Aligning Data Management Plans with Community Standards using FAIR Implementation Profiles

Navroop K. Singh[1][0000−0001−9131−3528], Shuai Wang[1][0000−0002−1261−9930], Angelica Maineri[2][0000−0002−6978−5278], Tycho Hofstra[1][0009−0009−7320−864X], Mark Bruyneel[1][0000−0002−9426−3439], Stephanie van de Sandt[1][0000−0002−9576−1974], Ronald Siebes[1][0000−0001−8772−7904], Jacco van Ossenbruggen[1][0000−0002−7748−4715], and Tobias Kuhn[1][0000−0002−1267−0234]

[1] Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, the Netherlands {shuai.wang|t.m.hofstra|m.bruyneel|s.van.de.sandt|r.m.siebes|jacco.van.ossenbruggen|t.kuhn} @vu.nl, n.k2.singh@student.vu.nl

[2] ODISSEI, Erasmus University Rotterdam, 3000 DR Rotterdam, the Netherlands angelica@odissei-data.nl

Abstract. Data Management Plans (DMPs) are often required by organizations and funding agencies for research projects. One of the goals of DMPs is to capture how researchers plan to comply with some aspects of the Findability, Accessibility, Interoperability, and Reusability (FAIR) principles. When writing DMPs, taking into account community standards for managing and publishing research data can be a challenge for researchers. Community standards are often documented informally or communicated by word of mouth. The introduction of FAIR Implementation Profiles (FIPs) offers a structured way to capture such standards. This paper investigates with a user study, whether FIPs can serve as suggestions for aligning research data management with community standards. Through a customized interface with the related information extracted from FIPs as suggestions, we study whether participants take such suggestions into account when writing DMPs.

Keywords: FAIR Implementation Profile · FAIR Principles · Data Management Plans

1 Introduction

To make research data findable and readily reusable by others, researchers are often mandated by funding organizations and universities to create *Data Management Plans* (DMPs). A DMP is the result of a questionnaire (i.e. DMP template) with each question accompanied by some explanations. When completed, DMPs are formal documents consisting of answers to questions that outline how data is handled throughout and after a research project. Answering these questions, researchers specify the details and methods of data collection, data

repositories, responsibility, accessibility, licenses, etc. These answers can be influenced by many factors: the requirements of conferences and publishers, the recommendations by departments and universities, the suggestions by colleagues, the community standards, researchers' willingness to follow the FAIR principles (Findability, Accessibility, Interoperability, and Reusability) [13], etc.[3] Following community standards is a requirement of the FAIR principles (principle R1.3).[4] When uncertain about community standards, many choose to consult data stewards and colleagues. However, not all data stewards and colleagues can be aware of the standards of every community. Moreover, community standards are often in the word of mouth or informally documented (and often inaccessible beyond the community/organization). Adding more complexity to this alignment is when such standards evolve as members of the community adopt new tools, repositories, registries, licenses, etc. Thus, aligning DMPs with community standards remains an unsolved problem.

The introduction of *FAIR Implementation Profiles* (FIPs) offers a structured way to capture community standards [9]. FIPs serve as structured templates about decisions and guidelines by experts and members of communities of practice [9]. Moreover, FIP comes with related tools and metrics that make comparison of community standards and statistical analysis easy [9]. The alignment of DMPs with community standards captured by FIPs has the potential to offer substantial benefits to both researchers and their respective communities. For example, this alignment would make data findable in uniform repositories and promote the standardization of some machine-interpretable format, which makes it easier to integrate into a web framework and automatically compare with other schemas. However, the realization of effective alignment faces several obstacles. The DMP templates universities/institutes use can vary significantly. Some can have multiple versions for faculties and funding agencies. Determining which questions in the template could align with specific community standards can be ambiguous. Moreover, some datasets could be of interest to multiple communities, which further complicates researchers' efforts when selected for reuse.

The idea of using FIPs as suggestions for DMPs was initially proposed by K. Hettne et al. [4]. However, they did not conduct any user study to validate this idea. In this paper, we take an empirical approach and explore the workflow to extract information from FIPs as suggestions on the DMP interface. We study the following research questions. **RQ1:** Which questions in the DMP template can take community standards in FIPs as suggestions? **RQ2:** How can we build a user interface that takes community standards as suggestions? **RQ3:** How do users take advantage of suggestions from FIPs while writing their DMPs?

