1st Conference on Research Data Infrastructure

Poster presentations II

https://doi.org/10.52825/CoRDI.v1i.338

Published: 07 Sept. 2023

Data Management Plan Tools: Overview and Evaluation

Carina Becker [1][https://orcid.org/0000-0001-7595-2012],
Carolin Hundt [2][https://orcid.org/0000-0002-7237-965X],
Claudia Engelhardt [1][https://orcid.org/0000-0002-3391-7638],
Johannes Sperling [2][https://orcid.org/0000-0001-7413-1730],
Moritz Kurzweil [2][https://orcid.org/0000-0003-1413-6673],
Ralph Müller-Pfefferkorn [1][https://orcid.org/0000-0001-8719-5741]

[1]Center for Information Services and High Performance Computing (ZIH), TU Dresden, Dresden, Germany
[2]Directorate – Library and Digital Services, Leibniz Institute for the History and Culture of Eastern Europe (GWZO), Leipzig, Germany

Abstract: Data Management Plans (DMPs) are crucial for a structured research data management and often a mandatory part of research proposals. DMP tools support the development of DMPs. Among the variety of tools available, it can be difficult for researchers, data stewards and institutions to choose the one that is most appropriate for their specific needs and context. We evaluated 18 DMP tools according to 31 requirement parameters covering aspects relating to basic functions, technical aspects and user-friendliness. The highest total evaluation scores were reached by Data Stewardship Wizard (703.5), DMPTool (615.5) and RDMO NFDI4Ing (549.5). The tools evaluated satisfied between 10 % and 87 % of the requirement parameters. 11 tools cover at least half of the parameters. In terms of correlation among the tools, which indicates to which degree their scores in the different requirement parameters are alike, we found the highest correlation for ezDMP and GFBio DMPT. Regarding the relatedness between the tools, 85 % of the DMP tools were positively and 16 % negatively correlated. Accounting for the recent developments in the area of DMP tools, this study provides an up-to-date evaluation that can support tool developers in identifying potential improvements, and hosting institutions to select a tool suited to their specific needs.

Keywords: Data Management Plan, DMP Tools, FAIR, Research Data Management

1 Introduction

Data Management Plans (DMPs) are the basis for a structured research data management throughout the data lifecycle [1]. DMPs facilitate especially the administration of data in research groups by describing the joint handling of the data. Additionally, funders require DMPs in research proposals more frequently [2], [3]. There are a variety of tools to support the development of DMPs. From interdisciplinary DMP tools, which are used to write a generic draft DMP, to discipline-specific DMP tools, which support the creation of a DMP in different research fields, such as psychology, biodiversity, engineering, or life sciences. The manual creation of DMPs is very time-consuming, since researchers have to start from scratch and run the risk of not meeting the funder requirements. By using tools, DMPs can be effectively developed and managed. In view

of the large number of offers, the selection of a suitable tool poses a great challenge for researchers. Therefore, it is crucial to analyze these tools [4]. To support the decision of institutions planning to host a DMP tool, this evaluation can also be helpful. Thus, the objectives of this work are as follows:

A. Identify requirement parameters to evaluate existing DMP tools
B. Evaluate DMP tools based on the identified parameters
C. Determine the relatedness between the DMP tools

2 Materials and Methods

We evaluated 18 mainly open access DMP tools (table 1) based on the identified requirement parameters, of which seven provide discipline-specific and eleven generic templates. Eight tools were developed and hosted in Germany. The remaining DMP tools originate from other European countries, USA and Australia. Although DMPTool is based on DMPonline, we consider these tools individually as they differ in their rating scores.

Table 1. Evaluated DMP tools.

