Check for updates

RESEARCH ARTICLE

REVISED Data Management Plans in Horizon 2020: what beneficiaries think and what we can learn from their experience

[version 2; peer review: 2 approved, 1 approved with reservations]

Daniel Spichtinger

University Library, University of Vienna, Vienna, 1010, Austria

v2 **First published:** 23 Apr 2021, **1**:42
https://doi.org/10.12688/openreseurope.13342.1
Latest published: 21 Feb 2022, **1**:42
https://doi.org/10.12688/openreseurope.13342.2

Abstract

Background: Data Management Plans (DMPs) are at the heart of many research funder requirements for data management and open data, including the EU's Framework Programme for Research and Innovation, Horizon 2020. This article provides a summary of the findings of the DMP Use Case study, conducted as part of OpenAIRE Advance.

Methods: As part of the study we created a vetted collection of over 800 Horizon 2020 DMPs. Primarily, however, we report the results of qualitative interviews and a quantitative survey on the experience of Horizon 2020 projects with DMPs.

Results & Conclusions: We find that a significant number of projects had to develop a DMP for the first time in the context of Horizon 2020, which points to the importance of funder requirements in spreading good data management practices. In total, 82% of survey respondents found DMPs useful or partially useful, beyond them being "just" an European Commission (EC) requirement. DMPs are most prominently developed within a project's Management Work Package. Templates were considered important, with 40% of respondents using the EC/European Research Council template. However, some argue for a more tailor-made approach. The most frequent source for support with DMPs were other project partners, but many beneficiaries did not receive any support at all. A number of survey respondents and interviewees therefore ask for a dedicated contact point at the EC, which could take the form of an EC Data Management Helpdesk, akin to the IP helpdesk. If DMPs are published, they are most often made available on the project website, which, however, is often taken offline after the project ends. There is therefore a need to further raise awareness on the importance of using repositories to ensure preservation and curation of DMPs. The study identifies IP and licensing arrangements for DMPs as promising areas for further research.

Open Peer Review

Approval Status ✓ ✓ ?

	1	2	3
version 2 (revision) 21 Feb 2022			
version 1 23 Apr 2021	✓ view	✓ view	? view

1. **Ron Dekker**, Consortium of Social Science Data Archives, Bergen, Norway
2. **Daniela Adele Hausen**, RWTH Aachen University, Aachen, Germany
   Giacomo Lanza, Physikalisch-Technische Bundesanstalt (PTB), Braunschweig, Germany
3. **Daniel Mietchen**, University of Virginia, Charlottesville, USA

Any reports and responses or comments on the article can be found at the end of the article.

Keywords
data management plans, data management, Horizon 2020, research data management

This article is included in the Research on Research gateway.

This article is included in the Horizon 2020 gateway.

Corresponding author: Daniel Spichtinger (daniel.spichtinger@univie.ac.at)

Author roles: Spichtinger D: Conceptualization, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Validation, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The "DMP Use Case" study and this article were financially supported by the European Union's Horizon 2020 research and innovation programme under the grant agreement No 777541 (project OpenAIRE-Advance). OpenAIRE-Advance continues the mission of OpenAIRE to support the Open Access/Open Data mandates in Europe. This publication reflects only the author's view and the Research Executive Agency and/or the Commission and/or OpenAIRE are not responsible for any use that may be made of the information it contains.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2022 Spichtinger D. This is an open access article distributed under the terms of the Creative Commons Attribution License , which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Spichtinger D. **Data Management Plans in Horizon 2020: what beneficiaries think and what we can learn from their experience [version 2; peer review: 2 approved, 1 approved with reservations]** Open Research Europe 2022, **1**:42 https://doi.org/10.12688/openreseurope.13342.2

First published: 23 Apr 2021, **1**:42 https://doi.org/10.12688/openreseurope.13342.1

REVISED **Amendments from Version 1**

Main changes to previous version (as per peer review comments)

- Added reference to opt-out options in Horizon 2020 open data requirements
- Clarification that this publication is primarily about study results but also feeds into larger literature
- Reference to underlying data for qualitative analysis
- More information on data source and methodology
- More information on screening process
- Revised pie chart
- Reference to new study on copyright provisions
- Revised Figure 2 (difficulty of partner feedback)
- Added that one stop shop could also train existing local staff as multipliers
- Expanded section on DMP licencing and recommendations for consortia to make DMPs available
- Reference to a number of organisations to which the study could be useful (suggestions from reviewer 3)
- Spelling corrections
- Restructuring of the presentation of the underlying data, including availability of more underlying data
- Update of acknowledgements

Any further responses from the reviewers can be found at the end of the article

Introduction

The importance of data and Horizon 2020 data management provisions

Data has been described as the 21st century's most valuable resource[1]. Issues related to data are also a priority of the von der Leyen Commission: to leverage the potential of data, the European Commission (EC) published a data strategy in February 2020[2], with the overall aim of creating a single market for the free flow of data within the EU and across sectors[3] and, as a first step towards implementation, the EC proposed a Regulation on European data governance (Data Governance Act) in late 2020. Despite these ambitions, however, discrepancies remain amongst EU member states as concerns the maturity of their data economies[4].

In a parallel and interlocked development, *research* data are also increasingly conceptualized as inherently valuable products of scientific research, rather than components of the research process that have no value in themselves[5]. Consequently, research funders both on the international and on the national level increasingly include requirements for data management (including openness). One of the trend-setters in this regard was the EC, which developed open research data and research data management requirements in its multiannual framework programme for research and innovation, Horizon 2020 (2014 to 2020). In Horizon 2020, the EC initially ran an open research data pilot scheme (ORD Pilot) in selected thematic areas which was subsequently extended to the whole of Horizon 2020 as of the work programme 2017 (under the principle of "as open as possible, as closed as necessary", i.e. allowing for opt-outs when justified)[6]. A key component is the obligation to create a Data Management Plan (DMP). In recent years, the objective to make data not only open but FAIR (findable, accessible, interoperable and reusable), has been gaining prominence as an important principle for data management and DMPs[7]. This trend will continue in the new Framework Programme Horizon Europe (2021–2027).

The DMP Use Case study goals

In the "DMP Use Case" study[a] we aimed to identify good practices but also common challenges amongst a number of DMP use cases across different disciplines. Our goal was to support researchers with their DMP obligations throughout their own European projects; furthermore, DMP analysis is an important resource for a variety of other purposes, for instance training activities but also for further scientific research into data management and re-use practices. The study was undertaken from April 2020 to February 2021 on behalf of the EU funded OpenAIRE-Advance consortium by the University of Vienna Library, with support from the OpenAIRE-Advance RDM Task Force. This publication primarily reports the findings of this project; it also feeds into literature about DMP good practice in a broader FAIR ecosystem and on making them machine actionable.

Methods

There were two main components to the study: a qualitative part, which consisted of an analysis of six DMPs and interviews with six cases studies, and a quantitative part, which involved a manual and automated screening process to establish a white list of DMPs, as well as a survey of the DMP experiences of H2020 projects. These components are described in more detail below.