For RQ1, we first map DMP questions to their corresponding FAIR principles. We then filter out which questions from the mapping can be answered using suggestions from FIPs. For RQ2, we create a knowledge model (KM), a template file that specifies a tree-like structure of the DMP with its questions and some additional information. The KM is then used to generate the DMP interface

[3] https://www.go-fair.org/fair-principles/
[4] The R1.3 principle: `https://www.go-fair.org/fair-principles/`.

on the FAIR Wizard platform[5] with text-based suggestions. Finally, for RQ3, a user study is conducted followed by a survey to understand how participants take suggestions from FIPs. This paper made the following research contributions[6]: 1) a generic workflow for using FIPs in the interface as suggested in DMPs, 2) a mapping between the chosen DMP template and the FAIR principles, 3) an analysis of the relationship between the DMP questions and the FIP questions, 4) a reusable and extendable knowledge model that is used to generate the interface in the FAIR Wizard platform, and 5) a user study aimed to understand how researchers can effectively use FIPs as suggestions while creating DMPs.

2 Related Work

Despite the potential impact of FAIR community standards on researchers' choices over management and publication of research data, the connection between DMPs and FAIR principles has been empirically examined only in a few studies. Henning et al. [3] analyzed 10 DMP templates and concluded that DMPs fail to capture detailed community-specific implementations, especially the principle of interoperability, and do not cover metadata sufficiently. In a study by Mannheimer, DMPs associated with grant proposals were analyzed along with interviews with the Principal Investigators who wrote them [6]. It was found that the more technical parts of the DMPs (including questions on FAIR) were the least detailed, and that they would need more training and guidance on more specialized concepts such as FAIR and metadata. Likewise, a report by OpenAIRE on the Horizon 2020 template for data management plans highlighted the need to clarify issues and terms around FAIR implementation [2]. Finally, regarding the use of Open Science Framework (OSF) platform for DMPs, Sullivan et al. [11] indicated the importance of referring to best practices applied to different research contexts. These studies suggest that clearer guidance on FAIR implementation and standards can be beneficial to researchers when filling in their DMPs. By following FAIR community standards, for instance, published metadata could also be more easily harvested, refined, or enriched by platforms such as the Data Europa[7], and the ODISSEI portal[8].

To our knowledge, the only attempt that explicitly link the DMPs with the FAIR community standards declared in FIPs used the DMP template of Leiden University. The authors identified seven questions in their DMP template that could be linked to the FIP questions [4].[9] They proposed to develop a knowledge model of the Leiden University DMP template and import answers from a FIP as

[5] The FAIR Wizard uses the Jinja template engine: `https://fair-wizard.com/`.

[6] The DMP template, the knowledge model, the mock DMPs, the survey, Python code, the analytical results, and a demo are in a repository on Zenodo (`https://doi.org/10.5281/zenodo.10285647`). The Python scripts for the analysis of survey results are at `https://github.com/FAIR-Expertise-Hub/FIP2DMP`.

[7] `https://data.europa.eu/en`.

[8] `https://portal.odissei.nl/`

[9] The mapping is at `https://osf.io/5jsfp`.

pre-filled answers for the DMP, leaving it to the user to select the most relevant ones [8]. A KM is a structured document template with questions, descriptive text accompanying each question, and the type of answers specified. The template includes examples of good data management practices and guidance on how to meet the requirements of various funding agencies and institutions. Our research is inspired by this work. Their proposal takes into account one FIP, while in reality, there could be multiple communities that could be relevant to a research project. They proposed to have imported decisions from the FIP as pre-filled answers in the DMP. However, since the efficacy of using the information captured by FIP has not been evaluated, it remains debatable if such information can be used as pre-filled answers as researchers' data management decisions in their DMPs, not to mention that not all the answers in the FIPs are correct. Moreover, our examination of the FIPs in social science shows that some entries could be missing or incomplete for some FIPs. A question in a selected FIP could correspond to multiple resources. If a user decides to align with a community by taking the resources indicated by its FIP as the answer, then which one should be taken as the pre-filled answer in the DMP? This could lead to confusion. Moreover, they proposed to allow the import of information from a FIP while writing a DMP. This would overwrite the user's answers, which can lead to confusion as the user would have to look into the version history. Thus, their proposed approach could have some lack of consideration in practice. Moreover, they did not include a user study. Our approach is inspired by their work, but differs in the DMP template used, the development of KM, and how (multiple) FIPs are handled as suggestions. In addition, we include a user study and evaluate the efficacy. Details of the user study are in Section 6.