DMP tool	Discipline	Hosting/Developers
UWA-DMP	Interdisciplinary	University of Western Australia, Australia
DMP Canvas Generator	Life sciences	Swiss Institute of Bioinformatics, Switzerland
Clarin-d DMP	Humanities/social science	Eberhard Karls Universität Tübingen, Germany
ARIADNE	Archeology	Vast-Lab, Italy
ezDMP	Interdisciplinary	Columbia University, Rutgers University, University of Illinois, USA
GFBio	Biodiversity	GFBio, Germany
TUDD DMP	Interdisciplinary	TU Dresden, Germany
RDMO	Interdisciplinary	Leibniz Institute for Astrophysics Potsdam, University of Applied Sciences Potsdam, Germany
DataWiz	Psychology	Leibniz Institute for Psychology Information, Germany
TUM Workbench	Interdisciplinary	TU München, Germany
QUT	Interdisciplinary	Queensland University of Technology, Australia
ARGOS	Interdisciplinary	OpenAIRE AMKE, EUDAT CDI, Europe
easyDMP	Interdisciplinary	EUDAT, Finland, Norway
NFDI4Plants Dataplan	Plant science	Eberhard Karls Universität Tübingen, Germany
DMPonline	Interdisciplinary	Digital Curation Centre, University of Edinburgh, United Kingdom
RDMO NFDI4Ing	Engineering	University and State Library Darmstadt, Germany
DMPTool	Interdisciplinary	California Digital Library, University of California, USA
Data Stewardship Wizard	Interdisciplinary	Czech Technical University, Dutch Techcentre for Life Sciences, Czech Republic, Netherlands

Based on the findings of 19 expert interviews and a subsequent discussion among the project partners, we identified requirement parameters for the evaluation of existing DMP tools. The parameters were grouped into main categories to show a more detailed view of the rating scores. We focused on the technical requirements in order to identify

DMP tools, which are easy to host and maintain to ensure their adaptability to the specific needs of researchers, institutions, and funders. Furthermore, a weight factor between zero (not relevant) and three (high priority) was assigned to every parameter. For this purpose, the weight factor was determined individually by each member of the research team, and afterwards the arithmetic mean was calculated.

The DMP tools were rated by two different researchers independently according to a fixed rating scheme from zero (poor) to ten (excellent). In a next step, we calculated the arithmetic mean for each requirement parameter. To calculate the final score, the score for each parameter was multiplied by the weight factor. Then, the sum of the rating scores was calculated per main category and for the total score. Furthermore, the percentage of DMP tools with a score greater equal five was calculated. To identify and compare the linear relationships between the tools, that indicates to which degree their scores in the different requirement parameters are alike, the Pearson correlation coefficients based on the scores of the requirement parameters were computed.

3 Results and Discussion

Table 2 shows the 31 identified requirement parameters, which are important for the easy hosting and maintenance of a DMP tool. The parameter 'text modules' is of high importance, since many researchers prefer pre-fabricated text passages, which are automatically generated by the DMP tool. Although such a text might need some refinement by the researchers, it can serve as a first draft of a DMP. The important aspect of machine-actionability [5]–[7] is represented by 'export/import of DMP in tool format' and 'various export formats'.

Figure 1 shows the rating scores of the evaluated DMP tools. The evaluated tools satisfied between 10 % and 87 % of the requirement parameters. 61 % covered at least half of the parameters. The highest total rating scores were attained by Data Stewardship Wizard (DSW) (703.5), DMPTool (615.5) and RDMO NFDI4Ing (549.5). In the main category 'basic functions', DSW also performed best (239.5) followed by DMPonline (220) and EasyDMP (205). The three best performing tools in terms of 'technical aspects' were again DSW (190), DMPTool (190) and RDMO NFDI4Ing (181). The most user-friendly ones were DSW (274), DMPTool (225.5) and RDMO QUT (220). Comparing the results of our study with those of Gajbe et al. [4], who analyzed 14 tools, there are quite different results in the evaluations. 72 % of the evaluated tools provided source code, compared to 64 % in Gajbe's study. The majority (79 %) of tools from Gajbe were open access, however our findings showed 100 % open access. 67 % of our analysed tools provided user-friendly guidance, compared to 86 % of Gajbe's results. Most of the tools evaluated by Gajbe (86 %) had a user guide, while our results could confirm this for only 44% of the tools. In Gajbe's study, 64 % provided an option to share the DMP with others, whereas our results showed 56 %. We found that 67% of the DMP tools provided more than one export format, compared to 57 % in Gajbe's results. Pre-formulated filterable answer options were supplied by 64 % of the tools in Gajbe's study, while our study resulted in 56 %. In both studies, all tools provided open text fields.

