Exposing DMPs WG
Recommendations

Angus Whyte, Natalie Meyers, Fiona Murphy, Marie-Christine Jacquemot, Kathryn Unsworth

APRIL 2021
DRAFT UNDER REVISION Version 0.4

Introduction

Background to the Working Group

The Exposing DMPs working group presents their recommendations for further comment. These recommendations respond to growing interest in exposing data management plan content to other actors (human/machine) in the research lifecycle. By 'exposing' we mean sharing with stakeholders other than the DMP author, funding body and institutional staff who would normally have access to information on a research project or proposal. Our working group aims to identify effective, efficient and ethical practice in this area. With a new standard for expression and interchange of DMPs available, we aimed to better understand user needs, and the benefits and risks to stakeholders of different modes of action.

The working group has run a survey, interviews, plenary session discussions and polls to inform its draft recommendations. This work has surfaced the following main themes:
- Benefits from sharing exemplars to help learn data management planning, and from the better availability of information about data management costs
- Risks of sharing sensitive information from the DMP, concerns about scooping and about changes to plans being perceived negatively
- Applying FAIR principles to DMPs. This may be to improve DMPS as tools for making research outputs FAIR. Or to aid transparency, by making DMPs FAIR as records of planning and execution.

- The need for DMPs to be machine-actionable, to enable their integration into the research workflow, and interoperability with institutional or external systems and services

DMP exposure may involve the private or public sharing of DMP content, or metadata about a DMP. This could in principle happen at any time during research activity, or after a research project has been finalised. When we refer to 'private' sharing we include interaction between the DMP author(s) and the Principal Investigator(s) of the research, if they are not themselves the author. Private sharing would also include communication with the Research Funder, and any Research Support Services normally involved in reviewing and advising on a DMP, possibly including a Data Steward.

The distinction between private and public sharing of a DMP is not, in practice, well defined. Content might be shared with a Data Service Provider (e.g. a repository). This could be internal to an Institution, outsourced to an external provider, or a combination of both, and could include DMP Platform provider(s). Public DMP sharing is likely to involve these and other service providers - publishers, repositories and catalogues, for example. These could also support some measure of private sharing, e.g. to facilitate peer review, or facilitate restricted access.

There is potential value in exposing plans for a variety of the stakeholders involved in their production and consumption. These stakeholders will benefit significantly from adoption of the recommendations outlined below, which address shared interests in using Data Management Plans to demonstrate that research products have been managed according to research community standards and generic principles (e.g. that the research products should be FAIR), and in giving recognition to researchers and others for their efforts in making this happen.

Methods used

The first step involved analysing the DMP Common Standards User Stories.[1] This analysis resulted in eight Use Cases:

- **Deposition:** Submit DMP to a repository or registry
- **Estimation:** Mine individual or collected plans for requirements planning
- **Evaluation:** Review DMP for completeness and policy alignment
- **Integration:** Integrate DMP in research workflows

[1] ' Use cases for Exposing DMPs - for Plenary-13 discussion: https://www.rd-alliance.org/system/files/documents/Use%20cases%20for%20Exposing%20DMPs%20-%20background%20paper.pdf

- **Notification:** To notify services of anticipated resource and support needs
- **Publication:** Publish DMP for research visibility
- **Resourcing:** Costing the planned data management activities
- **Transparency:** Ability to see (updated) record of output which describes management intentions and actions

The Use Cases informed the design of the survey instrument - https://www.rd-alliance.org/system/files/documents/ExposingDMPsQuestionnaire.pdf. Qualtrics Survey Software was used to develop and administer the survey. The survey was distributed via various channels available to the co-chairs, e.g. DMPOnline users, DMPTool users, coordination fora such as CODATA, French Committee for Open Science, Swiss DLCM. There were 571 responses: 409 complete, 42 at 52% complete and 120 at 13% complete.

The data were de-identified and uploaded to the RDA Exposing DMPs file repository - https://www.rd-alliance.org/system/files/documents/ExposingDMPsSurveyData.txt and the results data is visualized online

The Survey underpinned the interviews that followed. Stakeholders interviewed included funders, service providers, institutions and repositories. These interviews have further refined our understanding of the landscape and why, how, when and where DMPs can or should be shared.

Importantly, one of the major information collection and feedback points has been successive plenaries, starting at Plenary 11 in Berlin through to Plenary 15 in Melbourne. A Mentimeter poll run at Plenary 14 in Helsinki provided the working group with a prioritised list of themes for recommendations.