3 Workflow

In this section, we provide the details of the workflow of our approach. Given a selected DMP template, to provide suggestions for the right questions in DMP, we need to find the mapping between the questions in DMP and FIPs. To do this, we first associate a list of questions with the FAIR principles. We can further narrow this list by removing questions that do not correspond to any question in the FIP. This results in a mapping from a subset of questions in DMP to a subset of questions in FIP. Note that the mapping is not always one-to-one given that the questions in DMPs are typically developed without considering the correspondence to the FAIR principles, respectively. This mapping is then used to extract community standards captured by the answers to the selected questions in the FIPs.[10] The extracted information can be structured as a table of the communities and their resources for each question (see Section 4).

[10] In this paper, due to the limited number of FIPs in the domain of social science, we do not filter out any existing FIP in social science. The selection/recommendation of FIP regarding a project could be further explored when there are enough FIPs in social science.

Next, we customize the FAIR Wizard platform to include the extracted information. A knowledge model (KM) of the FAIR Wizard is a customizable template that can be imported into the platform. It is a file of computer readable and actionable statements about questions in chapters with additional information that explains the questions, the specification of API calls, and the expected type of answers (string, yes/no, etc.). A KM can then be loaded into the FAIR Wizard to form a customized interface as specified. For each question in the DMP template, the interface displays the question, additional descriptive information, and suggestions from FIPs. The interface is then used to create DMPs. More details can be found in Section 5. To evaluate the efficacy, we perform a user study. Each study participant is asked to create a DMP and complete a survey. The DMP and the survey results are then used for analysis (see Section 6).

4 Knowledge preparation: Connecting DMPs with FIPs

4.1 Data Management Plans

DMPs have become a standard in recent years and are required more and more by funding organizations. Miksa et al. describe them as 'awareness tools', shedding light on data management practices employed in research projects [7]. In practice, they assist researchers in ensuring proper management, documentation, and preservation of data, while also meeting funders' requirements. DMPs are implemented from DMP templates, which consist of questions with specified answer types. DMP templates can be very different from each other, as they are often tailored to best facilitate the data management of researchers and meet the requirements of the funding organizations. Various tools can be used to instantiate such templates with a user interface, such as DMPOnline[11], DMP Tool and Data Stewardship Wizard [10]. For this paper, we use the DMP template by the Vrije Universiteit Amsterdam (VU). The template is hosted on DMPOnline whose strength lies in the convenience for researchers to request feedback from data stewards. Furthermore, DMPOnline dynamically displays or omits questions based on the researcher's responses regarding dealing with personal data. By March 2023 when this project started, the DMP in use at VU was identified as '*1 - VU DMP template 2021 (NWO & ZonMW certified) v1.3*', which consists of 53 questions. The VU DMP template is used by all faculties except the Medical Faculty. In this study, we focus on researchers in the Faculty of Social Sciences. The questions in the template cover a wide range of topics including authorship, legal and privacy ethics, funding number, etc.

When comparing the VU DMP template with that of Leiden,[12] there are notable differences. The Leiden DMP consists of 48 questions (compared to 53 in the VU template), mostly in multiple-choice format, in contrast to VU's text-based answers. Both address findability, accessibility, and reusability. That of Leiden places more emphasis on privacy concerns and security risks.

[11] `https://dmponline.vu.nl/`

[12] The Leiden DMP template used is at `https://zenodo.org/records/4423065`.

4.2 Mapping the DMP Template to the FAIR principles

Since the FAIR principles are widely endorsed for good data management, most DMPs also include questions on the implementation of FAIR [3]. Hence, to answer RQ1, we begin by analyzing the 53 questions in the chosen VU DMP template and exclude questions that do not correspond to any FAIR principle. After consulting the team at the University Library that develops and maintains the DMP and compare it against the FAIR principles, we identified a total of 17 questions relevant to the FAIR principles.[13] More specifically, 14 of the questions are about the Findability principle, and two questions are related to the Accessibility principle. Three other questions focus on the Reuse principle. Interoperability is not addressed explicitly. Since the template was designed without exactly following the FAIR principles, questions can correspond to multiple principles. For example, Question 5.1 corresponds to both F2 and R1.2 principles. Among all the questions that have to do with the FAIR principles, we observe that 82.4% of the questions are about Findability since a significant amount of questions have to do with persistent identifiers. Question 5.3 can cover much broader topics than one FAIR principle and leads to ambiguity.

