

Data Management Planning: How Requirements and Solutions are Beginning to Converge (post-print)

Authors: Jones, Sarah, Pergl, Robert, Hooft, Rob, Miksa, Tomasz, Samors, Robert, Ungvari, Judit, Davis, Rowena, I, Lee, Tina, Pergl, Robert

Date: 2022-11-16T00:00:00+00:00

Abstract

Effective stewardship of data is a critical precursor to making data FAIR. The goal of this paper is to bring an overview of current state of the art of data management and data stewardship planning solutions (DMP). We begin by arguing why data management is an important vehicle supporting adoption and implementation of the FAIR principles, we describe the background, context and historical development, as well as major driving forces, being research initiatives and funders. Then we provide an overview of the current leading DMP tools in the form of a table presenting the key characteristics. Next, we elaborate on emerging common standards for DMPs, especially the topic of machine-actionable DMPs. As sound DMP is not only a precursor of FAIR data stewardship, but also an integral part of it, we discuss its positioning in the emerging FAIR tools ecosystem. Capacity building and training activities are an important ingredient in the whole effort. Although not being the primary goal of this paper, we touch also the topic of research workforce support, as tools can be just as much effective as their users are competent to use them properly. We conclude by discussing the relations of DMP to FAIR principles, as there are other important connections than just being a precursor.

Full Text

Preamble

Addressing FAIR principles F, A, I, R

Data Management Planning: How Requirements and Solutions are Beginning to Converge

Sarah Jones¹, Robert Pergl^{2†}, Rob Hooft³, Tomasz Miksa⁴, Robert Samors⁵, Judit Ungvari⁵, Rowena I. Davis⁵ & Tina Lee⁵

¹Digital Curation Centre, University of Glasgow, Glasgow G12 8QQ, Scotland, UK

²Czech Technical University in Prague, Faculty of Information Technology (FIT CTU), 160 00 Prague 6, Czech Republic

³Dutch Techcentre for Life Sciences, Utrecht 3521 AL, The Netherlands

⁴Technische Universität Wien, Wien 1040, Austria

⁵Belmont Forum, Western Australia WA 6105, Australia

Keywords: FAIR; Data management; Data stewardship; Machine-actionable; DMP tools

Citation: S. Jones, R. Pergl, R. Hooft, T. Miksa, R. Samors, J. Ungvari, R.I. Davis & T. Lee. Data management planning: How requirements and solutions are beginning to converge. *Data Intelligence* 2(2020), 208–219. doi: 10.1162/dint_a_{00043}

Abstract

Effective stewardship of data is a critical precursor to making data FAIR. The goal of this paper is to provide an overview of the current state of the art of data management and data stewardship planning solutions (DMP).

We begin by arguing why data management is an important vehicle supporting adoption and implementation of the FAIR principles, describing the background, context and historical development, as well as major driving forces, including research initiatives and funders. We then provide an overview of the current leading DMP tools in the form of a table presenting their key characteristics. Next, we elaborate on emerging common standards for DMPs, with particular focus on machine-actionable DMPs. Since a sound DMP is not only a precursor of FAIR data stewardship but also an integral part of it, we discuss its positioning in the emerging FAIR tools ecosystem. Capacity building and training activities are crucial ingredients in this entire effort. Although not being the primary goal of this paper, we also touch upon the topic of research workforce support, as tools can only be as effective as their users are competent to use them properly. We conclude by discussing the relations of DMP to FAIR principles, as there are other important connections beyond simply being a precursor.

† Corresponding author: Robert Pergl (E-mail: perglr@fit.cvut.cz, ORCID: 0000-0003-2980-4400)

1. Introduction

Effective stewardship of data is a critical precursor to making data FAIR, which is why researchers should develop Data Management Plans (DMPs) from the early stages of research. It is obviously desirable to share data wherever possible. This requires obtaining necessary permissions (either via consent agreements or for third-party data), choosing appropriate formats and standards, and providing rich documentation to ensure data are meaningful to other stakeholders, including machines.

This paper covers the convergence in data policy, tools, and standards for DMPs, highlighting opportunities to facilitate the planning process and make better use of the information gathered.

1.1 Convergence: Finding the Common Ground

DMPs are becoming commonplace across the globe. Expectations have been established by several UK and US research funders for well over a decade, and an increasing number of governments, funding agencies, and research organisations are releasing expectations that either require or encourage the development of data plans.

