Ten simple rules for machine-actionable data management plans

Tomasz Miksa[1*◐], Stephanie Simms[2◐], Daniel Mietchen[3◐], Sarah Jones[4◐]

1 SBA Research & TU Wien, Vienna, Austria
2 California Digital Library, Oakland, United States of America
3 Data Science Institute, University of Virginia, Charlottesville, United States of America
4 Digital Curation Centre, Glasgow, United Kingdom

◐These authors contributed equally to this work.
* miksa@ifs.tuwien.ac.at

Abstract

Data management plans (DMPs) are documents accompanying research proposals and project outputs. They describe data and tools employed in scientific investigations, mostly in free-form text. DMPs are often seen as an administrative exercise and not as an integral part of research practice.

There is now widespread recognition that the DMP can have more thematic, machine-actionable richness with added value for all stakeholders: researchers, funders, repository managers, research administrators, data librarians, etc. 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. This will enable parts of the DMP to be automatically generated and shared, thus reducing administrative burdens and improving the quality of information within a DMP.

This paper presents 10 simple rules outlining specific steps to put machine-actionable DMPs into practice and realize their benefits.

Introduction

Data Management Plans (DMPs) are documents accompanying research proposals and project outputs. They describe the data that is used and produced during the course of research activities, where the data will be archived, which licenses and constraints apply, and to whom credit should be given. The fundamental purpose of a DMP is to help researchers manage their data. DMPs are typically created manually, mostly by researchers using checklists and online questionnaires. They are required by funding bodies and institutions all over the world, e.g. the National Science Foundation (NSF) in the USA, the European Commission in Europe, and the National Research Foundation (NRF) in South Africa.

The current manifestation of a DMP—a static document often created before a project begins—only contributes to the perception that DMPs are an annoying administrative exercise and do not support data management activities. Questions can remain unanswered, or the answers can be overly generic due to the use of free-form text. What DMPs 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 (often in addition to physical materials, which can also be described in a DMP). We continue to need a human-readable narrative, but there is now widespread recognition that the DMP could have more thematic, machine-actionable richness with added value for all stakeholders. This includes researchers, funders, repository managers, administrators, data librarians, and so on (as visualised in Fig 1)—in short, everyone who is part of the larger ecosystem in which data is produced, transformed, exchanged, and reused.

What we propose

In this paper, we describe 10 rules for machine-actionable DMPs (maDMPs). 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. This will enable parts of the DMP to be automatically generated and shared, e.g., with collaborators and funders. Furthermore, researchers whose data is reused in other experiments will gain recognition and credit because their data can be located, reused, and cited more easily.

To achieve this goal, all stakeholders must coordinate efforts to realize a new generation of machine-actionable DMPs that contain an inventory of key information about a project and its outputs. The basic framework requires common data models for exchanging information, as well as a shared ecosystem of services that send notifications and act on behalf of humans. Other essential components of the machine-actionable DMP vision include machine-actionable policies, persistent identifiers (PIDs) used in new settings—e.g., ORCID IDs, funder IDs, and new initiatives such as Org IDs [1]—in addition to the removal of barriers for information sharing. By implementing and experimenting with these components, we believe that the global research community can reduce the administrative workload on all stakeholders and enhance the quality of recorded information.

For example, new and/or existing services could consume information provided by a researcher on the amount and type of data they will produce and automatically suggest a proper license, estimate costs of storage, and notify a repository operator to reserve space for a future data deposit. In this manner, we can reduce the input needed from researchers and make their decisions actionable, rather than just describing them.

What we do not propose

These 10 rules outline specific steps that must be taken to put machine-actionable DMPs into practice and begin to realize their benefits. The rules are independent of any tool or technology, and not related to any specific DMP template or funding organisation. They do not contain guidance for researchers writing a traditional DMP, as those rules exist already and can be found in [2].

Target audience

This paper is addressed to a wide range of stakeholders involved in research data management workflows (Fig. 1). The primary audience is those with the greatest ability to bring this machine-actionable DMP vision to life, i.e. policy makers, funders, and institutions. Broad adoption by all stakeholders is required to achieve the benefits, but researchers cannot follow the rules if the infrastructure providers do not provide supporting systems.