Concerning the relatedness, 85 % of the DMP tools showed a positive and 16 % a negative correlation. The Pearson correlation of the tools (figure 2) is highest for ezDMP and GFBio (0.9). We did not find such a strong correlation as Gajbe et al. for DMPonline and DMPTool (1), but only a correlation coefficient of 0.7. The correlations for the other tool pairs differ from Gajbe as well.

The differences between our and Gajbe's results might be explained by the different tools evaluated, although eight tools were the same. Furthermore, the tool properties have changed over the years between Gajbe's analysis (2020) and our study (2023). The individual way of evaluating the tools could have been divergent as well.

Table 2. Requirement parameters are grouped into main categories (dark gray) and subcategories (light gray). Priority 3 = high, priority 0 = not relevant. DSGVO - General Data Protection Regulation.

Parameter	Priority
1 BASIC FUNCTIONS	
1.1 Access	
1.1.1 Open access with login	3
1.1.2 Open access without login	3
1.1.3 Encryption	2
1.1.4 DSGVO compatibility	3
1.2 Storage and Export	
1.2.1 Saving	3
1.2.2 Export/import of DMP in tool format	3
1.2.3 Various export formats	3
1.3 Collaboration	
1.3.1 Share DMP with collaborators	3
1.3.2 Track changes	1
1.3.3 Commenting function	2
1.3.4 Control levels	2
2 TECHNICAL ASPECTS	
2.1 Editing	
2.1.1 Editor access (CMS with roles)	3
2.1.2 Modularity ('generic' and 'institution specific')	3
2.1.3 Frontend/backend access	2
2.1.4 Easy maintenance of content	3
2.1.5 Sustainability of the software (updates and development)	3
2.2 Transparency	
2.2.1 Open source	3
2.2.2 FAIRness	2
3 USER FRIENDLINESS	
3.1 Assistance	
3.1.1 User friendly guidance	3
3.1.2 Pre-formulated filterable answer options	3
3.1.3 Text modules	3
3.1.4 Text sections (short DMP)	3
3.1.5 Preview of text modules (what you see is what you get)	2
3.1.6 User guide	3
3.1.7 User feedback	2
3.2 Design/Structure	
3.2.1 Layout/usability	3
3.2.2 Progress	2
3.2.3 Breadcrumbs (navigation)	2
3.2.4 Highlighting unanswered questions	3
3.2.5 Skipping questions	3
3.2.6 Open text fields	3