Despite the fact that many funders have individual templates and guidelines for DMPs, there is considerable alignment in content. The Digital Curation Centre (DCC) analysed different UK and international requirements to agree on a set of DMP themes representing the main aspects addressed. These cover topics such as “data formats,” “ethics,” “data sharing,” and “preservation.” The themes are used in DMPonline to allow questions and guidance to be associated. In 2018, the California Digital Library and the Digital Curation Centre revisited this exercise to apply it to many new US funder templates and to consult with the wider Research Data Management (RDM) community on convergence. This resulted in 14 themes that act as a baseline for expectations and were used as initial input to inspire the RDA Working Group on Common Standards for DMPs.

Convergence is becoming increasingly important when projects are executed at multiple institutes and/or funded by multiple different funders. It becomes impractical to create data management plans for each institute and for each funder separately. The DMPRoadmap platform utilized in DMPonline, DMPTool, and other national DMP services addresses this by allowing local requirements to be added to existing funder templates, preventing researchers from having to write one DMP for their funder and another for their institution.

DMPRoadmap is an open source codebase for a Data Management Planning tool, jointly managed by the Digital Curation Centre and The University of California Curation Center (UC3). It represents a major effort to converge on a single solution that married together the best features from earlier versions of DMPonline and DMPTool.

Policy harmonisation can also be a solution to differing requirements, and various groups such as UK Research and Innovation, Science Europe, the Belmont Forum, and the Research Data Alliance funders forum are promoting convergence. Science Europe ran a workshop in 2018 to bring together European stakeholders on RDM requirements and DMP. There are two strands of resulting activity: promoting policy harmonisation and agreeing on a common set of DMP requirements which can be extended by domain-specific expectations (DDPs –domain data protocols). Similarly, the Belmont Forum, an international consortium of 29+ science funding agencies, has worked closely with data organizations such as RDA, CODATA, and science publishers to develop and align its data and digital outputs management requirements template with FAIR and open data best practices. The Belmont Forum is integrating its open data requirements with its online proposal and review processes.

Some research funders use the outcomes of DMP reviews to improve their guidelines. The Economic and Social Research Council (ESRC), Belmont Forum, Health Research Board (HRB) in Ireland, and ZonMw, a medical science funder in the Netherlands, have already released several iterations of their guidelines for researchers with various adjustments to increase clarity and improve the quality of responses.

Funders such as ZonMw, HRB, and Wellcome Trust are exploring ways to enable easier monitoring of DMPs by providing more structured questions that can be automatically assessed and connecting DMPs to local grants systems. Some designated data centres supported by the National Environment Research Council (NERC) in the UK and National Science Foundation (NSF) in the US, such as Woods Hole, are also discussing with tool providers how DMPs can be aligned with local systems and support processes.

2. DMP Tools

Many DMP tools are available worldwide. The earliest date from 2010–2011, when DMPonline and DMPTool were launched in the UK and USA, respectively. The Digital Curation Centre and California Digital Library, which operate these services, converged on a joint open source community-led codebase called DMPRoadmap in 2018. This codebase is used in several international services including DMPAssistant in Canada, DMPTuuli in Finland, DMPiDOR in France, PGDonline in Spain, and DEIC's version of DMPonline in Denmark. In recent years, many other new DMP tools have been released. Most provide functionality to create, share, export, and review DMPs; however, different aspects have been emphasized.

Several focus on providing closed questions instead of free text and underlying knowledge-bases to help prompt and guide users. Others take a project or dataset focus and also link to other tools for data documentation and storage allocation to support implementation. Below is a summary of the main services, highlighting differences in functionality and approaches to deployment

and sustainability.

3. Common Standards for DMPs

Given the range of tools emerging and the opportunity to connect DMPs with other research systems and processes, greater coordination and adoption of standards to enable data exchange and interoperability is needed. A request to establish the DMP Common Standards working group was articulated during the 9th Research Data Alliance (RDA) plenary meeting in Barcelona. The discussion was framed by a white paper by Simms et al. on machine-actionable data management plans (maDMPs). The white paper is based on outputs from the IDCC workshop held in Edinburgh in 2017 that gathered almost 50 participants from Africa, America, Australia, and Europe. It describes eight community use cases that articulate consensus about the need for a common standard for maDMPs (where machine-actionable is defined as “information that is structured in a consistent way so that machines, or computers, can be programmed against the structure”).