Fig 1. Target audience. Stakeholders with a role to play in realizing the machine-actionable DMP vision.

Methodology

We want to emphasize that machine-actionable DMPs are part of a global community effort to improve traditional DMPs and the quality of research data (and metadata) more generally through automation, while also reducing administrative overhead. The substance and inspiration for the rules is based on community-generated use cases from a workshop held at the International Digital Curation Conference (IDCC) in Edinburgh in 2017 that gathered almost 50 participants from Africa, America, Australia, and Europe [3]. The 10 rules themselves have gone through multiple drafts since then via consultations with Research Data Alliance and FORCE11 groups focused on DMPs. The current phrasing takes into account all of the feedback received through various channels in the research data management community.

How to read the rules

All 10 rules are equally important and can be read in any order (Fig. 2). Some rules depend on others, for example, to implement a common data model, we need PIDs and controlled vocabularies (i.e., Rule 6 depends on Rule 5). We indicate these dependencies and relationships between rules in the text.

The rules also vary in scope and specificity. Some are narrower, (e.g., Rule 3: Make policies [also] for machines, not just for people), and some are broader (e.g., Rule 8: Support data management evaluation and monitoring). This is because rules address a combination of technical, organisational and social issues that can be defined on different levels of precision.

Another important point is that we consider data and metadata jointly throughout the paper. This encompasses basic project metadata that should be part of any DMP (e.g., project title, abstract, institution, names of the people involved, and associated identifiers, as per Rule 5) as well as the research data that are described in the DMP and accompanied by appropriate metadata when preserved in a repository. It also extends to things like metadata about the repository and related policies. The idea is to apply these rules to any piece of information or infrastructure that supports effective and efficient management of research data.

The rules can be understood in a different manner by different stakeholders within the DMP ecosystem. When developing the rules, we kept in mind three roles that represent a majority of stakeholders:

1. policy making and infrastructure provision,

2. DMP authoring and updating,

3. using and reusing (DMPs directly, data indirectly through DMPs).

We did not distinguish the rules by stakeholders on purpose, as roles and responsibilities overlap and vary across domains, institutions, countries and projects as well as along the timeline of the research.

Rule 1: Integrate DMPs with the workflows of all stakeholders in the research data ecosystem

Good data management requires precise information on various aspects of data ranging from methodological and technical details on formats and infrastructure to legal and ethical aspects of data collection and reuse.

Authoring DMPs should not be the responsibility of a single person, but has to become a collaborative exercise, in which various stakeholders share their expertise who are knowledgeable in their domains and adjacent parts of the data management ecosystem. Only then can we ensure that the right information is provided and can be acted on by others.

Information provided in DMPs is also consumed by multiple stakeholders (see Fig. 3). For example, repository operators set embargo periods and assign licenses for repository content based on information in the DMPs that was provided by researchers, while research funders check whether research outputs that have been published or deposited in repositories follow relevant policies and guidelines, such as the FAIR principles [8].

Multiple stakeholders provide information in DMPs, and multiple stakeholders consume it. Coordination between them is key—otherwise, the current situation in which DMPs are considered as bureaucracy will continue. Traditional DMPs are typically written at the beginning of a project and hardly ever used later. This way, too many opportunities to use, update and reuse the information held within them are missed. Moreover, the many-to-many relationships of a wide variety of stakeholders contributing and/or consuming different elements of DMPs are not well supported by DMP-related infrastructure currently.

To change this, we need to involve all stakeholders throughout the data management lifecycle, starting from project planning, through project execution to project end and preservation (cf. Rule 9). The machine-actionable DMPs and their common model (cf. Rule 6) will facilitate the structuring of information, but this has to be complemented by organisational and technical means that involve the various stakeholders at all stages of data management who provide and reuse information from DMPs.

Organisational changes should ensure that tasks related to data management become routine and not ad hoc actions. For example, legal experts should be involved in selecting licenses, while IT experts should advise on the best tools and infrastructure to conduct research and manage data. This has to be complemented by technical solutions that provide means for collaborative work and exchange of information, for example, by sending automatic notifications to specific stakeholders when input or other actions are expected from their side. Technical solutions can help in automating routine data management tasks (cf. Rule 2). We want to emphasise that technical solutions must be complemented by human expertise and actions throughout the process.