Qualitative dimension

This part of the study first contained a qualitative evaluation of six Horizon 2020 DMPs (see *underlying data*). The evaluation was based on a modified version of the example rubric (CC0) from the DART project (https://osf.io/26b9r/), which was presented during the IDCC 2016 DMP Workshop in Amsterdam. This rubric was augmented by providing a score of 0 to 2 points for each of the 33 categories, resulting in a maximum of 66 points to be scored[b]. In a second step, an interview guide (*Extended data*[8]) was developed and six interviews with key personnel involved in DMPs were conducted[c]. For each interview a summary document (not a full transcript) was produced, additionally

[a]A description can also be found at https://bibliothek.univie.ac.at/en/openaire.html under the section "DMP Use Case Project".

[b]A limitation is that due to the available resources, the qualitative DMP assessment was only undertaken by one person. For future analysis, the four eyes principle is recommended.

[c]In two cases, interviews were refused or the relevant personnel did not come back to us. In this case, backup candidates were selected for the interview.

to the initial DMP assessment. The following selection criteria were used to identify the longlist and shortlist of projects to analyse and interview:

- *Balanced thematic representation:* although no European Research Council (ERC) DMPs were analyzed (see below), the ERC classification for its evaluation panels[9] was used as a convenient way to establish a thematic grouping. The aim was to have two DMPs represented from each ERC top level area classification, that is two DMPs from Social Sciences and Humanities (SH), two DMPs from Physical Sciences and Engineering (PE), and two DMPs from Life Sciences (LS). Within these areas, the research interests of the staff of the University of Vienna library were taken into account when making the specific selection.
- *Availability of more than one project specific DMP:* we also wanted to assess whether there have been major updates or progress during project duration, which is why we selected Horizon 2020 projects which submitted more than one version of the DMP.
- *Geographic balance:* we wanted to ensure that we analyse DMPs that reflect geographic balance. This was particularly important for the interview selection.
- *Gender balance:* this criterium was primarily of importance when selecting the interview candidates.

Taking into account these criteria, the potential participants were approached via email; the interviews took the form of video calls of about 30 minutes length each. Interviews were conducted with projects from the following specific ERC areas:

- *Social Sciences and Humanities (SH).* Education: systems and institutions, teaching and learning (SH 4_11); Linguistics: formal, cognitive, functional and computational linguistics (SH4_6).
- *Physical Sciences and Engineering (PE).* Web and information systems, database systems, information retrieval and digital libraries, data fusion PE 6_10 (2x).
- *Life Sciences (LS).* LS7_8 (Health services, health care research); LS7_2 Diagnostic tools (e.g. genetic, imaging).

Ethical considerations

Due to the low risk nature of the qualitative part of the study and the fact that it raised no significant ethical issues, approval from the University of Vienna's ethics committee was not required. Rather, approval of the survey was provided by the scientific supervisor. Each interviewee was contacted before the interview with information about the study and a consent form. Some participants returned the completed consent form. Where this was not done, participants were asked orally for their consent prior to the interview, which was provided in all cases. Consent obtained orally was noted in the internal summary document, which we created for each interview.

Quantitative dimension

The decision to include a quantitative dimension was based on the large amount of public DMPs that are available from CORDIS.

An initial list of public DMPs was obtained from the Commission (CORDA Support) on April 21, 2020. This was then supplemented by the University of Vienna's IT support by downloading the full list from https://cordis.europa.eu/data/cordis-h2020projects-xlsx.zip on May 18th, 2020. This data was enriched with further information about the projects (programme topics, funding schemes etc from the EU Data portal (data.europa.eu/euodp/de/data/dataset/cordisH2020projects and data.europa.eu/euodp/en/data/dataset/cordisref-data)

Because not all of these could be analyzed in this study, the decision was made to develop a curated collection for future use and re-use. In order to achieve this, a two-step process, combining manual and automated vetting of these DMPs was applied:

Stage 1 – manual vetting: with the help of volunteers from OpenAIRE's RDM taskforce the initial list of 1552 DMPs downloaded from CORDIS was manually screened according to the following questions:

- Is the document in question really a DMP? This screener was added in order to identify and remove documents that had been wrongly classified as a DMP in the system.
- Is the document really public? This screener was added in order to identify and remove documents that had been wrongly classified as public in the system but in fact were set to restricted and/or confidential by the beneficiary.

In total, 1053 DMPs passed the first stage of the screening: details about these projects , most notably acronym, full title, partners thematic area classification, start and end date, project website URL, project objective, budget, call and funding scheme are available in the supplementary data. For those DMPs that did not pass the first stage, the public nature of the DMP was not clear for 21%, to a significant extent ERC DMPs[d]. Furthermore, a minority of documents was listed as confidential (5%), followed by 3%, which were not DMPs as well as 2% of DMPs which technically could not be downloaded and another 2% which could not be classified. (see Figure 1).

Stage 2 – automatic vetting: an automatic search for the word "copyright" was conducted in each DMP document. Any DMPs that included copyright were excluded from the collection in order to ensure that only those DMPs which are not IP protected are published. It is clear that this processing is not very fine grained and an additional study is currently under way to look at copyright provisions in more detail.

[d]The ERC does not require its beneficiaries to develop deliverables and hence most ERC DMPs did not include a reference to the status of the document as being public or restricted access. Although a representative from the ERCEA did indicate that publication of these DMPs should in most cases not be problematic the decision was made to not release them at the time of writing.

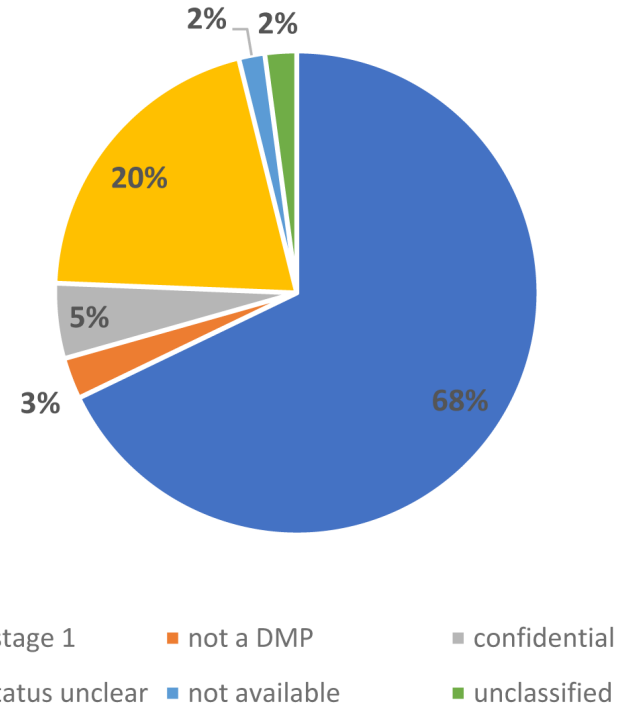


Figure 1. DMP screening stage 1.

This two-stage process resulted in a list of 840 DMPS which passed both stages. These DMPs are now publicly available for further use (e.g. analysis, training etc)[10]: https://phaidra.univie.ac.at/detail/o:1140797[11]. See *Underlying data* for full details of the final vetted list.

Based on the interview guide for the qualitative component (see *Extended data*[8]) a survey was developed and distributed to the contact persons indicated on the 840-white listed DMPs via Survey Monkey. This served to further enlarge and broaden the data collection. The survey was filled in by 108[e] projects - 87 of which provided the project acronym and 21 who filled in the survey anonymously. In total, 68 projects were still ongoing at the time of the survey, while 40 were completed (this compares with four completed and two ongoing projects in the qualitative interviews). At the beginning of the online questionnaire, participants were notified that if they proceed this indicates their consent to participate in the research.