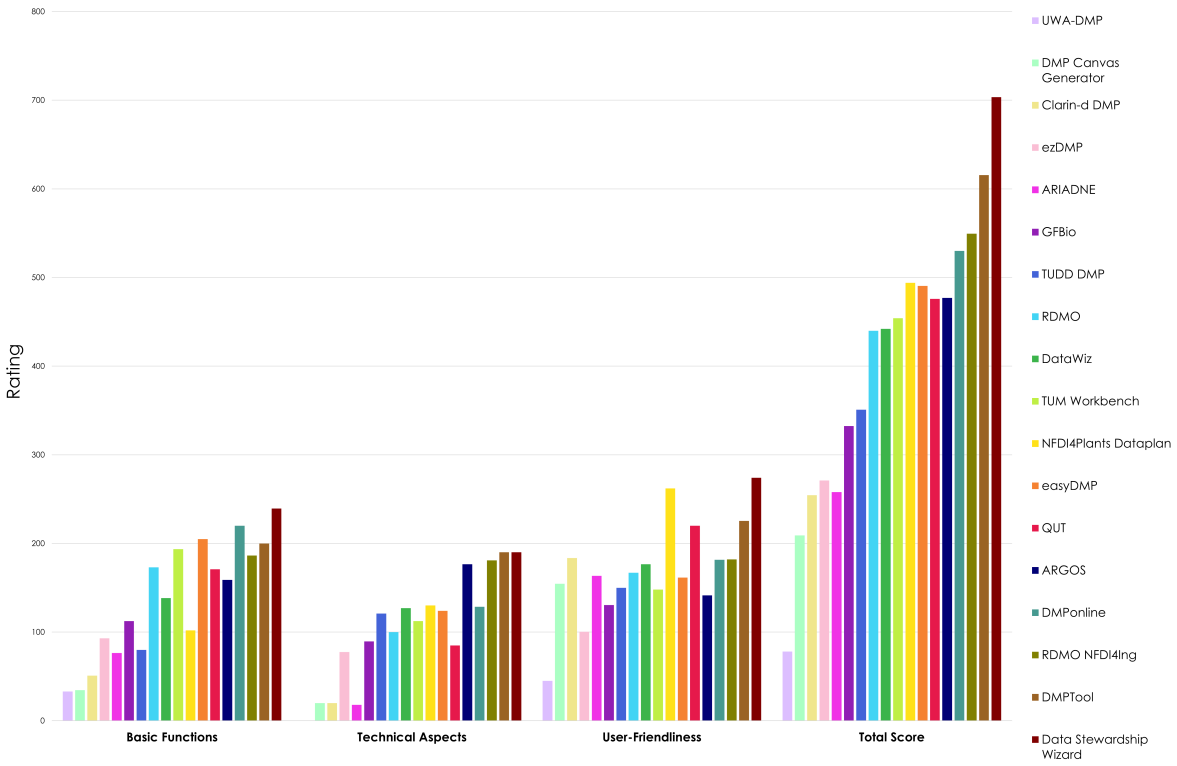

Figure 1. Evaluation results of 18 DMP tools. The rating scores are grouped by main categories (x-axis).