Rule 2: Allow automated systems to act on behalf of stakeholders

The full involvement of all stakeholders in research data management (cf. Rule 1) depends on having systems to automatically act on their behalf, thus reducing the need

1 Integrate DMPs with the workflows of all stakeholders in the research data ecosystem
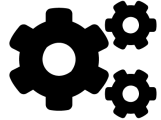
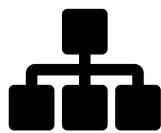**2** Allow automated systems to act on behalf of stakeholders

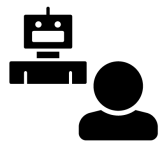**3** Make policies (also) for machines, not just for people

4 Describe—for both machines and humans—the components of the data management ecosystem

5 Use PIDs and controlled vocabularies

6 Follow a common data model for maDMPs

7 Make DMPs available for human and machine consumption

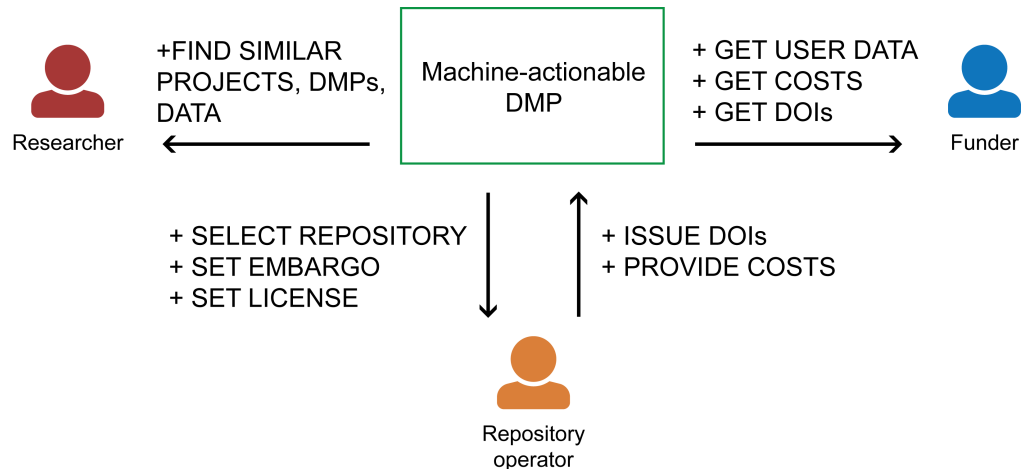**8** Support data management evaluation and monitoring

9 Make DMPs updatable, living, versioned documents

10 Make DMPs publicly available

Fig 2. 10 rules for machine-actionable DMPs at a glance.

Fig 3. Stakeholder interactions. Examples of stakeholder interactions within the ecosystem of machine-actionable DMPs. Stakeholders communicate with each other by exchanging information through DMPs. For example, a repository operator can select a proper repository, set an embargo period, and assign a correct license to data submitted by researchers. In return, a system acting on behalf of a repository operator provides a list of DOIs assigned to the data and provides information on costs of storage and preservation. This in turn can be accessed by a funder to check how the DMP was implemented. Researchers can browse DMP catalogues using a variety of filters that allows them to discover projects using similar methodologies or infrastructure or producing similar outputs.

for human interaction, while helping to focus the remaining human interactions on tasks that cannot readily be automated.

Some of the information to be captured in a DMP is already available electronically, so instead of entering it again, it would be helpful if the relevant bits could be fetched from appropriate sources, perhaps after consistency checks with other sources for quality assurance.

To make this happen, we need to integrate systems and allow stakeholders to expose services that automate their tasks and act on their behalf, for example:

- collating administrative data — a service that acts on behalf of researchers or other DMP authors and collects administrative information, such as affiliation, grant number and postal or email addresses from institutional databases like Current Research Information Systems (CRIS) or Research Information Management (RIM) systems more generally to prefill the DMP, thus reducing the necessity to provide the same information all over again. Information could also flow from the DMP into the CRIS, and if previous DMPs are in the system, relevant bits (e.g. about instrumentation or data formats) could be fetched from there to assist in authoring or reviewing another DMP.