Results

Knowledge about DMPs prior to the project

In the quantitative survey, 59 projects (54.63%) had been aware of DMPs before their Horizon 2020 project started. A significant minority (49 projects; 45.37%) were thus introduced to DMPs through the Horizon 2020 programme, pointing towards the influence of Horizon 2020 in introducing and spreading the practice of creating a DMP. This is also corroborated by the qualitative interviews: in most of the interviews the participants had been (sometimes vaguely) aware of data management before their Horizon 2020 project, but in several cases the Horizon 2020 project was the first time they actually had to write a DMP. In one case, a project did not in fact participate in the H2020 ORD pilot but still volunteered to do a DMP because they thought it would contribute to a positive evaluation of their proposal. Several interview partners indicated that since their initial involvement their knowledge about data management and DMPs has increased significantly.

Which work package was/is the DMP part of?

From a project management perspective, anecdotal evidence suggested that the DMP has been integrated into different parts (Work Packages or WPs) of Horizon 2020 projects, notably the general project management WP or the dissemination WP. In most of the qualitative interviews, the DMP been dealt with in the management WP. Given that 51.4% of surveyed projects consider this the appropriate setting for the DMP (see Table 1), there is an indication that this is becoming more standard practice, as DMPs are becoming more widespread. This said, in one project, data management was formally part of the management WP, but informally spread over three WPs

[e]In the results section below the number of respondents varies slightly, since questions could be skipped.

related to data acquisition and analysis. In another case DM was split between the management and the dissemination WP. In the quantitative survey, the second most popular work package for the DMP was dissemination (21.5%), followed by a distinct WP solely for data management (17.76%). The latter figure is surprisingly high and not corroborated by the qualitative interviews, in which not a single project had a dedicated DM WP.

Table 1. Which work package was/is the DMP part of?

Answer choices	Responses	
part of the project management work package	51.40%	55
part of the dissemination work package	21.50%	23
own work package for data management	17.76%	19
other	9.35%	10
	Answered	107
	Skipped	1

Feedback from partners

There were a lot of different answers concerning the ease or difficulty of obtaining feedback from project partners in the process of creating and/or updating the DMP, which is perhaps not surprising given the different size and composition of H2020 projects. In the quantitative survey, only a tiny minority thought this was very easy or very difficult, with most responses on the scale of a difficulty of 5–8 (where 1 is very easy and 10 very difficult) see Figure 2. In the qualitative interviews, challenges encountered with other partners concerned:

(i) personal data and GDPR,

(ii) the amount of time and resources needed and

(iii) coordination among geographically distant partners (though this is not necessarily limited to DM)

One interviewee also stated that ease and quality of feedback depended on the type of data. In the same project there was at least one person per partner involved in data management. In one project it was decided to sign user agreements, with the data belonging to those users but the project having limited usage rights. One participant also stressed that data management does not necessarily mean open: only some data in their project was opened for scientific conferences (e.g. deposited on Zenodo).

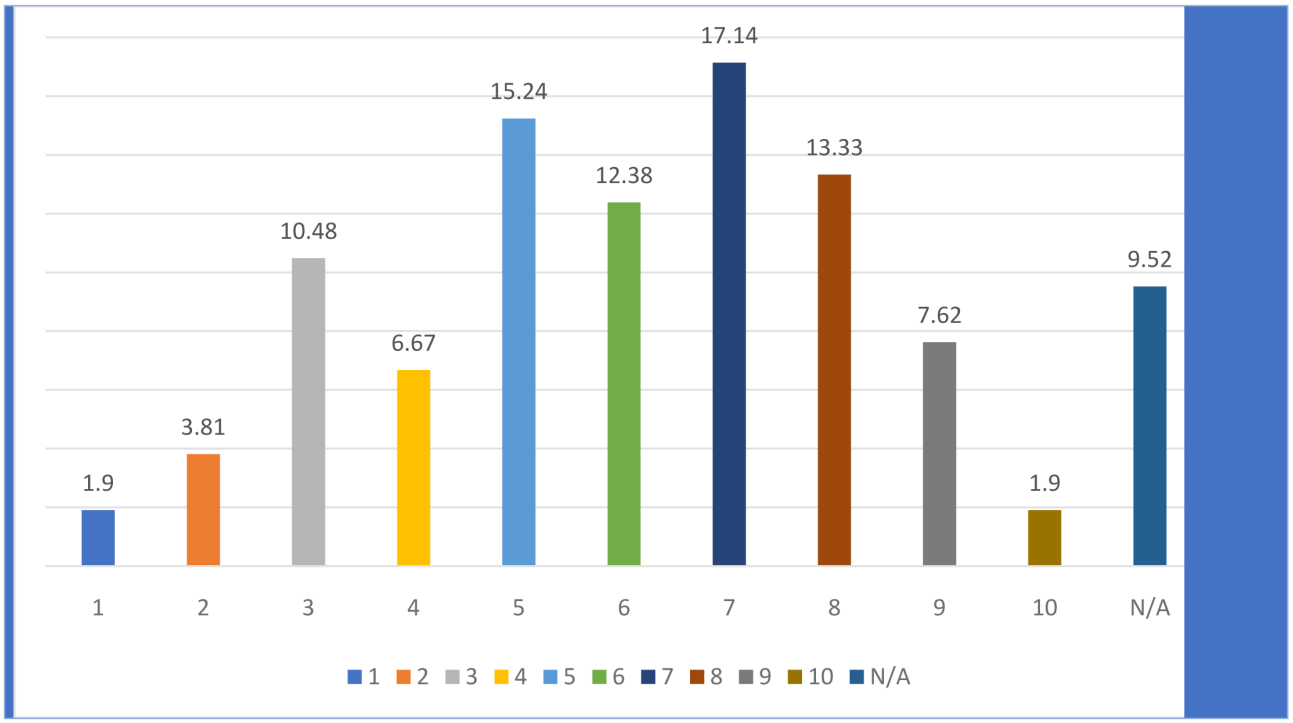


Figure 2. How easy/difficult was it to obtain feedback from the partners for the DMP? (1-very easy, 10 very difficult), numbers in %. *N/A = did not consult with partners N= 105.*

Use of templates & online tools

In the quantitative survey, 40% of the respondents (42 projects) indicated that they used the templates from the EC or the ERC, while 17.14% used another template and 8.57% used a digital tool (see Table 2). From the comments it became apparent that the most often used external tool is DMP online[f] from the Digital Curation Center. In total, 25% of respondents did not use a specific tool or template at all but developed their own template. In fact, in the qualitative interviews several interviewees stated that at the time of their project start the EC template was not yet available. Therefore, these projects did their own - sometimes extensive - background research or based their DMP on previous knowledge. Some also partially used the EC template and augmented it with information from other sources and their communities. In one case the library was involved in assisting with the DMP. In another case it was reported that each partner contributed to their own part of the DMP, while in a more recent project of the same partner this was changed to a more unified approach.

Table 2. Did you use a template or online tool when creating the data management plan for your project?