4.3 FAIR Implementation Profiles

A FAIR Implementation Profile (FIP) is a set of choices made by a FAIR Implementation Community (FIC) on how to implement the FAIR principles [9]. The community decision in a FIP is collected using a questionnaire[14] in which experts and members of the community collectively indicate their preferred FAIR Enabling Resources (FERs), that include tools, documentation, registries, licenses, standards, and other resources that are needed to achieve a specific aspect of FAIR implementation [9, 12].[15] Examples include the REST API, IISG Dataverse[16], CC-BY-NC 2.0[17], etc. Each question corresponds to a FAIR principle. FIPs have been conceptualized to foster convergence of FAIR implementation efforts across communities and domains [9]. FIPs are filled in using the FIP Wizard[18], providing an easy-to-use interface. FIPs can be published in a machine-actionable format as nanopublications.

There are already over a hundred FAIR Implementation Communities covering various domains. For this study, we focus on the six FIPs that pertain to communities in the social sciences [12].[19] The social sciences constitute an

[13] Details are given in the supplementary material.

[14] `https://bit.ly/yourFIP`

[15] `https://peta-pico.github.io/FAIR-nanopubs/fip/index-en.html#https://w3id.org/fair/fip/terms/FAIR-Enabling-Resource`

[16] `https://iisg.amsterdam/nl/data/datasets`

[17] `https://creativecommons.org/licenses/by-nc/2.0/`

[18] The FIP Wizard platform is a specified version of the general-purpose FAIR Wizard. It is available online at `https://fip-wizard.ds-wizard.org`.

[19] These were the only social science FIPs available by the time the project start in spring 2023.

interesting use case due to their longstanding tradition of data sharing, showed by the abundant availability of large-scale survey data, yet combined with a large heterogeneity in the standards adopted. Among the six FIPs, three come from communities that publish survey data: GESIS social Science Survey Research (GESIS SSSR), the European Social Survey (ESS), and the Australian correspondent (AUSSI-ESS). In addition, the Dutch Socio-Economic History (SEH) and the Media Content Analysis Lab (MCAL) are two communities in the Netherlands. Lastly, the LGBTQ+ Linked Open Vocabulary (LGBTQVoC) community creates multilingual LGBTQ+ controlled vocabularies for indexing digital records to represent LGBTQ+ objects in non-English languages. Details about these communities can be found in [12].

4.4 Assigning FIPs to DMP Questions

ID	DMP Question	FAIR principle	FIP Question
4.6	Where will you publish your data assets?	F4 Data	In which search engines are your datasets indexed?
4.8	How will you ensure your data assets get a persistent identifier (e.g. a DOI-code)?	F1 Data	What globally unique, persistent, resolvable identifiers do you use for datasets?
4.9	Will you register your datasets in an online registry other than PURE? If yes, where?	F1 Data	What globally unique, persistent, resolvable identifiers do you use for datasets?
4.13	Please indicate the license and/ or terms of use under which you share your data.	R1.1 Data	Which usage license do you use for your datasets?
5.1	What metadata and documentation will accompany the project?	F2 & R1.2 Data	Which metadata schemas do you use for findability? & Which metadata schemas do you use for describing the provenance of your datasets?
5.2	What metadata and documentation will accompany the data assets?	F2	Which metadata schemas do you use for findability?
5.3	What methods, software or hardware are needed to access and use your data?	R1.2 Data	Which metadata schemas do you use for describing the provenance of your datasets?

Table 1. Mapping of DMP Questions to FIP questions via the FAIR principles.

Next, we use the mapping in Section 4.2 and find the correspondence between questions in FIPs and DMPs. We manually examined questions that correspond to the same FAIR principle and identified eight DMP questions that may be successfully connected to FIPs. However, one question (Question 6.3) was excluded due to its ambiguity.[20] Table 1 illustrates the correspondence between these DMP questions to the FIP questions after manual examination. Question 4.6, 4.8, 4.9, and 4.13 have a clear one-to-one mapping to the corresponding FIP questions. [21] The case of the next section is more complicated. Question 5.2

[20] Question 6.3 'For data that are only available upon request, what methods will be used to handle requests for access and how will data be made available to those requesting access?' was linked to FAIR principle A1.2 in the previous step.