- cost estimation — a service that acts on behalf of repository operators and implements a cost model of a repository to provide automatic estimates of costs of storage and preservation based on input parameters such as amount of data, type of data, project duration, etc. There has been research on cost models and ways of comparing them [9], but there is still no such service in place.

- license selection — a service that acts on behalf of legal experts and proposes a license for data sharing, taking into account policies that apply to the project and

type of data. For example, if the institutional policy recommends open access publishing and the data do not contain sensitive information, then CC0 could be the default setting for data, and CC BY for text and media. There is already a wizard from EUDAT [4] that offers similar functionality.

- storage booking — a service that acts on behalf of an infrastructure operator and reserves storage space for the duration of a project if a repository suitable for the expected types and amounts of data and meeting relevant policy requirements can be found. Furthermore, such a service can help repository managers plan investments into infrastructure when knowing in advance how much new data is expected within the planning period.

- data deposit — a service that acts on behalf of a repository operator to deposit data and associated metadata, using information from the DMP, such as embargo periods, license types, and metadata standards to automatically set properties of ingested data.

- validation and compliance — a service that acts on behalf of a funder and checks compliance with its policies, for example, by checking whether data described in a DMP is accessible by the indicated time and under appropriate licenses.

These examples show that automation is possible for the majority of stakeholders during various phases of research data lifecycles. This helps to save time and reduce costs, while also providing more precise information.

Apart from services automating tasks, we need a system that triggers automated notifications when human intervention is needed (cf. Rule 7). For example, it can create a ticket and assign a human who will then either provide the missing information directly or contact the researcher if clarification is needed.

Rule 3: Make policies (also) for machines, not just for people

Interactions amongst humans as well as between humans and human-made systems are guided by cultural norms, some of which are formalised as legal documents like guidelines, contracts, policies or laws. For simplicity, we refer to them collectively as policies.

There may be various policies relevant to a given DMP, e.g. on data sharing, data quality, data security or ethical review. While policies usually agree on a broader goal, they often handle details in different fashions, which makes it hard for any of the relevant stakeholders to find out whether any particular data have been handled in a way that is compliant with applicable policies.

Policy statements may be very broad, e.g. "Research data will be managed to the highest standards throughout the research data lifecycle as part of the University's commitment to research excellence" [5], or they may be specific enough to be easily applied and tested. More specific requirements could be broken down into a set of rules checking certain properties (e.g. is the resource available? does it have a PID? is it registered?).

Data policies should themselves be machine actionable, at least at some basic level, to assist in the evaluation of data management practice. This can be achieved in several ways, e.g. by

- composing policies using machine-actionable policy elements (cf. [6])

- including a machine-actionable section into policy documents or

- complementing a policy by a closely associated machine-actionable document (e.g. an appendix).

The common feature of these three approaches is that the key requirements of the policy should be expressed in a format that machines can act on, i.e. using a consistent predefined structure and a controlled vocabulary. While people—free-thinking ones like researchers in particular—hate to follow policies, machines are happy to comply when properly instructed. Investing effort in making data policies less ambiguous, more discoverable and machine-actionable will pay dividends, helping funders, publishers and other stakeholders achieve much higher adoption.

Rule 4: Describe—for both machines and humans—the components of the data management ecosystem

A common problem faced by researchers is how to find a suitable repository for data sharing and preservation. There is a wide range of repositories differing in parameters like the types and amounts of content they accept, levels of trust, geographical location, costs, licensing and so on. Each repository provides this information in a different form or even language – sometimes, it is included in the terms of use, in other cases part of an FAQ, or it may not be specified at all and only provided upon request.

If we provide a common way to describe specific components of a data management ecosystem—such as repositories—then these components can be readily discovered by humans and machines. Specifically, in the case of machines, we would be able to create services (cf. Rule 2) that can suggest a repository using information already provided in a DMP. Thus, authors of DMPs would be presented with a list of repositories that fulfill their criteria, and the selection will be narrowed down to those that are relevant.

