survey about Horizon2020 template for Data Management Plans' OpenAire News May 22, 2017 <https://www.openaire.eu/openaire-survey-dmp>

Milestones and Deliverables

M1. Use Cases Consultation (Report and dataset– M6)

The only evidence of researchers' support for exposing DMPs is the fact that they do it. Many do so with only minimal exposure, e.g. to institutional support services in order to get advice. A few do it with maximal exposure, using various routes to make their DMP public. Their reasons, concerns and arguments for doing so (or not) have not so far been widely documented, and this consultation aims to correct that.

The use cases consultation will provide an interim report describing feedback elicited on a range of exposure scenarios, from the two main stakeholders in Data Management Planning; data policy-makers and the researchers those policies are directed at. The scenarios (8-10 in number) will range in degree of exposure from partial exposure of metadata to full exposure of plans they are responsible for producing, or interested in accessing.

1. 'Direct "Live" Deposit' Integration of DMP management tools (e.g. DMPonline¹⁶, DMPtool¹⁷, RDM Organizer¹⁸) with active data management and sharing platforms (e.g. Open Science Framework¹⁹), grant administration or Current Research Information Systems. Exposure in this use case may be limited to that required for the research team and direct collaborators to implement the plan and engage with their organisation's professional services.
2. 'Service Recommendation'; Users are offered further guidance on options available that match their research context, funder requirements and/or data sharing policies of the top journals in the field, from a broader range of services (e.g. cloud storage, data analysis) by disclosing elements of their planning at whichever stage in their DMP/project lifecycle they choose. As above, service recommendation might be performed within DMP management tools, or via intermediary applications.
3. 'Direct Deposit' Integration of DMP management tools (e.g. DMPonline, DMPtool) with journal and repository submission workflows, possibly via intermediary applications (e.g. Data2Paper²⁰). Users are offered guidance on options available that match their research context, - these may consist of open source, academic and/or commercial tools and other services - then taken through the submission and acceptance criteria, and can then submit directly to their chosen outlet.
4. 'DMP Paper-1' Submit a Data Management Plan as an article for publication. Similar in concept to a data paper, the DMP paper briefly describes the research context for the DMP, identifies aspects of the data management challenges of interest to the community, and cites any dataset that is available.
5. 'DMP Paper-2' As above, but the DMP itself is first submitted to a public repository and then cited in the DMP Paper along with any relevant dataset, using a citation format similar to that recommended by Force11 for datasets.

¹⁶ DMPonline. <https://dmponline.dcc.ac.uk/>

¹⁷ DMPtool <https://dmptool.org/>

¹⁸ RDM Organizer. <http://rdmorganiser.github.io/en/>

¹⁹ Open Science Framework. <https://osf.io/>

²⁰ Data2Paper. <https://www.data2paper.org/>

The final set of scenarios for the consultation will draw on the use cases identified in the IDCC17 workshop report on MaDMP by Simms et al. That report identifies 8 main use cases, each of which relies on some element of DMP exposure, and is based on priorities identified by participants in this workshop and previous RDA discussions. These will be further elaborated on, and an agreed set of up to 12 use cases will be used in the consultation.

The main instruments for gathering feedback on the use case scenarios will be: -

- 1) Online survey to elicit views on the benefits and risks of DMP exposure, targeting the following groups:
 - a) Researchers who have submitted DMPs to a funder with the aid of DMPtool, DMPonline or other tool for DMP authoring/ advice
 - b) Administrators/ professional support services responsible for providing localized advice on data or software management planning, including in institutions using DMPtool, DMPonline
 - c) Members of the Active DMP Interest Group
- 2) Phone or email interviews to elicit more detailed responses on usage scenarios, with survey respondents and with authors of data or software management plans that have been made public through a journal, repository, website, or DMP authoring tool
- 3) Phone or email interviews with data policy-makers in at least 4 funding bodies and 8 institutions (spanning the US, UK, Europe, and Australasia) that have policies requiring or recommending that researchers produce DMPs
- 4) Group discussion at RDA Plenary 11 or the IDCC18 conference; with stakeholders including the above; plus representatives of academic libraries, publishers, research infrastructures and other groups providing advice or other services to support data and software management planning

Response data from users of DMPTool or DMPonline will be anonymized before sharing if required by respondents, and deposited in Zenodo or another public repository. Interview data will also be deposited. The report will summarise responses to consultation questions relating to each scenario.

D1. Use Case Catalogue and Case Studies (Report/ database M12)

The Use Cases Catalogue and Case Studies will describe our shared analysis of the consultation responses, and provide a case study describing an implementation of each of the 8-10 use case scenarios, where examples can be found. While the consultation responses will describe the range of sentiment, benefits and concerns identified by each stakeholder group involved in the use case, the case studies will illustrate the planning lifecycle and relevant workflows used by adopters of pilots/early implementations. Adopters will also be invited to describe the outcome of their planning in terms of three main dimensions:

1. The extent to which the plan has been updated from an initial statement at the project application stage through to formal reporting and publication of outputs
2. The extent to which the plan has been shared by its primary author (e.g. researcher or data manager) with peer groups, institutional support services, domain-based support services, public repositories, and journals
3. The extent to which the latest version of their plan has altered choices proposed in the initial version

The studies will consider how these dimensions have, from the plan authors' perspective, shaped subsequent choices of platforms and tools to conduct the research, and impacted on their project's ability to meet FAIR data management guidelines.