[21] For a discussion on the mapping of question 4.9, see section 8.

focuses on metadata schemas for data assets. Question 5.3 addresses provenance models and methods. Question 5.1 bridges both, encompassing rich metadata and detailed provenance for the entire project, with metadata schemas. Other questions that are nontrivially related to the FAIR principles are 4.2 (F4), 4.3 (F4), 4.10 (A1.2), and 6.3 (A1.2). Thus, around 63.6% of the FAIR-related questions can have suggestions from FIPs.

5 Interface Customization

Recall that RQ2 is about building an interface with extracted community standards as suggestions. Currently, the chosen DMP template is hosted on the DMPOnline platform, which does not support customization of its interface, nor specifying FERs as an answer. Furthermore, the platform cannot convert the resulting DMP to machine-actionable formats and does not support loading content from other datasets using customized queries. Thus, we migrated the template to the FAIR Wizard. In this section, we explain how the knowledge model is constructed and can be used to create the corresponding interface on the FAIR Wizard platform. For a DMP, its KM is a template file based on Jinja[22] (a template engine) that specifies a tree-like structure of the corresponding DMP with its questions, some explanatory text, and the expected type of answer. Additionally, it allows some customized functions to retrieve external information as options for answers.[23] Suggestions extracted from chosen FIPs are included as additional information below the questions. In this paper, we limit our KM to questions chosen in Section 4.4. For each question, two types of information are expected: a list of FERs and a string description that explains how the chosen resources will be used as well as some additional information to cover resources not found in the current system. The created KMs can be modified, extended, downloaded, and reused. Finally, the interface is automatically generated when the KM is uploaded. A new DMP project can be initiated for each user.

6 User Study

After reaching out to 31 researchers in the Faculty of Social Sciences of the VU[24], a total of 6 researchers agreed to participate in this study. They have written DMPs or have sufficient knowledge of them. Participants first pick one out of four mock DMP that is closest to their research. These mock DMPs were inspired by some ideas of existing DMPs by university researchers. The participants were then asked to complete the DMPs with selected questions by using our customized interface, followed by a small survey about their experiences and the effectiveness of the suggestions. The survey consisted of 12 survey questions

[22] `https://palletsprojects.com/p/jinja/`

[23] `https://guide.ds-wizard.org/en/latest/about/introduction/knowledge-model.html`.

[24] Teaching staff, supporting staff, retired professors, external and visiting researchers were excluded. PhD students were included.

(SQs) in two parts (see the supplementary material for the complete list of questions). In Part A (SQ1-SQ7), participants assessed the relevance and usefulness of the suggestions: the relevance of communities for DMPs, decision alignment with communities, and the effectiveness of suggestions for specific DMP questions, along with the ease of locating FERs. Part B (SQ8-SQ11) focused on participants' background and experience with DMPs. Additionally, they were asked to rate the clarity of the study's goals. Finally, the survey ended with a question about how they consider aligning with community standards compared to other stakeholders. The user study was conducted in October 2023.

7 Evaluation

For our analysis, we used the answers of five participants[25]. The participants had a range of 3 to 10 years of experience in academia, counting from the start of their PhD (SQ8). With the exception of one participant, all participants had prior experience in writing DMPs (SQ9). The survey results indicated that participants, on average, found the objective of the study to be moderately clear, with a mean rating of 3.4 out of 5 (see SQ10 in Table 2). Some indicated the lack of knowledge of FIP and FER. Five key survey questions and an analysis of the corresponding results are included in Table 2. In SQ1, participants were asked to rate the relevance of research communities for their DMPs. The mean rating of 3.4 out of 5 suggested that *some* participants found research communities relevant to their DMPs. Responses of SQ2 spans from 2 to 4, with a mean rating of 3.4 and a median of 4, indicating that the participants perceived these suggestions relatively helpful. Regarding SQ3, participants found alignment with community decisions moderate important with a mean of 3.33.