Conversely, stakeholders who described their services and infrastructure using such standard terms could be informed of parties who selected their services in a DMP, and have greater confidence that those parties are aware of the associated conditions. In the case of repositories, such conditions could be matching data and metadata standards, and checking such matches reduces the effort required for ingesting and maintaining the data.

This rule goes beyond repositories to include all other components of the data management ecosystem that need to be discovered by humans and machines. It should not be confused with Rule 6 that recognises the need for a common data model for DMPs themselves, because the common way to describe specific components of the data management ecosystem enables service discovery, that is, finding resources that may be relevant for DMP creation or in terms of automated notifications, while a common data model for DMPs is a way to model information that is at least in principle known to the DMP authors.

Rule 5: Use PIDs and controlled vocabularies

To make DMPs explicit and understandable for all stakeholders (cf. Rule 1 and Rule 7), we need well-defined terms and precise identification of resources.

The free-form text fields dominating traditional DMPs can contain complex and/or ambiguous terms that leave a lot of room for interpretation. This can lead to situations where it is not clear what data was used in an experiment, where the data will be deposited, or to whom the provisions in the DMP apply.

Sometimes, the opposite is the case: the wording is specific and thus understandable in a very narrow context, requiring implicit knowledge on the part of reusing parties. This can become an issue when data is reused in a different domain or even when the DMP is co-created by various stakeholders (cf. Rule 1).

Furthermore, DMPs are living documents (cf. Rule 9), and the amount and granularity of information contained within them evolves over time—from high-level estimates and expectations down to precise descriptions of actions that have actually been taken.

For this reason, to implement machine-actionable DMPs, we need to use controlled vocabularies and PIDs whenever possible. Controlled vocabularies provide a list of common, well-defined terms that can be used to annotate data or to provide users with a limited list of options to choose from when describing their data or associated workflows. PIDs provide a way to identify and locate resources. They can be used to refer to people and publications, as well as datasets, file types, repositories, organisations, policies and other elements of the research data ecosystem. For example, principle investigators can be identified using their ORCID IDs, and their data using DOIs. Additional PID systems can also be developed to identify other resources, such as specific instances of a given repository software or a cell line.

In cases when a corresponding identification system does not exist, machine-actionable DMPs can employ controlled vocabularies instead. For example, researchers should be able to choose their affiliation by default from a controlled list of institutions. In a similar fashion, they should be able to select rather than type the appropriate metadata standard or a license for their data. This would alleviate generic and meaningless descriptions known from traditional DMPs, such as: "best community practices and standards will be used to document all outputs produced by researchers working in this project".

Rule 6: Follow a common data model for machine-actionable DMPs

A common data model is a medium for exchange of information between stakeholders (cf. Rule 1). It provides information in a machine-actionable form, thus enabling interoperability of tools and services that act on behalf of stakeholders (cf. Rule 2).

The common data model is not a prescriptive template or a questionnaire, but provides a reusable way of representing machine-actionable information in a structured way on themes covered by DMPs. It models information, which contrasts with the free-text information gathered by the questionnaires known from traditional DMP tools.

Due to a wide range of topics covered by the DMPs used in different disciplinary, national or other contexts, the model should be modular. It should have a core model common for all DMPs and a clear mechanism for including extensions that describe specific aspects of data management or that address specific domain requirements. It should also reuse existing standards, controlled vocabularies, and models to organize information in a systematic way (cf. Rule 5).

The common data model does not affect the internal architecture of specific components within the data management ecosystem—each component (e.g. a repository) can model the information internally in the way that is best for its purpose, but when information is exchanged across components, then this information must be modelled using the common data model.

The common data model remains transparent to the stakeholders authoring and updating DMPs: when their input is needed, they will be notified and presented with relevant information (cf. Rule 2). The common data model is used by these tools to read

and write information to and from the machine-actionable DMP and to automatically take actions based on the information therein. To implement this rule, all stakeholders involved in the data management ecosystem must adopt a common data model.