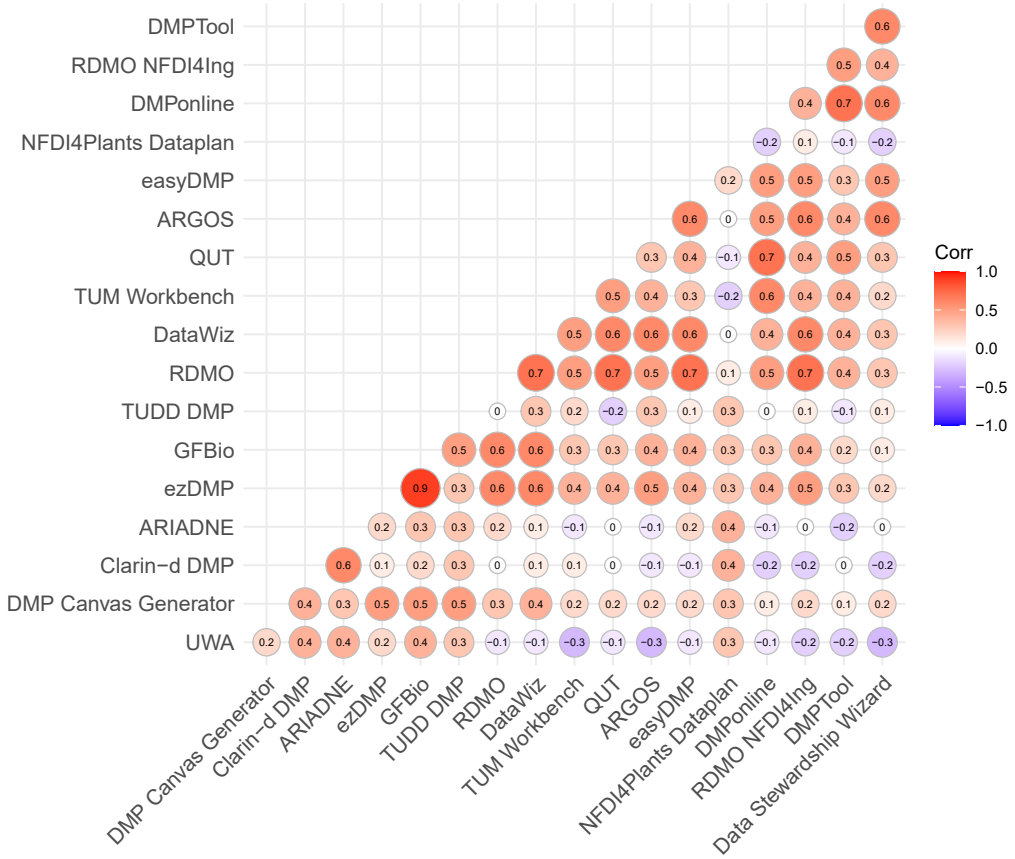

Figure 2. Correlation of 18 DMP tools based on the requirement parameters.

4 Conclusion

Our results show that Data Stewardship Wizard, DMPTool and RDMO NFDI4Ing are the highest rated DMP tools and can be recommended for researchers and institutions as flexible tools for hosting. In the light of recent developments in the area of DMP tools, this study provides an up-to-date evaluation of 18 DMP tools according to 31 parameters covering basic functions, DMP contents, technical aspects and user-friendliness. The results can support tool developers to identify potential improvements and hosting institutions to select a tool suited to their specific needs.

Data availability statement

The data supporting the results of our contribution can be accessed once the more extensive work is published in a peer reviewed journal.

Underlying and related material

There is no supplementary material.

Author contributions

Carina Becker: conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, validation, visualization, writing - original draft

Carolin Hundt: conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, validation, investigation, writing - review and editing

Claudia Engelhardt: investigation, writing - review and editing

Johannes Sperling: investigation

Moritz Kurzweil: investigation, writing - review and editing, funding acquisition

Ralph Müller-Pfefferkorn: conceptualization, investigation, funding acquisition, methodology, supervision

Competing interests

The authors declare that they have no competing interests.

Funding

This study is co-funded with taxes based on the budget passed by the Saxon state parliament (research proposal number 100607005).

References

[1] A. Ball, *Review of Data Management Lifecycle Models (version 1.0)*. Citeseer, 2012.

[2] W. K. Michener, "Ten simple rules for creating a good data management plan," *PLoS computational biology*, vol. 11, no. 10, e1004525, 2015. DOI: https://doi.org/10.1371/journal.pcbi.1004525.

[3] European Research Council and Scientific Council, *Open research data and data management plans - information for erc grantees*, https://erc.europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_Management_Plans.pdf, Accessed: 19.04.2023, 2022.

[4] S. B. Gajbe, A. Tiwari, R. K. Singh, *et al.*, "Evaluation and analysis of data management plan tools: A parametric approach," *Information Processing & Management*, vol. 58, no. 3, p. 102 480, 2021. DOI: https://doi.org/10.1016/j.ipm.2020.102480.

[5] T. Miksa, P. Walk, P. Neish, *et al.*, "Application profile for machine-actionable data management plans," *Data Science Journal*, vol. 20, no. 1, 2021. DOI: https://doi.org/10.5334/dsj-2021-032.

[6] T. Miksa, S. Oblasser, and A. Rauber, "Automating research data management using machine-actionable data management plans," *ACM Transactions on Management Information Systems (TMIS)*, vol. 13, no. 2, pp. 1–22, 2021. DOI: https://doi.org/10.1145/3490396.

[7] N.-M. Pham, H. Moulaison-Sandy, B. W. Bishop, and H. Gunderman, "Data management plans: Implications for automated analyses," *Data Science Journal*, vol. 22, no. 1, 2023. DOI: https://doi.org/10.5334/dsj-2023-002.