Potential benefits identified

- mutual learning about data management practice
- feedback on planning for FAIR research outputs, including cost estimation
- better quality of data management if some tasks are automatized, distributed and taken in charge by relevant experts
- planning opportunities for repository managers (in terms of resources required, timing, and ability to advise researchers on best practices before the data are collected)
- opportunities for funders to compare DMPs with datasets deposited on project completion

Potential risks identified

- getting scooped on active research
- disclosing information to enable re-identification of personal data
- disclosing confidential or sensitive data
- impacting data security (if data transfer or storage security measures exposed)
- picking up/following bad data management practice if DMP quality not sufficient
- researcher perception that DMPs are a burdensome administrative obligation

Priority themes for recommendations

Results from a poll during P.14 session

Consent from DMP authors to share content	35
Metadata about DMPs	33
Persistent identifiers for DMPs	41
Exposing DMP landing pages or stubs	17
Licensing DMP content for reuse	27
Controlled vocabularies in DMPs	56
Serialisations of Common Standard	22

Summary and Conclusions

Revised recommendations are described in this report following community review, and are summarised below.

WG had intended to provide a *Use Cases Catalogue* to describe use cases, descriptive scenarios, workflows, and further detail the benefits and risks described by their stakeholders in interviews and plenary sessions of the WG and Active DMP IG.

Workflows to be further described

- internal e.g. integration with ethics process (U.Manchester, TU Delft), storage provisioning and management of project outputs (UQRDM, CSIRO RDP, Haplo, FAIR Island) publishing (HRB-> F1000)

Future work, e.g. under the Active DMP IG, is needed to extrapolate a Reference Model from the examples, to describe generic components and workflows for exposing plans (and metadata about them). Following further community review, such a Reference Model would provide a community endorsed approach to using plans for demonstrable advancement in data sharing practice.

Recommendations

Notes:

1. For the purpose of these recommendations the terms ‘data management plan’ (DMP) and ‘data stewardship plan’ (DSP) have a similar meaning. We use DMP as the more widely used term, but recognise that DSP is a preferred term in the GO-FAIR community.
2. DMPs can be ‘active’ (machine actionable - updateable, versioned, integrated, interoperable) during the life of a research project, but can also be ‘static’ snapshot representations of that research project, e.g. attached to a funding proposal at the beginning, or publicly available in a repository or catalogue at the finalisation stage.

Stakeholder groups targeted by the recommendations

Stakeholder categories listed below are based on those used in the report Turning FAIR into Reality. In the interest of clarity these are modified slightly to add DMP authors and DMP platform providers as distinct groups. Both groups overlap with others, e.g. research communities and data service providers.

- **Research communities**: practitioners from all research fields, clustered around disciplinary interests, data types or cross-cutting grand challenges.
- **DMP authors**: practitioners responsible for creating, updating and exchanging plans relating to specific research projects, for management and stewardship of data, software or other research outputs.
- **Data service providers**: domain repositories, research infrastructures (e.g. ESFRIs) and e-infrastructures, institutional, community and commercial tools and services.
- **DMP platform providers**: research infrastructures, institutions, community or commercial providers of tools for creating, updating and exchanging plans for the management and stewardship of data, software, other outputs.
- **Data stewards**: support staff from research communities and research libraries, and those managing data repositories.
- **Standards bodies:** organisations and consortia coordinating data standards and governing procedures relevant to FAIR, e.g. repository certification, curriculum accreditation (e.g. W3C, NIST).
- **Coordination fora**: global and national bodies such as the Research Data Alliance, CODATA, WDS Communities of Excellence, GO FAIR, German Data Forum (RatSWD), Dutch Coordination Point (LCRDM) and similar initiatives.
- **Policymakers**: governments, international entities like OECD, research funders, institutions, publishers and others defining data policy.
- **Research funders:**, national and international research funders, charitable organisations and foundations, and other funders of research activity.
- **Institutions**: universities and other research performing organisations, e.g. institutes.
- **Publishers**: not-for-profit and commercial, Open Access and paywall publishers of research papers and data.
- **Research support services:** Legal experts, risk managers, ethics reviewers, institutional administrators.

The 12 recommendations cover the following 5 areas:

1. FAIR DMPs for FAIR data production
2. Ethical exposure of DMP content
3. Standardised metadata for DMPs

4. Controlled vocabularies in DMPs
5. Persistent identifiers in DMPs and for DMPs

1. FAIR DMPs for FAIR data production

Recommendation 1.1: Make DMP content as open as possible and as closed as necessary.