Answer Choices	Responses	
YES- template from the European Commission/European Research Council	40.00%	42
YES - template from another organisation	17.14%	18
YES - online tool	8.57%	9
NO	25.71%	27
Not applicable	8.57%	9
If you used a template or online tool from another organisation please specify which template or tool?		26
	Answered	105
	Skipped	3

Support

When asked whether they received support when creating their DMP, the majority of those surveyed indicated that they received support from other partners (39%), and 27% indicated that they did not receive any support. For the rest, several projects (11.5%) indicated that they received support from the library, 3.8% from OpenAIRE and 2.9% from the IT department (see Table 3). A more fine-grained picture emerges from the qualitative interviews: here one partner mentioned support from the university's data protection officer and technical input as regards data security. OpenAIRE was also explicitly mentioned several times, as was the advantage of having partners with experience in this area. The library and a data archive was also mentioned in one interview. It was stressed that it would be helpful to have a designated contact at the EC for inquiries.

[f]https://dmponline.dcc.ac.uk/

Table 3. Did you receive support when creating your data management plan?

Answer Choices	Responses	
YES- from other partners	39.42%	41
YES-from the library	11.54%	12
YES-from the IT departement	2.88%	3
YES- from OpenAIRE	3.85%	4
NO	26.92%	28
YES Other - please specify	15.38%	16
	Answered	104
	Skipped	4

Feedback from the EC/the Agency

A total of 55.8% of survey respondents did not receive feedback from the EC or the implementing agency, while 22.1% received feedback from the Project Officer, and 22.1% received feedback from the reviewers (see Table 4). The comments indicate that those that did receive feedback largely considered it useful for their further work. Within the qualitative interviews, none of the six interviewees received any content related feedback from the Commission or Agency but several received feedback from their reviewers. In one case it was mentioned that the Commission itself seemed unsure on how to handle this deliverable.

Table 4. Was there feedback from the European Commission/the Agency you submitted the plan to?

Answer Choices	Responses	
YES from the Project Officer	22.12%	23
YES from the Reviewer(s)	22.12%	23
NO	55.77%	58
	Answered	104
	Skipped	4

DMP project specificities (qualitative interviews only)

Within the qualitative interviews we were able to delve deeper into some of the project specific DMP related issues the interviewees encountered:

- *GDPR compliance:* with older projects this was initially not an issue but became relevant once the GDPR entered into force. One early DMP dealing with privacy issues

included a privacy impact assessment; the data in this DMP was not in fact open: "the objective was to be accountable, not open".

- One project was concerned with vulnerable groups and therefore has a strong focus on personal data consent forms, data security and ethical issues in their DMP.
- One DMP did not consider Creative Commons very useful – the data was not considered an original work in the sense of the German word "Urheberrecht".
- One project primarily used pre-existing open data. There was therefore no problem in using an open license – however, the business partners in the project were somewhat critical and saw open data more as an obstacle, rather than as an opportunity.
- One project illustrates the progressive evolution of DMPs from one version to the next, with some questions only being able to be answered in the final iteration of the document (while in other project DMPs there is little change over time)
- One project explicitly mentions the lack of community standards as a major barrier.

Major challenges (qualitative interviews only)

The following major challenges were raised by the interviewees in the qualitative interviews:

- reading and analyzing partner input and turning it into one understandable document, in particular at the beginning of the project, when there was little experience
- where to put the focus and how much details to give – internal procedures or output; also whether to tackle any data or data underlying publications (the latter strongly preferred)
- understanding the technicalities
- how to create the DMP from scratch with zero experience
- understanding the requirements and convincing partners to submit thorough information (done through peer pressure). This is easier in newer projects since DMPs are more accepted
- covering all partners, some of them in non-EU countries where different national policies apply (e.g. on protecting vulnerable groups)

Usefulness of the DMP, beyond it being an EC requirement

Anecdotal evidence suggested that DMPs may be considered a tick boxing exercise or an unnecessary burden by at least some of the Horizon 2020 beneficiaries. It was therefore somewhat surprising that the survey respondents as well as the interviewees did not share this view[g]. A total of 53.3% considered the DMP useful beyond it being a EC requirement and an additional 29% considered it somewhat useful, resulting in 82.2% with a generally positive attitude. Only 17.8% did not consider a DMP useful (see Table 5). There were, however, a number of diverging views in the comments. This rather positive view, with some caveats, was also present in the interviews, where the interview partners had the following to say on the usefulness of the DMP beyond it being an EC requirement:

- We turned something that was initially a chore into a Socrative work and learned a lot from it
- Not every project needs a DMP; a simple checklist would suffice
- It's challenging having to write a DMP through a pre-existing template because you need to fit your project to pre-existing guidelines. At the same time, it is also useful because during the work one can lose sight of FAIR data and the exercise reminds you to remain on track. A more advanced project might benefit from having its own template though-
- Very important to be done for each project (regardless of EU funding) but needs different approaches and categories based on the size and the nature of the project (currently not much of a distinction whether it is a project with 500 partners with a lot of shared data or 5 "friends and family")
- The only thing it was useful for was to clarify in project meetings which datasets we were talking about. For the overall objective of the project a DMP was not very important
- DMPs are very useful, also for projects which deal with vulnerable and marginalized groups and long-term curation and preservation. We are switching from destroying data to archiving data after the project end.

Table 5. Do you consider the development of a data management plan useful beyond it being a requirement from the side of the European Commission?

Answer Choices	Responses	
Yes	53.27%	57
Partially	28.97%	31
No	17.76%	19
if you chose "partially", can you specify?		31
	Answered	107
	Skipped	1

[g]Both survey respondents and interviewees were not selected from a data management "in-group". However, they were taken from the sample of projects that delivered a public DMP (see Methods section).

Publication of the DMP

Interestingly, most of the qualitative interviewees were initially not sure whether their DMP had been published somewhere (except it being submitted to the EC). After checking, many stated that it had been published on the project website, which, however, was no longer online in some cases. In one instance the project was contacted by OpenAIRE and uploaded the DMP to OpenAIRE (alongside the other project deliverables). In the survey, the project website was also the most popular location for publishing the DMP (30.5%); however, the majority of projects did not publish their DMP at all (38.1%). Only 22.9% deposited their DMP in a repository (Table 6).

Table 6. Did you publish your data management plan somewhere?

Answer Choices	Responses	
YES - in a repository	22.86%	**24**
YES - on the project website	30.48%	**32**
YES - somewhere else	8.57%	**9**
NO	38.10%	**40**
If you chose the answer "yes - somewhere else", please specify where		**13**
	Answered	**105**
	Skipped	**3**

Time and resources

Both the survey respondents and the interviewees struggled with this question and a number of different answers were given. One user pointed to the fact that the existence of a template has made it easier and less resource intensive to create a DMP. Other good advice shared related to spending time on data management in the planning phase, which makes it easier to implement once the project has started ("something well planned is half done"). The normalization and routinization of DMPs also makes it less resource intensive.

As to the time and effort needed, no uniform answer emerged, which is perhaps not surprising given the different size of Horizon 2020 projects and the different thematic areas covered. The same holds true not only for data management plans but also for data management more generally.

What kind of support is needed and who should provide it?