The RDA working group on DMP Common Standards has developed a common model for machine-actionable DMPs that enables exchange of information between systems acting on behalf of stakeholders involved in the research life cycle, such as researchers, funders, repository managers, ICT providers, librarians, etc. The group has also implemented prototype workflows to demonstrate how machine-actionable DMPs can be implemented by connecting them to various systems, such as CRIS, repositories, or funder systems.

The model is independent of specific funder requirements and provides a common set of concepts needed to represent DMP-specific information in most settings. Since it is meant as a format for exchange of information between systems, it is also independent of an internal software architecture adopting the common standard. Furthermore, the model can be serialized into different representations, for example, JSON, XML, OWL. In 2019, the maDMPs were summarized into 10 principles.

As of March 2019, the RDA DMP Common Standards Working Group is documenting the model and creating examples in JSON to facilitate adoption. It is also engaging with pilot users who wish to deploy the model. These include tool providers, such as DMPonline, but also universities such as TU Delft. The next steps for DMP Common Standards are to develop further serializations of the model, reach out to new pilot users, and maintain the model by incorporating feedback from its deployments. All DMP tool providers are encouraged to adopt the common data model to promote interoperability across services and reuse of information held in DMPs.

4. DMP Tools in the Context of FAIR Tools Ecosystem

As data management encompasses the whole life cycle of data, all tools related to achieving FAIRness are relevant, particularly the discovery of existing datasets, evaluating FAIRness, and publishing. Tight integration between DMP and FAIR tools is not currently in place, at least not at an appropriate level of adoption and maturity. The first efforts have been formulated in the “FAIR Funder Pilot Programme,” which will attempt to use DMPs as an indication of the prospective FAIRness of data by connecting FAIR metrics to perform an automated assessment. DMPonline has also integrated the RDA Metadata Standards Directory to help researchers identify and adopt relevant standards, and forthcoming tools such as OpenDMP are intended to focus on API integrations to connect different tools in the research system.

The DMP Common Standards work lays important cornerstones for interoperability in the tooling ecosystem.

Another dimension of such an ecosystem is internationalisation. Individual DMP tools offer translations to various European and world languages; however, FAIR through its focus on machine-actionability and strong semantics starts to show a way to perform language-neutral science—a prospect very interesting also for data management planning with possibilities not yet fully fathomed.

5. Research Workforce Support

Capacity building activities are an important complement to data management tools and guidance. Training resources exist in a variety of formats to help researchers successfully prepare a DMP, and more importantly, to implement these plans. As data needs vary across domains, methods, and research objectives, flexibility is an important capacity building consideration. The Belmont Forum has developed a Data Skills Curricula Framework to guide research teams and agencies in developing sustainable DMPs, putting together the components of a comprehensive program for data-enabled research, or to be used as a resource to discover the types of training that exist to develop a customised plan. The Framework emphasizes full-path data management and role-specific approaches to help teams identify who needs which skills, and when to turn to a data professional for assistance.

The RDA/CODATA Research Data Science Schools, which have been running since 2016 to equip early career researchers with core data handling, visualisation, computational infrastructure, and open science skills, are now exploring parallel data stewardship courses supported by the FAIRsFAIR project where researchers and data stewards will be co-taught. The curriculum will promote collaboration and the range of skills and inputs needed to effectively manage and share FAIR data. Resources such as the Data Management Training Clearinghouse, the Belmont Forum e-Infrastructures and Data Management Toolkit, and the FOSTER Open Science portal have compiled globally-oriented free online training materials, many of which have been collaboratively developed in

line with the principles of FAIR and Open Educational Resources.

In ELIXIR, the “Towards Data Stewardship in ELIXIR: Training & Portal” Implementation Study running in 2018–2019 has brought advancements both in the Data Stewardship Wizard tool and in training materials on data stewardship curriculum.

6. Conclusions and Discussion

The following points summarise the relation between the FAIR principles and data management planning and the respective tools:

1. A sound and elaborate data management plan is a necessary precursor to making data FAIR.
2. DMP tools can help become more FAIR by providing feedback—FAIR metrics calculations indicating the assumed FAIRness of data if the DMP is followed (e.g., Data Stewardship Wizard).
3. DMPs can themselves be FAIR. The RDA DMP Common Standards Working Group has delivered a prototype model which should now be implemented by the growing number of DMP tool providers.