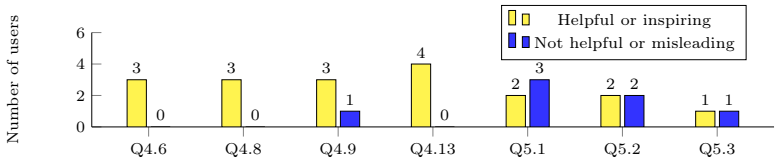

Fig. 1. Comparing the impact of suggestions on DMP questions: helpful or inspiring vs. not helpful or misleading.

Next, we evaluate the effectiveness of the suggestions for each DMP question (corresponding to SQ4 and SQ5). Figure 1 shows that questions 4.6, 4.8, 4.9, and 4.13 in the VU DMP template were perceived by the participants as helpful or inspiring to some extent. However, a closer look reveals varying responses to Question 5.1. This question, which inquired about the types of documentation to be produced during the research project, was frequently perceived as not

[25] The PDF file of one of the mock DMPs downloaded from the DMPOnline had missing pages. This was not reported until the end of the user study. One participant chose that DMP and was therefore excluded from the analysis

Survey Question ID	Survey question	Mean	Median	Std
Q1	On a scale of 1 to 5, how relevant are communities for this DMP? 1 indicating that no community is relevant and 5 indicating that many communities are relevant.	3.4	3	0.89
Q2	On a scale of 1 to 5, please evaluate whether the suggestions provided in this DMP are helpful for the communities in answering their corresponding questions. 1 indicating that it is not helpful and 5 indicating that it is very helpful.	3.4	4	0.89
Q3	On a scale of 1 to 5, how much would you consider aligning the decisions in this DMP with those made by the relevant community? 1 indicating minimal alignment and 5 indicating complete alignment.	3.4	3	1.14
Q7	On a scale of 1 to 5, how easy was it for you to find the FAIR-Enabling Resource in the search bar? 1 indicating extremely difficult and 5 indicating extremely easy.	3.6	4	1.14
Q10	On a scale of 1 to 5, how clear was the goal of the study to you? 1 indicating not clear at all and 5 indicating very clear.	3.4	4	0.89

Table 2. Survey questions and their results together the range, mean, median, and standard deviation.

helpful or misleading. This might be attributed to the inherent ambiguity of the question, but also the weak link with the FIP question, which does not include the documentation part. The DMP question demands descriptions of documentation, including codebooks, lab journals, read-me files, research logs, and protocols. The challenges arise because the FIP question can only provide FERs. Questions 5.2 and 5.3 were less attended with neutral results.

Our survey includes participants' experience with FERs. Although most participants cannot find all the FERs desired through the search bar, they find the search bar easy to use to find resources with only one participant reported successfully finding all the FERs they intended to specify (SQ6) and a mean rating was 3.6 out of 5 when asked to rate how easy it was to find FERs using the search bar. This removed authors' concerns about switching platforms from DMPOnline to the FAIR Wizard. Some further analysis of the resulting DMPs shows that all participants managed to specify at least one FER for Questions 4.6 and 4.13. In Questions 4.9 and 5.1, some users struggled to locate the desired FERs through the search bar. This was reflected in SQ12: some are confused about the terminology; some are not familiar with metadata standards.

8 Discussion

The pilot study, along with the analysis, elicits discussion on the approach. The mapping between the questions in the DMP template and the FIP questions is not unambiguous and can depend on interpretation. This is because the chosen DMP template did not take the FAIR principles into account by design. Our analysis shows that out of the seven identified DMP questions, suggestions from FIPs to three questions (5.1, 5.2, and 5.3) can be unhelpful or misleading. This is because the Questions 5.1 and 5.3 cover multiple issues and are only weakly linked to the FIP questions. 5.1 and 5.3 both correspond to R1.2 where

three FIPs lack resources to contribute to the suggestion. Moreover, Question 4.9 is formulated in a way that excludes the PURE system, the default choice required by the university, as an answer[26]. At further inspection, it emerged that principle F4 Data would match this DMP question better, yet this was missed due to the way the FIP question was formulated, mentioning search engines. It is only recently that the FIP question was re-formulated and now focuses on services (registries). This shows that an accurate mapping is crucial for FIP suggestions in DMPs to make sense. For instance, the way FIPs intend 'Accessibility' focuses mostly on machine accessibility. In contrast, DMPs focus more on human accessibility. Thus, a unified vocabulary for DMP and FIP is essential. Another limitation worth noting is that, despite the detailed introduction, some participants expressed that they found it difficult to comprehend certain terminologies and lacked knowledge of metadata standards. This could also be a cause of confusion for Question 5.2. In practice, metadata handling is typically the responsibility of data repositories or data stewards.