We will solicit the use cases from members of the Working Group, the Active DMP Interest Group, known contacts and other contacts made through the use cases consultation. The approach taken is based on the co-chairs experience in inviting the participants in a previous RDA/WDS Working Group on Publishing Data Workflows to present a structured walk through of their processes, during RDA plenaries and conference calls.

D2. Reference Model and Recommendations for Exposing Data and Software

Management plans (Report, M18)

A Reference Model will be produced to offer practical guidelines for institutions and service providers. The model will illustrate how exposure helps researchers and others to use Data or Software Management Plans as resources for better management, stewardship, and preservation, rather than as a box-ticking exercise. Specific recommendations or guidelines will address aggregation of metadata from plans to inform capacity planning, submission of plans to publishers, and the use of shared plans in policy-making to evaluate how data policies impact on data management practice.

Drawing on the approach taken by the RDA/WDS Publishing Data Workflows WG, which produced a Reference Model for Data Publishing, the Reference Model will describe the workflows needed to connect the various components of plan exposure, and address versioning issues. The content will initially be abstracted from the case studies describing the planning lifecycle and workflows used for each of the use case scenarios.

The Reference Model will be platform agnostic, allowing for the possibility that alternative combinations of components may be used (for example a DMP might be 'published' through alternative routes). The model will assume that components interact through published APIs and open standards (e.g. SWORD). As far as possible it will also use outputs of the DMP Common Standards WG as these become available.

A set of recommendations for relevant stakeholder groups will be included in this final report, based on the analysis of the consultation responses presented in the Use Cases Catalogue and Report, and feedback from the Active DMP IG and broader RDA community on the earlier deliverable.

Working group operation, communication and engagement

The working group co-chairs will ensure the group operates by consensus. Where consensus cannot be achieved decisions will be taken by majority decision of the group, or its co-chairs, and at all times according to RDA principles of openness, balanced representation, technology neutrality, and harmonization across communities and technologies.

The working group will meet on a monthly basis, online and at RDA Plenaries and other events, including at the IDCC (International Digital Curation) conference in February 2017, 2018. Advance notice of meeting arrangements will be circulated online to the group's RDA mailing list, and meeting notes will be circulated by the same route.

The co-chairs and members have access to a broad range of other means of communication that will be used to engage with the broader community, initially to seek comments on the case statement. These

include websites, mailing lists and social media associated with DMP authoring tools, journals and stakeholder organisations that have an international and multi-sector scope.

Co-chairs will develop a detailed workplan with further milestones. This and all further discussion documents and related material resulting from the group's work will be posted in a file repository on the RDA website, and made available under a CC-BY 4.0 licence.

Adoption Plan

The co-chairs and initial membership is well positioned to ensure that outputs are considered for adoption by the broad range of stakeholders interested in plan exposure. These include for example:

- DMP authoring, guidance and recommendation services: Australian National Data Service, Digital Curation Centre, California Digital Library, Elsevier RDM solutions
- Active data management services: Center for Open Science
- Journals and publishers: Elsevier RDM Solutions, Digital Curation Centre (IJDC), F1000Research and eLife
- Institutions: University of Edinburgh, University of Hagen, University of Lincoln, University of Lisbon, MIT, University of Porto,

Initial Membership:

Initial co-chairs (*) and members of the WG are listed below.

Mike Brown mjbr@ceh.ac.uk (Centre for Ecology and Hydrology, UK)

João Cardoso - joao.cardoso.geral@gmail.com University of Lisbon / INESC-ID Portugal

David Carr - d.carr@wellcome.ac.uk The Wellcome Trust

João Aguiar Castro - joaoaguiarcastro@gmail.com University of Porto

John Chodacki - John.Chodacki@ucop.edu CDL - DMPtool

Leighton Christiansen - leightonlc@gmail.com , NTL USDOT

John Faundeen, Earth Resources Observation Center, USGS

Heike Görzig - heike.goerzig@fernuni-hagen.de Univ of Hagen

Bev Jones - bjones@lincoln.ac.uk University of Lincoln, UK

Peter McQuilton - peter.mcquilton@oerc.ox.ac.uk (Oxford, UK/BioSharing.org/FAIRsharing)

* Natalie Meyers - natalie@cos.io Center for Open Science, Open Science Framework Collaborations Manager

* Fiona Murphy - fionalm27@gmail.com Publishing/Research Data Consulting

Amy Nurnberger anurnberger@columbia.edu (Columbia University), (MIT from 22 June: nurnberg@mit.edu)

Cristina Ribeiro - mcr@fe.up.pt University of Porto/ INESC TEC Portugal

* Kathryn Unsworth, kathryn.unsworth@ands.org.au Australian National Data Service

* Angus Whyte - a.whyte@ed.ac.uk Digital Curation Centre - DMPonline – Associate editor at IJDC

Elena Zudilova-Seinstra E.Zudilova-Seinstra@elsevier.com (Elsevier RDM Solutions)