If we have a standard expression for the content within DMPs, we will be able to exchange information more easily with the wider ecosystem of FAIR services and provide greater value to researchers, funders, and service providers using DMPs. This covers “I” and “R”. “F” and “A” still need to be achieved by storing DMPs in suitable repositories, which does not generally happen yet. A question, of course, is what information may be needed to be FAIR and what can be shared (DMPs tend to be inherently confidential)—a decision the community has yet to make.

Convergence in research data policy and the adoption of standards for DMP tools is desirable to help clarify the landscape and enable interoperability. We are unlikely to achieve a single DMP template or one tool, and neither is that desirable. DMPs should be tailored to the context in which research is conducted. For some, ethical issues will be prevalent, while for others it will be challenges of scale or timely data sharing. Also, even in the same human language, different research disciplines may use different words to express data management topics; using the right terminology is critical to facilitate usage of the tools and templates.

Author Contributions

R. Samors (miksa@ifs.tuwien.ac.at), J. Ungvari (jungvari@gmail.com), R. Davis (rowenaidavis@email.arizona.edu), and T. Lee (tinal@email.arizona.edu) collectively wrote and revised information in Section 5 on the Belmont Forum’s research workforce support activities and resources and made additional contributions to information in the remainder of the article.

Acknowledgements

Many thanks to the DMP tool providers who shared information on their services and validated the tabular summary. These include Brian Riley of DMP-Tool, Harry Enke, Jochen Klar, and Jens Ludwig of RDMOrganiser, Maria Iozzi, Hanne Moa, and Adil Hasan of EasyDMP, Helen Connick and Andrew Janke of UQRDM, Erich Weichselgartner and Ronny Bölter of DataWiz, Kerstin Lehnert of ezDMP, and Georgios Kakalettris of OpenDMP.

The development of Data Stewardship Wizard was funded partially by ELIXIR, the European research infrastructure for life-science data, and the Institute of Organic Chemistry and Biochemistry AS CR. Several hackathons and workshops were organised and funded by GO FAIR International Support and Coordination Office. Considerable work was done in kind by the Dutch Techcentre for Life Sciences and the Faculty of Information Technology, Czech Technical University in Prague.

References

- [1] S. Jones. Developments in research funder data policy. *International Journal of Digital Curation* 7(2012), 114–125. doi: 10.2218/ijdc.v7i1.219.
- [2] National Institutes of Health, NIH Data Sharing Policy and Implementation Guidance, 2003. Available at: https://grants.nih.gov/grants/policy/data_{sharing}/data_{sharing}_{guidan
- [3] SPARC Europe, Latest update to European Open Data and Open Science Policies 2018. Available at: <https://sparceurope.org/latest-update-to-european-open-data-and-open-science-policies-released/>.
- [4] Digital Curation Centre, Data Management Planning Themes, 2016. Available at: <http://www.dcc.ac.uk/sites/default/files/documents/publications/DMP-themes.pdf>.
- [5] S. Jones. A common set of themes for DMPs: seeking input. Digital Curation Centre, 2016. Available at: <http://www.dcc.ac.uk/blog/common-set-themes-dmps-seeking-input>.
- [6] Digital Curation Centre, DMPRoadmap: DCC/UC3 collaboration for a data management planning tool –Codebase, 2010. Available at: <https://github.com/DMPRoadmap/roadmap>.
- [7] P. Wittenburg, H. Pergl Sustkova, A. Montesanti, S.M. Bloemers, S.H. de Waard, M.A. Musen ...& E.A. Schultes. The FAIR funder pilot programme to make it easy for funders to require and for grantees to produce FAIR data, 2019. Available at: <http://arxiv.org/abs/1902.11162>.
- [8] M. Donnelly, S. Jones & J.W. Pattenden-Fail. DMP online: The digital curation centre' s Web-based tool for creating, maintaining and exporting data management plans. *International Journal of Digital Curation* 5 (2010), 187–193. doi: 10.2218/ijdc.v5i1.152.