In contrast to an earlier question that was designed to elicit where support *currently* comes from, this question was designed to elicit where it *should* come from and who should deliver such support. In the qualitative interview the following issues were raised:

- Reference contact in the Commission to provide training and advice
- A support paper which contains the requirements from the EU as concretely as possible – a matrix then just needs to be applied. Research support organizations should execute that, single researcher should have an overview and operational support
- Sustainability questions are important, including how to pay for data management after the end of the project; what are the limits to make data FAIR but at the same time sustainable. Larger infrastructures (ERICs) can help
- The best approach is to have someone in the data community with expertise to help; the data community should be more approachable for everyone
- For bigger organizations the library can provide support (and sometimes also the data archive)

Final thoughts from the interviewees

The interviewees were provided with the opportunity to flag up any other issues they would like to mention in the context of DMPs and data management. The following aspects were mentioned in the qualitative interviews:

- The need for awareness raising (in particular as concerns legal regulations for personal data)
- Taking issues connected with AI into account (very new and not always included in DMPs)
- Zenodo is a useful tool and collaboration with OpenAIRE works well – we need sufficient political will to continue that
- When project ends data tends to disappear, people save data in different repositories which make it very dispersed – the best solution would be to have one repository, although a monopoly can also pose problems. We may need a global agreement to releasing open data (COVID could be an opportunity)
- We often fail at longevity both as concerns tools and repositories (will they still be here in 5 years?) – they don't always allow you to take data out in accessible formats in an easy way
- Templates are very useful; it helps to think about data collection but also use (even after project end)

Discussion and recommendations

For a significant number of projects, Horizon 2020 was the first time they had to develop a DMP. This underlines the importance of the Commission's policy and its impact in using the framework programme to promote research data management practices. In a wider context, it generally illustrates the importance of funders to provide clear and ambitious open science requirements in their programmes. Results from the qualitative interviews also indicated that the Horizon 2020 DMP development process has been a learning journey for many of the interviewees: several indicated that they have significantly

developed their knowledge on data management, underlining the fast development of data management practices in recent years. Importantly, 82% of survey respondents see data management as useful or partially useful beyond it being just an EC requirement.

As regards project management, having a DMP as part of the WP on management (as opposed to e.g. dissemination) seems to be the most widespread practice, in particular for those projects which do not have data science as their focus. We would therefore generally recommend projects to follow this approach, if there are no good reasons to do otherwise, but at the same time to also ensure links with dissemination activities. In general, having one person per project partner responsible for data issues is a good practice (except for small projects or coordination and support actions where no data is generated). However, there also needs to be a person that takes overall responsibility for the project DMP, so that it is not simply a collection of input delivered by partners but forms a coherent whole.

Templates are clearly important: 40% of the survey participants used the EC and ERC template and these templates seems to have helped to dispel some confusion at the beginning of the H2020 ORD pilot when there was little information available and projects had to do an enormous amount of research themselves. However, some ask for a more tailor-made approach, since one DMP template may not fit all the different kinds of projects funded under Horizon. This could be done through an EC online system for creating DMPs (and not just a pdf template) or through the further development of existing tools such as ARGOS[h], which could then be endorsed by the Commission. For details about potential improvements of the template the reader is also referred to the specific OpenAIRE/FAIR Data Expert Group report on the Horizon 2020 template[12].

Support in creating DMPs was in most cases received through the project partners (if at all); in some cases the library or a data archive or OpenAIRE were also mentioned as sources of support. In the qualitative interviews, none of the participants received content feedback on the DMP from the Commission or Agency but some did receive feedback from the reviewers. Similarly, in the qualitative survey, the majority of respondents (55%) did not obtain feedback from either, but those that did, found it helpful. Especially beginners report a feeling of being lost and, in particular before the template was available, had to do a significant amount of self-learning (qualitative interviews). A number of interviewees ask for a dedicated contact at the EC to help with the data management plan. I would therefore recommend to set up a "EU one-stop-shop for Horizon research data management", akin to the IP helpdesk (e.g. through a public procurement procedure or a grant to a named beneficiary). OpenAIRE would be an institution with a lot of knowhow to run such a one stop shop, potentially as part of the European Open Science Cloud (EOSC), which OpenAIRE is an important part of. This should also include further guidance on resources and costing[13]. This one stop-shop could also train existing local staff as multipliers to support researchers.

Both respondents in the qualitative interviews and the quantitative survey point towards the project website as the main place where they published their DMP. This is somewhat problematic, since project websites tend to be shut down after project end and thus the DMPs (and other deliverables) are not available for long term preservation and curation. On the one hand this points to the importance of CORDIS/CORDA as a source for public DMPs (and other public deliverables). However, there is also a need to further raise awareness of the need to deposit DMPs in repositories to ensure preservation. The deposition process can be combined with assigning a persistent identifier to a DMP, which improves their findability. Moreover, there is also a need to ensure that DMPs themselves are clearly licenced, preferably with a creative commons licence. As Horizon DMPs are classified as deliverables, it is currently up to the consortium whether DMPs are provided open access or not. Projects should be encouraged to apply the "as open as possible, as closed as necessary" principle to DMPs as well.

This study has also identified a number of areas for further research. As concerns the establishment of the DMP white list, a number of DMPs were not included in the list due to their mentioning "copyright" (see Methods section). DMPs would merit closer analysis on what exactly licensing arrangements and any copyright restrictions are - in some cases first evidence indicates that this can be confused, Furthermore, an analysis of ERC DMPs - not included in the curated collection due to the fact that they do not mention whether they are public or not - would also be interesting.

The results reported in this publication and the data underlying it could be useful for the ongoing work of the research data management community, such as the "Exposing Data Management Plans WG" and/or the Active Data Management Plans IG" of the Research Data Alliance (RDA).[i] Furthermore, this study also contributes evidence important for funders and policy makers: as more and more funder mandates for data management move from aspirational to hard requirements[14], it will be important to monitor in detail not only the uptake but also the experience of the project beneficiaries with current and new requirements. These funder requirements have real live impact on what researchers are required to do with underlying research data and this in turn also has implications for those institutions and organisations that take care of research data. However, rather than seeing funder requirements as an administrative burden they should be considered a strong incentive

[h]https://argos.openaire.eu/splash/

[i]I am grateful to reviewer 3 for this suggestion.

for change towards a better managed and more transparent scientific system.

Data availability

Underlying data

PHAIDRA: Horizon 2020 DMPs what beneficiaries think and what we can learn from their experience ("DMP Use Case Project"), https://phaidra.univie.ac.at/detail/o:1165751[8]. Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0)

This project contains the following underlying data:

a) Individual DMP analysis

- DMP analysis EURHISFIRM (Data management plan analysis of the CORDIS project: EURHISFIRM)
- DMP analysis Edu MAP (Data management plan analysis of the CORDIS project: Edu MAP)
- DMP analysis FREME (Data management Plan analysis of the CORDIS project: FREME)
- DMP analysis READ (Data management Plan analysis of the CORDIS project: READ)
- DMP analysis AFRI ALLIANCE (Data management Plan analysis of the CORDIS project: AFRI ALLIANCE)
- DMP analysis Apollo (Data management Plan analysis of the CORDIS project: Apollo)
- DMP analysis CAREGIVERSPRO-MMD (Data management Plan analysis of the CORDIS project: CAREGIVERSPRO-MMD)
- Evaluation rubric for DMP analysis
- Responses summary (Results of quantitative survey of "DMP Use Case Project")

b) Interview Guide

c) Collection of vetted DMP

- DMP Case Study Stage 1 Manual Vetting Project DMP List
- DMP Case Study Stage 2 Copyright Vetting Final DMP List
- Individual DMPs at PHAIDRA: DMP Use Case Project, https://phaidra.univie.ac.at/o:1140797[11] -The DMPs available in the collection are clearly marked as public documents (designation PU) and not copyrighted (as far as the latter information could be verified through automation). The CORDIS license allows re-use (see https//cordis.europa.eu/about/legal). Nevertheless, users of the collections are advised to consult the individual DMPs for any restrictions in re-use; PHAIDRA and OpenAIRE Austria accept no liability. The *content* of the DMPs has not been quality reviewed – they are published as is and should not necessarily be taken as good practice cases.