Target groups: All

Background: Research funders and other influential stakeholders are interested in making DMPs more useful as instruments for FAIR data production. To that end, some now expect DMP content to be treated as a research output, to itself be made more FAIR. This does not necessarily mean that DMP content should be entirely open, but (in keeping with the principle adopted by the European Commission) 'as open as possible, and as closed as necessary' to facilitate effective, efficient and ethical research data management processes.

The EC Expert Group on FAIR recommended in their report *Turning FAIR into Reality* that "The FAIR principles - and related concepts and policies - should be applied not just to data, but to metadata, identifiers, software and Data Management Plans (DMPs) that enable data to be FAIR." (p.11).

EC planning for the EU Horizon Europe programme includes that DMPs will become mandatory project deliverables. This will lead to a large increase in the DMPs that are published as projects funded reach completion, compared with those currently becoming available from Horizon 2020 projects where the DMP requirement was recently strengthened.

National research funders have also …(HRB Ireland/ FAIR Funders study)

Recommendation 1.2 Review DMPs on criteria for FAIR data

Target groups: Research funders, Institutions, Coordination fora, DMP Platform providers, Data service providers, Publishers

Background: Research funders and institutions vary widely in what they expect researchers to provide in a DMP, at what point in the research cycle they require a plan to be submitted and reviewed. There is, however, growing interest in aligning this review process with the criteria used later in the research lifecycle to assess whether data is FAIR. These may offer a basis for assessing a DMP to offer constructive suggestions on improving it, by identifying whether steps are being taken to make the planned data FAIR according to the accepted criteria.

The recommendation is agnostic about who should provide this constructive feedback, or at what stage in the lifecycle.

- LIBER catalogue demonstrates an approach to reviewing DMPs that have been made publicly accessible
- FAIR funders implementation study illustrates interest in better supporting DMP review early in the lifecycle of funded projects, in the interest of increasing the quality of data stewardship practices.
- Used in conjunction with DMP platforms, FAIR assessment tools could help DMP authors, research funders, and data stewards work together to assess the FAIRness of their planned outputs, as well as the FAIRness of their DMPs.

Recommendation 1.3 Ensure that DMP content is clearly licensed for reuse

Target groups: DMP platform providers, Institutions, Data service providers, Research funders

Background: Reuse of any content according to FAIR principles depends on clarity on licensing and on non-restrictive licence choices (CC-BY, CC-0 or equivalents). For any DMP published via an open-access repository or publisher this may be expected similar arrangements as are applied to any other published work. DMP platform providers, Institutions and Data service providers should ensure that DMP content that may be exposed with the intention of fulfilling pre-publication use cases (e.g. for review, or for analysis of anticipated storage requirements) is similarly covered by non-restrictive reuse terms. DMP metadata should preferably be made CC0 to enable this metadata to be mined in aggregate.

Recommendation 1.4 Advocate DMP content exposure

Target groups: Research communities, DMP authors, Data service providers, DMP platform providers, Data stewards, Standards bodies, Coordination fora, Policymakers, Research funders, Institutions, Publishers, Research support services

Background: Survey responses indicate which benefits are mostly widely anticipated by our respondents. The survey options reflected use cases that have previously been discussed in RDA plenary sessions of this WG and DMP Common Standards WG, which in turn reflect published findings from prior community dialogue. Our interviews found some, though limited, reports of benefits actually experienced from sharing DMP content. These benefits are mainly for the peer review of plans, especially to improve information available on cost budgeting for RDM, and for teaching and learning purposes. They will be summarised in the forthcoming WG report summary. Further work is needed to communicate the potential benefits to DMP authors and stakeholders from exposing DMP content, e.g. to link plans to actual research outputs, or to encourage knowledge exchange about current practices in making data FAIR and keeping it FAIR.

Upcycling DMPs project interview, FAIR funders pilot report, Health Research Board (Ireland) interview > gathering cost estimates from DMPs, and benefit to researchers in learning how others in their community make data FAIR and ensure it is kept FAIR.

Donaldson, D. R. (2019, April). “Upcycling Data Management Plans.” Drexel-CODATA FAIR-RRDM Workshop, Philadelphia, PA.