Rule 7: Make DMPs available for human and machine consumption

So far, DMPs are written with the expectation that they will be read a maximum of about once by a human.

By converting DMPs into living documents (cf. Rule 9), they become more likely to be consulted multiple times throughout the course of research. This works best if not just the most current version is readily accessible but differences between versions can be assessed by both humans and machines.

It would also be helpful if interested parties could subscribe to automated notifications of changes to a specific DMP, ideally in a way that allows for different levels of granularity. For instance, project collaborators may be interested in the full content of the DMP, whereas the repository named as the destination of a specific subset of the data may only be interested in changes to the amount, licensing, deposition date or format of that specific data subset.

Enabling such granular notifications requires the DMP to be machine actionable at corresponding levels of granularity. This necessitates avoiding free text and providing structured information whenever possible. Some form of human-readable narrative will remain necessary but DMP content that is structured, machine-readable and actionable increases the potential for reuse.

By turning DMPs into public documents (cf. Rule 10), they are more likely to be consulted by multiple humans and machines. Having machine-actionable DMPs would also facilitate the aggregation of DMPs at the available levels of granularity. For instance, infrastructure providers or funders may be interested in dashboards aggregating project-based DMPs on an ongoing basis and reslicing them in various ways, e.g. by the institutions associated with these DMPs, by the designated infrastructure, by the funding mechanism or by the kinds of data.

Rule 8: Support data management evaluation and monitoring

If DMPs are to be taken seriously, they must be evaluated along with grant proposals, and during active stages of research. Reviewers and other stakeholders still need a human-readable narrative, but providing policies in machine-actionable formats (cf. Rule 3) would also assist in automated monitoring, e.g. of research outputs or compliance with applicable policies. DMPs should be explicit about the policies they are meant to comply with, and include version numbers and PIDs to avoid ambiguity.

Involving stakeholders in the process of DMP authoring (cf. Rule 1) and use of controlled vocabularies, PIDs (cf. Rule 5) and a common data model (cf. Rule 6) improves the quality of information contained in DMPs. This is because fine grained information will be provided in a structured way, and many associated tasks can be automated.

For example, in an early phase of a DMP creation, the tools can check whether a selected license for data sharing is compliant with a funder policy. In a later phase, when data was created and was supposed to be deposited in a repository, the tools can automatically check whether the data in question was deposited there and is accessible

and licensed as prescribed by applicable policy. This would considerably change the way that data management actions are reviewed by the relevant stakeholders.

However, machine-actionable DMPs should never be an evaluation means on their own. DMPs only provide means to monitor data management practices. DMPs must reflect reality (or realistic planning), even if that differs from best data management practices. DMPs also cannot impose limits on research methodology and must allow investigations to be implemented using any technology of choice.

Rule 9: Make DMPs updatable, living, versioned documents

It is unhelpful to think of DMPs as static documents. They should not just be seen as a "plan" but as updatable, versioned documents representing and recording the actual state of data management as the project unfolds. The notion of Data Management Records [7] to move beyond a plan has been put forward in this vein. The act of planning is far more important than the plan itself, and to derive value for researchers and other stakeholders, the plan needs to evolve. DMPs should track the course of research activities from planning to sharing and preserving outputs, recording key events in the lifecycle of the data to become an evolving record of activities related to the implementation of the plan.

Changes to machine-actionable DMPs should trigger notifications at configurable levels of granularity to inform interested stakeholders accordingly (cf. Rule 7). For example, such notifications could inform research communities about amendments to the conditions under which forthcoming datasets will be made available, or alert them as the datasets are deposited. As well as issuing notifications, systems could exchange updated data directly. As a new event is recorded in one system, it could automatically pass the new entry to CRIS / RIM platforms, grant management systems, repositories or other related tools.

Updating the DMP might not always need human intervention. Some of the changes could be done automatically, triggered by events elsewhere in the research ecosystem, e.g. when data was deposited, the DMP could be updated with timestamp and PID of the dataset. Conversely, some of the changes to a DMP (e.g. personnel changes) may need to be made by hand, but could trigger notifications elsewhere in the system. In both cases, this requires that the information is available in a machine actionable manner and that the notification mechanism is linked to some tracking tool that is aware of the relationships of the given DMP with relevant external resources and actors.