Acknowledgements

I wish to thank the volunteers of the OpenAIRE Advance RDM taskforce and the University of Vienna Library IT Service for their support in establishing the DMP white list. I also wish to thank my qualitative interview partners and those that filled in the quantitative survey as well as my supervisor at the University of Vienna Library Gerda McNeill. I also wish to thank the University of Vienna Library for additional funding to revise this article based on the peer review comments (which were obtained after project funding ended). Finally, my thanks also go to the reviewers for their thorough work and insightful suggestions.

References

1. The Economist: **Regulating the internet giants. The world's most valuable resource is no longer oil, but data**. 2017.
Reference Source
2. European Commission: **A European strategy for data.** COM(2020) 66 final. 2020a.
Reference Source
3. European Commission: **Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on European data governance (Data Governance Act).** COM/2020/767 final. 2020b.
Reference Source
4. European Data Portal: **Open Data Maturity in Europe. Report 2020.** European Data Portal/CapGemini *et al.* 2020.
Reference Source
5. Leonelli S: **Why the Current Insistence on Open Access to Scientific Data? Big Data, Knowledge Production, and the Political Economy of Contemporary Biology.** *Bull Sci Technol Soc.* 2013; **33**(1–2): 6–11.
Publisher Full Text
6. Spichtinger D, Siren J: **The Development of Research Data Management Policies in Horizon 2020.** In *Research Data Management - A European Perspective.* Berlin/Boston: De Gruyter SAUR, 2018; 11–23.
Publisher Full Text
7. European Commission Horizon 2020 Online Manual: **Data Management**.
Reference Source
8. OpenAIRE Austria: **Horizon 2020 DMPs what beneficiaries think and what we can learn from their experience ("DMP Use Case Project").** *PHAIDRA.* [dataset].
https://phaidra.univie.ac.at/o:1165751
9. European Research Council: **ANNEX 1: ERC PEER REVIEW EVALUATION PANELS (ERC PANELS)**.
Reference Source
10. OpenAIRE Blog: **Establishing a collection of 841 publicly available Horizon 2020 Data Management Plans**. 2021.
Reference Source
11. OpenAIRE Austria: **DMP Use Case Project.** *PHAIDRA.* [dataset].
https://phaidra.univie.ac.at/o:1140797
12. Grootveld M, Leenarts E, Jones S, *et al.*: **OpenAIRE and FAIR Data Expert Group survey about Horizon 2020 template for Data Management Plans.** (Version 1.0.0) [Data set]. *Zenodo.* 2018.
Publisher Full Text
13. OpenAIRE: **Guides for Researchers How to identify and assess Research Data Management (RDM) costs**.
Reference Source
14. Neylon C: **Compliance Culture or Culture Change? The role of funders in improving data management and sharing practice amongst researchers.** *Res Ideas Outcomes.* 2017; **3**: e14673.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔ ?

Version 1

Reviewer Report 25 August 2021

https://doi.org/10.21956/openreseurope.14412.r27181

© **2021 Mietchen D.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Daniel Mietchen
School of Data Science, University of Virginia, Charlottesville, VA, USA

General comments
This study provides a useful introduction to practical aspects of Data Management Plans (DMPs) in the context of Horizon 2020-funded projects and presents an analysis of associated trends with respect to data management more broadly, particularly in the context of Horizon Europe. While the overall conceptualization is solid, the description of the methodology would benefit from greater detail (e.g. sharing of the code involved), as would the discussion of policy implications for key stakeholders like projects, research fields, institutions or the European Commission.

I read the reviews by Ron Dekker[1] and by Daniela Adele Hausen & Giacomo Lanza[2], who expressed much of what my comments would be. In the following, I will thus concentrate on aspects of the paper that they did not discuss in detail, or where I would put another emphasis.

The present work uses references only scarcely but would benefit from better contextualization, since many aspects of it overlap with work on making DMPs better in general[3], on integrating them with a broader FAIR ecosystem[4] and on making them machine actionable[5].

Specific comments
The study is based on two datasets (cited as references 8 and 11) that have been created and shared in the framework of the project. For the dataset in ref. 11, some methodological details are missing, e.g. the precise source of the DMPs (e.g. a query URL) or the set of instructions (e.g. an API call) used to retrieve and filter the DMPs represented in the collection. Likewise, no version information or URI was given for the H2020 DMP templates (of the European Commission or European Research Council) being discussed.

Given the "[u]sefulness of the DMP, beyond it being an EC requirement", it would have been instructive to read the DMP for the DMP Use Case Project itself.

As stated in footnote b, following the four-eyes principle would indeed have been a good choice. There are other quality control mechanisms that could have been considered, e.g. the study

explicitly focused on beneficiaries of funded proposals, whereas inclusion of DMP-related comments by grant reviewers or by applicants whose proposals were rejected might have added some new insights.

The filtering of non-ERC DMPs by simple mention of the string "copyright" seems very crude, both because of the overlap with Creative Commons licenses (as alluded to in the discussion), of which some would be compatible with the kind of aggregation and sharing described here, and because some DMPs might express their copyright status using other means (e.g. the copyright sign) or not at all. Because of this, the goal of the filtering - "ensure that only those DMPs which are not IP protected are published" - was likely not achieved with much precision.

The discussion of these copyright issues (and the uncertainty around the copyright status of ERC DMPs) does highlight the need for stating the copyright and licensing terms in a FAIR manner (for DMPs and other documents, ERC or not), but policy implications like this are only occasionally being touched upon in the present work, even though they would be a useful addition.

Licensing-related confusion also surfaced in other areas of the text, e.g. in the comment "One project primarily used pre-existing open data. There was therefore no problem in using an open license – however, the business partners in the project were somewhat critical and saw open data more as an obstacle, rather than as an opportunity." Without further context, this is ambiguous at best and needs unpacking - if they "used pre-existing open data", it is hard to understand how they could "[see] open data more as an obstacle, rather than as an opportunity".

As pointed out by Ron Dekker, some descriptive statistics of the DMP corpus would have been useful. In addition, the presentation of the statistics that are given merits some brushing - is the number of significant digits in the tables really appropriate?

The need for better contextualization was already pointed out in the general comments, and while this happens on occasion (e.g. the mention of "OpenAIRE/FAIR Data Expert Group report on the Horizon 2020 template"), here are some more specific examples: the section "Publication of the DMP" could point to the "Exposing Data Management Plans WG" of the Research Data Alliance, whose "Active Data Management Plans IG" would also be relevant for considerations regarding the "Availability of more than one project specific DMP".

The section "What kind of support is needed and who should provide it?" is focused very much on the DMPs as an end in and of themselves, but misses thoughts regarding how data management plans can actually be leveraged for improved data management, and how the documents containing such plans can be made more readily reusable.