Recommendation 1.5 Investigate risk-benefit trade-offs

Target groups: All

Background: The survey conducted by the working group explored risks to DMP authors and stakeholders from exposing DMP content. Knowledge exchange benefits were reported and these need to be further substantiated (see recommendation 1.3). Further work is also needed to identify actual experience of risks that some respondents perceived, for example of adverse consequences from DMP content being prematurely shared.

[summary of survey findings on risks/concerns…]

2. Ethical exposure of DMP content

Recommendation 2.1 Obtain consent

Target groups: DMP platform providers, DMP authors, institutions, data service providers

Background: There is a common expectation that consent should be obtained from the DMP author or other appropriate person before exposing DMP content and metadata to parties other than the principal investigator or institution performing the relevant research. By default, DMPs must be treated confidentially until consent is obtained.

Survey concerns; interviewed funders expect DMPs to be treated in accordance with the GDPR; EasyDMP interview; CSIRO RDP interview

The recommendation does not distinguish between reasons for keeping DMP content confidential, or define any time limitation. Those reasons may include (for example) :
- sensitive personal data,
- data that are commercial in confidence
- biological/biodiversity related data - endangered species
- culturally sensitive data, etc.

Recommendation 2.2 Identify sensitive DMP content

Target groups: Research support services, DMP platform providers, DMP authors, Institutions, Data service providers, Publishers

Background: The DMP Common Standard describes a minimal set of information for the exchange of DMP content. It is up to the various parties to that exchange to define how access permissions will be dealt with in their business processes and workflows. Assuming that a DMP is confidential by default, people and processes involved in downstream processing of the DMP content, e.g. to fulfil storage needs identified in the plan, need to be able to identify anything in the plan content or metadata that may not be processed further unless specified ethical or legal/statutory requirements are met. For example a research project may identify its plan to use an identified secure storage service for sensitive data. This part of the plan is likely to itself be sensitive, and treated as such by the organisation's information security policies.

The workflows for dealing with any sensitive DMP content will depend on the nature of the sensitivity, e.g. commercial vs personal risks of disclosure. The recommendation implies that legal and ethical requirements are identified before any further exposure of the DMP content. This highlights the desirability of DMP platforms offering support to tag sensitive DMP content, and for integration of DMP workflows with those for ethical oversight and commercial IP advice. These capabilities would enable further integration with downstream workflows e.g. for fulfilling requirements for storage or other research support services.

3. Standardised metadata for DMPs

Recommendation 3.1 Use the RDA Common Standard to expose DMP content

Target groups: DMP platform providers, Institutions, Research funders, Data service providers, Publishers

Background: The Common Standard WG has provided a metadata profile that may be used to exchange DMP content between DMP platforms and other systems. This should enable data services providers and others, potentially within the institution or beyond, to reuse information about the DMP. Pilots are sought of the Common Standard, testing the utility of various serialisation formats. A first move in this direction is the JSON schema developed by Sotiris Tsepelakis that corresponds to the version 1.0 of the DMP Common Standard. https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/tree/master/examples/JSON/JSON-schema/1.0

4. Controlled vocabularies in DMPs

Recommendation 4.1: Employ published terminologies to describe elements of the DMP common standard.

Target groups: DMP platform providers, Data service providers, Institutions, Research funders, Publishers

Background: DMP platform providers recognise the need to use controlled lists of terms in DMPs to aid machine-actionability. Some platforms employ controlled lists for specific DMP elements, for example to enumerate the range of data products that the plan may encompass. Others allow institutional administrators to define their own controlled lists. However further work is needed to apply community endorsed terminologies to describe the various elements of the DMP Common Standard. Some publishers are, for example, taking steps towards applying the CRediT taxonomy to describe data management roles of DMP authors in published DMPs. Research funders have expressed interest in defining common terms for research grant information, following promising steps in this direction to define standard identifiers for people and organisations (see recommendation 5).

Ref. Easy DMP interview, GO-FAIR Funders IN, and discussions with HRB, Wellcome, F1000

List relevant vocabularies as examples - e.g. CRediT

5. Persistent identifiers in DMPs, and for DMPs

Recommendation 5.1: Use DMPs to document project output identifiers

Target groups: DMP platform providers, Institutions, DMP authors, Data service providers

Background: Some DMP platform providers and research funders are interested in the potential benefits to be gained from DMP authors using their DMP throughout their project to record any persistent identifiers assigned to its outputs. Sharing this content, whether publicly or with funding bodies or other stakeholders would help fulfil a number of evaluation use cases, e.g. monitoring the impact of data policies by identifying the extent to which data is shared from projects that have DMPs. Use cases that involve monitoring of individual projects and their outputs are more likely to cause concern among DMP authors,

given the evidence of our survey that potential negative perceptions of changes in their plans are a concern to them.
Datacite is a partner with California Digital Library in current research that seeks to establish mechanisms for DMPs to be automatically updated with the PIDs of related project outputs.