- [9] Digital Curation Centre, DMPonline, 2010. Available at: <https://dmponline.dcc.ac.uk/>.
- [10] Digital Curation Centre, DMPOnline -Help for administrators, 2010. Available at: <https://github.com/DigitalCurationCentre/roadmap/wiki>.
- [11] Digital Curation Centre, DMPRoadmap -API Documentation, 2010. Available at: <https://github.com/DMPRoadmap/roadmap/wiki/API-Documentation>.
- [12] California Digital Library, DMPTool, 2010. Available at: <https://dmptool.org/>.
- [13] California Digital Library, DMPTool -Quick start guide, 2011. Available at: <https://dmptool.org/help>.
- [14] EUDAT, UNINETT Sigma2, EasyDMP, 2015. Available at: <https://easydmp.sigma2.no/>.
- [15] EUDAT, UNINETT Sigma2, EasyDMP -User Documentation, 2015. Available at: <https://www.sigma2.no/easydmp/how-to>.
- [16] EUDAT, UNINETT Sigma2, EasyDMP -Codebase, 2015. Available at: <https://github.com/hmpf/easydmp>.
- [17] Dutch Techcentre for Life Sciences, Czech Technical University in Prague, Data Stewardship Wizard (DSW), 2016. Available at: <https://ds-wizard.org/>.
- [18] Dutch Techcentre for Life Sciences, Czech Technical University in Prague, Data Stewardship Wizard (DSW) -Documentation, 2016. Available at: <https://docs.ds-wizard.org/>.
- [19] Leibniz-Institut für Astrophysik Potsdam, Library of the Karlsruhe Institute of Technology, Research Data Management Organiser (RDMO) -Webpage, 2017. Available at: <https://rdmorganiser.github.io/en/>.
- [20] Leibniz-Institut für Astrophysik Potsdam, Library of the Karlsruhe Institute of Technology, Research Data Management Organiser (RDMO) -Tutorials, 2017. Available at: <https://rdmorganiser.github.io/en/tutorials/>.
- [21] Leibniz-Institut für Astrophysik Potsdam, Library of the Karlsruhe Institute of Technology, Research Data Management Organiser (RDMO) -Codebase, 2017. Available at: <https://github.com/rdmorganiser>.
- [22] The University of Queensland, UQ Research Data Manager (UQRDM), 2017. Available at: <https://cloud.rdm.uq.edu.au/>.
- [23] The University of Queensland, UQ Research Data Manager (UQRDM) -Documentation, 2017. Available at: <https://guides.library.uq.edu.au/for-researchers/uq-research-data-manager/set-up-project>.
- [24] Leibniz Institute for Psychology Information and Documentation, DataWiz, 2017. Available at: <https://datawiz.leibniz-psychology.org/DataWiz/>.
- [25] Leibniz Institute for Psychology Information and Documentation, DataWiz -User Guide, 2017. Available at: <https://datawizkb.leibniz-psychology.org/index.php/project-start>.

- [26] Leibniz Institute for Psychology Information and Documentation, DataWiz -Codebase, 2017. Available at: <https://github.com/ZPID/DataWiz>.
- [27] OpenAIRE, EUDAT, OpenDMP, 2019. URL: <https://opendmp.eu>.
- [28] OpenAIRE, EUDAT, OpenDMP -Codebase, 2019. Available at: <https://gitlab.eudat.eu/dmp>.
- [29] T. Miksa, P. Walk, P. Neish, DMP Common Standards WG -Group's Web Page, 2017. Available at: <https://www.rd-alliance.org/groups/dmp-common-standards-wg>.
- [30] S. Simms, S. Jones, D. Mietchen & T. Miksa. Machine-actionable data management plans (maDMPs). *Research Ideas and Outcomes* 3 (2017), e13086. doi: 10.3897/rio.3.e13086.
- [31] RDA DMP Common Standards WG, Official outputs from the RDA DMP Common Standards WG, 2019. URL: <https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard>.
- [32] RDA DMP Common Standards WG, Mockups: Machine-actionable Data Management Planning Application, 2019. Available at: <https://oblassadors.github.io/dmap-mockups/>.
- [33] T. Miksa, S. Simms, D. Mietchen & S. Jones, Ten principles for machine-actionable data management plans. *PLOS Computational Biology* 15(3) (2019), e1006750. doi: 10.1371/journal.pcbi.1006750.
- [34] V. Lucas. Skills gap analysis and curricula for data intensive digital skills in global environmental change research, technical report. Zenodo. 2017. Available at: <https://zenodo.org/record/1404652>. doi:10.5281/zenodo.1404652.
- [35] CODATA, CODATA-RDA School of Research Data Science. 2019. Available at: <http://www.codata.org/working-groups/research-data-science-summer-schools>.
- [36] FAIRsFAIR, FAIRsFAIR -Fostering Fair data practices in Europe -project web page, 2019. Available at: <https://www.fairsfair.eu/>.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv –Machine translation. Verify with original.