Rule 10: Make DMPs publicly available

The DMP is the earliest concrete indication of what data will be created in the framework of a research project and how it will be managed. Sharing and co-creating the DMP within the project team during the ideation and planning stages helps to specify and clarify the research methodology, to estimate required resources, and to produce a plausible timeline for data release.

Sharing it beyond the project team—e.g. within an institution, with repositories, funders or ethical review boards—from early on (as per Rule 1) helps streamline data-centric interactions between the various stakeholders over the course of the project.

Stakeholders with access to multiple DMPs (or consistent sections thereof) can aggregate them and—particularly for the subset that is machine actionable—mine the information contained therein and reslice it by the different parameters of the DMP data model (cf. Rule 6 and Rule 7). This informs RDM service delivery, facilitates

monitoring and evaluation (cf. Rule 8) and stimulates the development of tools to explore such DMP corpora and to enable humans and machines to interact with them (cf. Rule 2 and Rule 7).

Ideally, DMPs should be shared early and often (cf. Rule 9) throughout the research process and as broadly as possible. Where this is not feasible, they could be shared with a delay (e.g. at project end) or in limited contexts (e.g. within an institution) or in part (e.g. project metadata such as grant number, abstract, related outputs). The reasons for not sharing earlier, in full or more broadly should be stated in a machine-actionable manner, for example, through a standardized template in which the opt-out is justified using a controlled vocabulary. This would allow to gather data about such circumstances and could inform future data management policies.

If machine-actionable DMPs are shared in public and under an open license, anyone can aggregate them, reslice the corpora, use and re-share the resulting information. Such front-ends to machine-actionable DMP collections could be generic—which would help with the standardization and spread of good data management practices across domains—or tailored for specific audiences, e.g. to facilitate discovery in a given area or education about research in the domain, including associated data management practices.

Another important use case for sharing DMPs in public is to accompany data that are described by the DMP and deposited in a repository. Since different sets of data may differ in parameters like their thematic scope, their file types, size, or sharing restrictions, they are often not shared in the same way, and it is hard to get an overview of what data has been shared by a given project. If each dataset or other research output—irrespective of where it was deposited—would always point to the appropriate version (cf. Rule 9) of the DMP in a machine actionable manner, users who discover any part of that project's output could easily use the DMP to find the other parts.

This way, individual DMPs would act as a hub to project-level research outputs, and aggregations of DMPs as hubs to research more generally, including to planned or ongoing research and to research infrastructure.

While making even single DMPs machine actionable or versioned or public is beneficial in terms of data management and discovery, the real benefits come once *many* DMPs are machine actionable *and* versioned *and* public.

Acknowledgments

References

1. OrgID — Organization Identifier;.
   https://orcid.org/content/organization-identifier-working-group.

2. Michener WK. Ten Simple Rules for Creating a Good Data Management Plan. PLOS Computational Biology. 2015;11(10):1–9. doi:10.1371/journal.pcbi.1004525.

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

4. EUDAT. License selector tool;.
   https://eudat.eu/services/userdoc/license-selector.

5. University of Edinburgh. Research Data Management Policy;.
   https://www.ed.ac.uk/information-services/about/
   policies-and-regulations/research-data-policy.

6. Bunakov V. Data policy as activity network. In: XIX International Conference on
   Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL
   2017), Moscow, Russia; 2017.Available from:
   https://epubs.stfc.ac.uk/work/35227831.

7. University of Queensland. Data Management Records;.
   https://research.uq.edu.au/project/research-data-manager-uqrdm.

8. Wilkinson, Mark D. et al., 2016. The FAIR Guiding Principles for scientific data
   management and stewardship, Scientific Data 3 Mar, p.160018,
   http://dx.doi.org/10.1038/sdata.2016.18.

9. Kejser UB, Davidson J, Wang D, Strodl S, Miksa T, Johansen KHE, et al.. State
   of the Art of Cost and Benefit Models for Digital Curation; 2014. Available from:
   http://eprints.gla.ac.uk/92479/.