Ref. FAIR Island project, FAIR funders IN, CDL project

Recommendation 5.2: Assign a resolvable PID to the project DMP

Target groups: DMP platform providers, Institutions, DMP authors, Data service providers

Background: DMPs need to be findable if they are to be considered FAIR research outputs in their own right. Machine actionable DMPs need to be uniquely identifiable to fulfil use cases that involve their exchange across organisational systems, and where this occurs early in the research lifecycle it may be desirable for access to some or all of the DMP content to be limited. DMP platform providers therefore need to support institutions to expose DMP metadata using identifiers that resolve to a landing page or stub record containing metadata defined by the RDA Common Standard. There is an implicit need for registries of DMP metadata, to fulfil use cases that involve systems and processes acting on exposed DMP content.

Institutional repositories were the preferred registry mechanism for sharing DMP content among our survey respondents, followed by 'DMP Catalogue', and 'Funder repository'. Examples of the latter categories were the subject of several interviews, and will be characterised in the WG case studies.

Current practice in assigning PIDs to DMPs is limited to the few examples of these being shared publicly in journals and repositories. Of these, the Zenodo repository is probably the most frequently used outlet. DMPs uploaded to Zenodo may be assigned a Datacite DOI and, as with any other item uploaded to Zenodo upload, this will resolve to a landing page associated with that DOI.

Ref. Freya's PID Graph project, CDL project

Recommendation 5.3. Consult further on appropriate PIDs for active DMPs. .

Target groups: Coordination fora, DMP platform providers, Data service providers

Background: There is emerging practice around using DOIs to identify DMPs that are shared at the end of the research cycle, as noted for recommendations 5.1 and 5.2. However there are too few examples of maDMP content being dynamically updated in keeping with the 'active DMP' vision to point to a consensus on how changes in a DMP metadata record should be managed alongside the identifier for that record.

One possible scenario would be that Handles are minted for active DMPs, and DOIs only minted at the point a DMP is exposed for relevant stakeholders or services.
For example in the UQRDM system a document format of the DMP is rendered in pdf and can be exported/shared with relevant stakeholders, but is time stamped as it represents either a snapshot during the active phase of the project, or the final view of the project's data management.

Ref. Freya's PID Graph project, CDL project

Glossary of Terms

Exposing DMPs - DMP content sharing by the DMP author, funding body or institutional staff who would normally have access to information on a research project or proposal, with any other stakeholder.

FAIR Data Principles - are designed to support knowledge discovery and innovation by both humans and machines, support data and knowledge integration, promote the sharing and reuse of data, and can be applied across multiple disciplines. These principles provide guidance for research data management and stewardship and are relevant to all stakeholders listed above.

FAIR DMPs - are created based on the FAIR principles, findable, accessible, interoperable and reusable and adhere to these principles as an output of a research project. They are DMPs that adopt common standards and support 'active' and iterative updating to enable information exchange across the FAIR data ecosystem.
(Adapted from Turning FAIR into reality, 3.4 Data Management Plans and FAIR)

Machine-Actionable DMPs - provide the ability to report on the intentions and outcomes of a research project, enabling information exchange between relevant parties and across associated systems.
(Adapted from DMPTool blog - posted July 9, 2018 by Stephanie Simms, Scoping Machine-Actionable DMPs)

PIDs - a persistent identifier is a long-lasting reference to a digital resource. (ORCID website - https://support.orcid.org/hc/en-us/articles/360006971013-What-are-persistent-identifiers-PIDs-)

Sensitive data - Sensitive data is data that must be protected against unwanted disclosure. Access to sensitive data should be safeguarded. Protection of sensitive data may be required

for legal or ethical reasons, for issues pertaining to personal privacy, or for proprietary considerations. https://www.openaire.eu/sensitive-data-guide

Serialisation - is the process of converting an object into a stream of bytes so that it can be transferred over a network or stored in a persistent storage location. Serialisation formats, e.g. JSON, XML.