Hausen and Lanza wrote: "It would be helpful for the reading flow to put the ethical considerations in footnote". I agree that the reading flow is interrupted by having the "Ethical considerations" section between the "Qualitative dimension" and "Quantitative dimension" sections but think the footnotes - at least as implemented in the journal - interrupt the flow even more, and a simple reordering of the sections would suffice.

Personally, I would like the ethical considerations to be expressed in a more explicit and more FAIR way (e.g. as per https://doi.org/10.5281/zenodo.2559998), but that is clearly beyond the scope of the present paper.

Linguistic matters

- There are a number of typos and other issues that need copyediting, e.g. "six cases studies" and "the DMP been dealt with"
- "with was established" (in Ref. 11)
- "creative commons" should be capitalized

References

1. Dekker R: Peer Review Report For: Data Management Plans in Horizon 2020: what beneficiaries think and what we can learn from their experience [version 1; peer review: 2 approved]. *Open Research Europe*. 2021; **1**. Publisher Full Text
2. Hausen DA, Lanza G: Peer Review Report For: Data Management Plans in Horizon 2020: what beneficiaries think and what we can learn from their experience [version 1; peer review: 2 approved]. *Open Research Europe*. 2021; **1**. Publisher Full Text
3. Michener WK: Ten Simple Rules for Creating a Good Data Management Plan.*PLoS Comput Biol*. 2015; **11** (10): e1004525 PubMed Abstract | Publisher Full Text
4. Turning FAIR into reality. Final report and action plan from the European Commission expert group on FAIR data. 2018. Publisher Full Text
5. Miksa T, Simms S, Mietchen D, Jones S: Ten principles for machine-actionable data management plans.*PLoS Comput Biol*. **15** (3): e1006750 PubMed Abstract | Publisher Full Text

Is the work clearly and accurately presented and does it engage with the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

Are all the source data and materials underlying the results available?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: My focus is on the integration between web-based research workflows and research policy, particularly around open science, data management and FAIR data.

I confirm that I have read this submission and believe that I have an appropriate level of

expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 20 Oct 2021

Daniel Spichtinger

Thank you for your thorough review and for already taking into account the comments from reviewer 1 and 2. I am grateful for the additional suggestions of relevant literature which I will gladly include.

Generally, the article focuses on this specific project and is not intended as a general discussion of good research data management practice.

I will get in touch with Uni Wien IT for the technical details requested (query URL, API), since they were in charge of the download. I will add links to the DMP templates.

As for the inclusion of non-funded proposals and reviewer comments these were not within the scope of the study and would have required significant additional resources (finding out which proposals were rejected); obtaining Evaluation Summary Reports from grant reviewers would also be difficult, since this is not public information. However, these are valuable suggestions for future research.

Copyright had to be considered, *inter alia to* avoid potential lawsuits of the University of Vienna, in whose repository the list of DMPs was made accessible. However, the reviewer is correct that a mere filtering according to the mention of copyright was indeed a blunt instrument, used also due to the limited resources available. This is why follow up research in a new project (results to be published) has focused on the use of creative commons licences. Future work on copyright related aspect in DMPs is certainly a promising avenue.

As to the comment on a project using open data but industry partners still seeing open data as an obstacle, this is indeed ambiguous and shows the different approaches and interest even within one and the same project.

As stated in response to the previous reviewers, statistical information about the DMPs will be added in a separate supplementary file. I will add the policy relevant activities the reviewer mentions.

As stated in response to a previous reviewer the ethics section is implemented according to the guidelines of ORE.

Copyediting will be addressed.

Competing Interests: No competing interests were disclosed.

Reviewer Report 28 July 2021

https://doi.org/10.21956/openreseurope.14412.r27179

© 2021 Hausen D et al. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Daniela Adele Hausen
RWTH Aachen University, Aachen, Germany
Giacomo Lanza
Physikalisch-Technische Bundesanstalt (PTB), Braunschweig, Germany

The article "Data Management Plans in Horizon 2020: what beneficiaries think and what we can learn from their experience" provides a good preparation and evaluation of the quantitative and qualitative methodology. Its content is very well designed and written in an understandable way. With this contribution, a current topic has been chosen that is well received both in the RDM community and among researchers from all disciplines.

Nevertheless, we have some suggestions for improving the structure and comprehensibility:

- General:
  - Quantitative information is often reported with nonprecise wording such as “most”, “several”, especially when it comes to the interviews. A quantification of the statements (“5 out of 6”) would be desirable and helpful for the content.
  - Tables are never linked in the text; instead, they should be referred to during presentation of the results.
  - Overall, it should be checked whether a table or even a diagram underlines better the content. From our point of view, it would be very well possible, for example, to present table 2 as a bar chart.
- Page 3, “Introduction”:
  - Openness of research data is a very sensitive topic and is interpreted differently in different disciplines or is not even possible in some cases. Therefore, there are opt-out reasons in the EU and many other funders deal vague with the term openness, so I think openness is formulated too generally.
  - Data FAIRness is a more appropriate request than openness, as it cushions precisely the difficulty of some disciplines just mentioned. Openness can thus emerge in the disciplinary environment and be adapted to the circumstances.
- Page 3, “Qualitative dimension”:
  - A short explanation (2-3 sentences) of the project and the background to the project and the rubric would be very helpful.
  - Concerning the metric from the DART project, please insert the direct link to the presentation / publication and possibly to the relevant page, instead of the repository link.

- Is there a particular reason to swap the qualitative and quantitative analysis in the methods compared to the order in the results?

- Page 4:
  - It would be helpful for the reading flow to put the ethical considerations in footnote.
  - "CORDIS": Please provide the link to the website.
  - Results: It would be interesting to get to know where the "prior knowledge of a DMP" came from.

- Pages 4-5:
  - When reading the results, it seems that the terms "data management" and "DMP" are mixed up or are sometimes used in the same way. There is no clear distinction. Please solve this ambiguity.

- Page 5:
  - Table 1: from the fourth row "other" please remove ", namely" or possibly replace it with the additional free-text information collected, if relevant. For rows 1-3 consider shortening the text (e.g. remove "part of").
  - Figure 1: The legend shows more options than in the diagram. Also, the description of the options can only be guessed in some cases "not available 27" or "others". Please add also a detailed, self-explaining caption.

- Page 7:
  - DMP project specificities: It would be interesting to get to know which answer is coming from which discipline, in order to fix the first clues of a discipline-specific effect. Also correlations with other side information (project size, partner composition...) might be interesting.

- Page 9:
  - Final thoughts from the interviewees:
    - It would be advisable to discuss these points already here, rather than in the following "discussion and recommendations" section. In particular the very relevant arguments about the "one repository" and "findability and availability of the data". The findability of DMPs could be improved by assigning to the DMP a persistent identifier such as a DOI or a DataCite DMP ID (https://support.datacite.org/docs/datacite-dmp-ids).
  - Discussion and recommendations:
    - Some parts are redundant to the above-mentioned results. Our suggestion: Integrate the redundant discussion and further statements above in the respective text passages. In our opinion, the concluding section should convey just the main statements and recommendations in a brief and crispy way.
    - We underline the importance of a one-stop-shop and really like the idea. On the other hand, most universities and research centers already offer inhouse (personal and technical) support to researchers. The best effect could arise with an intelligent coordination of these two offers, exploiting the high

acceptance of the inhouse support and the call-specific competence of the EU-Horizon RDM helpdesk service: so one other possibility could be that the EU-Horizon RDM team trains the inhouse support as multiplicators to support researchers in the best way.

- Page 10, “Data availability”:
  - The purpose of the presentation of the used data in the present form is not very clear. Some suggestions to get it a bit clearer:
    - The paragraph “Data availability” normally requires only a concise statement of whether and where the data are available, or which reason speak against data sharing.
    - Additionally, if you wish to describe the data in detail, then please write a concise paragraph for each of the two archives with some relevant information, dropping all redundant information (double file titles, long license statements, ...).
    - Additionally, to justify the citation, you are encouraged to recall the underlying data and as references or footnotes in all relevant positions in the previous sections.
    - Since the “interview guide” belongs to the first archive reported, there is no reason to report it separately. That way the distinction between “underlying data” and “extended data” is not applicable and the two sub-titles can be removed.

Is the work clearly and accurately presented and does it engage with the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

Are all the source data and materials underlying the results available?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: DMP, metadata, terminology, data stewardship, research data coordination,

active in NFDI4Ing, NFDI4Chem, RDA (DMP groups)

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 20 Oct 2021

Daniel Spichtinger

Thank you for the detailed review and apologies for the late response – due firstly to the fact that I needed to wait for the other reviews and secondly to the fact that the project has finished and it is difficult to find resources to continue the work on the next version of the article.

“most, several”: The text is meant to further elaborate on the quantitive information provided in the tables, often providing the more qualitative input from the interviews. I will check whether table 2 can be presented differently, since two reviewers have remarked on this.

EU policy which uses both “openness” and (in recent years) FAIRness and good data management: The article is not intended as an in-depth description of the ECs policies in this regard, but these are detailed in reference 6. A reference to the opt-outs will be added.

The DART project was eventually not essential to the work undertaken in the study, see also my response to reviewer 1.

In the methodology, the qualitative dimension is described first since time-wise the project started out with this. The quantitative dimension was then added.

The ethics paragraph was added on specific request from Open Research Europe. It is therefore not possible to move it to a footnote.

CORDIS – exact linked need to be checked with Vienna University IT services.

It is true that some questions referred to data management and some to data management plans. I will review this to make sure that the right terminology for the right question is used.

Table 1: “namely” will be removed, the other text is important to retain (rows 1-3).

Figure 1 will be revised, also based on the comments from reviewer 1.

Discipline specific analysis of the survey responses: this was outside the scope of the project and, given the fact that the project is closed, cannot be implemented at this stage.

A separate excel file with project information (size, budget) will be provided, as indicated also in the response to reviewer 1.

Discussions are for the discussions and recommendations section.

Assigning a persistent identifier to a DMP is a welcome suggestion that will be added to the recommendation's sections.

The part on data availability has been written according to the guidance from ORE and will therefore not be changed.

Competing Interests: No competing interests were disclosed.

Reviewer Report 11 May 2021

https://doi.org/10.21956/openreseurope.14412.r26801

© 2021 Dekker R. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ron Dekker
Consortium of Social Science Data Archives, Bergen, Norway

The study aims to identify good practices and common challenges when producing Data Management Plans (DMPs).
The hybrid approach of qualitative and quantitative empirical work is well thought out and pays off in the analysis.
Some striking results:

- 18% of the DMPs had an own Work Package for data management. However, there is no further analysis whether these projects have 'better DMPs', that is: score higher on the DART items (see below).

- data in one project was not considered as an original work (p. 7). It would have been interesting to elaborate on the (potential) consequences.

- almost 40% of the respondents did not publish their DMP. The author provides some solutions (within EC) but publishing the DMP, assign a DOI to this publication and depositing at Zenodo or similar provides better findability and sustainability.

Still some basic statistics are missing, e.g. size of the projects (budget, partners), size of the DMP (pages, MB), structure of the DMPs.

And what ever happened to the scoring of DMPs on the DART rubric (33 items maximum 66 points)? This rubric from the DART project (p.3) could have been explained, incl. the structure of its 33 categories "to standardize the review of data management plans". In addition, the reference could have been directly towards the paper instead of the project website.

It would have been nice if the EC could have commented on the results, especially the 'Final

thoughts'.

Detailed remarks:
The name of the Chair of the European Commission is Von der Leyen (p. 3).
The bullet points on Major challenges (p. 7) are sometimes in capitals and sometimes not.
Typo on p. 8 'a lot of shared date' --> '... data'.
On the selection criteria for the interviews (p. 3): the research interests of the staff ... - although useful - is not a very objective criterion. It would have been nice to see the numbers for each of the filters.
Figure 1 (p. 5) has five parts, but in the legend there are six categories, incl. an unexplained 'not available 27'.
In the Use of templates section (p. 6) the reference is to the qualitative survey, but the numbers indicate that this is the quantitative survey.
At the top of p. 7 it is not explained what the Agency is.

In the selection there was the criterion of at least two versions of a DMP, but this is not used in the analysis. So why this criterion at all?
There could have been a check on bias when selecting the DMPs (stage 1 and 2) and on the non-response when going from 840 white listed DMPs to the 108 that responded.
Table 2 (p. 6) could have been a graphic to facilitate interpretation.
The second bullet in 'What kind of support is needed' is rather cryptic: what matrix? What should be executed by research support organizations? And the last one: why is the library only relevant for bigger organizations?
The ARGOS tool is briefly mentioned (p. 9) but not explained, nor whether there are currently other tools to support DMPs, like CESSDA's Data Management Expert Guide.

Is the work clearly and accurately presented and does it engage with the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

Are all the source data and materials underlying the results available?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Open science, EOSC, research data infrastructures, data catalogues.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 20 Oct 2021

Daniel Spichtinger

Thank you for the detailed review and apologies for the very late response – due firstly to the fact that I needed to wait for the other reviews and secondly to the fact that the project has finished and it is difficult to find resources to continue the work on the next version of the article. I agree with your highlighting of salient facts – given that the project as such is closed, these issues would be good follow up projects. As for the statistics that are asked for (size of project, DMP) I will look into providing this as additional information in a separate excel file. I do include the results of the DART analysis as supplementary data. However, this needs to be considered with a "pinch of salt" since, due to the limited resources in the project, only one reviewer (the author) reviewed the selected DMPs. This is why, as the project developed, it was chosen to switch the main emphasis to complementing the qualitative interviews with a quantitative survey. The DART based evaluation, however, was important in that it provided input into the preparation for the qualitative interviews, helping to identify interesting project specifics. In the future it would be preferable to use a "four eyes" approach for such analysis (this is mentioned in footnote b in the article text). The results were also sent to the EC for potential use in their policy making activities.

Detailed remarks – where I do not provide comments I have implemented them in the main text of the article.

- Figure 1 will be revised.
- The selection criteria of having at least two DMP versions was useful for the qualitative interviews, where this fed into the creation of the interview guide. The selection criteria for qualitative interviews was not used in the survey.
- The selection of stage 1 and 2 was based on the objective criteria outlined (is it a DMP? Is it public?) and then undertaken by a number of volunteers. The results were also checked by the author. In the view of the author, the existence of systemic bias is therefore unlikely.
- The "cryptic" bullet(s) is/are input from the interviewees. The aim of the paper is not to provide an extensive discussion of the available data management tools.

Competing Interests: No competing interests were disclosed.