Inspired by [5], we included in the survey a question (SQ11) about which stakeholders have the greatest impact on their decision-making. Our participants indicated that the department, faculty, and the university research data management team, as well as the ethics committee, have the most significant influence on their decisions in DMPs (13.95%). That of community shares the second place with the university I.T. team, as well as the data management platform (9.30%). Some other factors could also be taken into account. Despite the scale of the survey, it shows that the decision can be influenced by many factors. However, these suggestions and guidelines from other stakeholders are mostly formatted as textual information in PDF format, which cannot be easily imported into a DMP editing platform.

Finally, the pipeline faces some challenges. Note that FIP is not free from errors. Mistakes from FIP could propagate through the pipeline and eventually end up in the DMP if the suggestion were taken. The pipeline depends on the correct mapping between the DMP questions and FIP questions. Thus, a careful review of the questions in the DMP template and the mapping is essential in future work. As the number of FIPs grows, there could be many resources as suggestions that could be confusing, especially for interdisciplinary projects. Thus, a selection of FIPs could be offered to the user in future work.

9 Conclusion and Future Work

Alignment with FAIR community standards via the DMP, could facilitate the harvesting of metadata of published dataset from Portals [27]. This paper explored how FIPs can be used as suggestions for DMPs and whether researchers can align their DMPs with community decisions through the use of FIPs. To address RQ1, we constructed a mapping between DMP and FIP, and identified 7 DMP questions where community standards captured by FIPs could be used as

[26] `https://vu.nl/en/employee/research-data-support/research-portal-pure`
[27] See for an example: `https://data.europa.eu`.

suggestions. For RQ2, we constructed a KM tailored to the VU DMP template with the information of six distinct research communities' standards integrated into each question. Finally, for RQ3, a pilot user study was conducted, which revealed that, for some questions, some users find the suggestions from FIPs helpful or inspiring. Due to the fixed scope and the limited number of participants, the conclusion still needs to be validated in different research domains and on a larger scale. Moreover, alternative principles exist next to FAIR, such as the CARE principles (Collective Benefit, Authority to Control, Responsibility, and Ethics) [1], calling for an extension of our approach in future work. As a proof-of-concept, our primary focus was on the social science using six FIPs, but we aspire to broaden the scope of our work to include other domains and FIPs in the future.

Acknowledgments

The research for this paper was made possible by the Platform Digitale Infrastructuur SSH. The authors appreciate help from volunteers.

References

1. Stephanie Russo Carroll et al. The CARE principles for indigenous data governance. *Data Science Journal*, (1), 2020.
2. Marjan Grootveld et al. OpenAIRE and FAIR Data Expert Group survey about Horizon 2020 template for Data Management Plans. 2018. 10.5281/zenodo.1120245.
3. Patrícia Henning et al. The FAIRness of data management plans: an assessment of some european dmps. *Revista Eletrônica de Comunicação, Informação e Inovação em Saúde*, 2021.
4. Kristina Hettne et al. Fip2dmp: Linking data management plans with fair implementation profiles. *FAIR Connect*, 1:23–27, 01 2023.
5. L. Kvale et al. Understanding the data management plan as a boundary object through a multi-stakeholder perspective. *Int. J. Digit. Curation*, 16:16, 2021.
6. Sara Mannheimer. Toward a Better Data Management Plan: The Impact of DMPs on Grant Funded Research Practices. *Journal of eScience Librarianship*, 7(3), 2018.
7. Tomasz Miksa et al. Ten principles for machine-actionable data management plans. *PLoS Comput. Biol.*, 15(3):e1006750, 2019.
8. Fieke Schoots et al. Data management plan template leiden university (4.2). Zenodo, 2021. `https://doi.org/10.5281/zenodo.4423065`. Accessed June 2023.
9. Erik Schultes et al. Reusable FAIR Implementation Profiles as accelerators of FAIR Convergence. In *Advances in Conceptual Modeling*. Springer, 2020.
10. Jurrian H. Spaaks et al. The data stewardship wizard: A tool bringing together researchers, data stewards, and data experts. *Int. J. Digital Curation*, 14(1), 2019.
11. Ian Sullivan et al. Open and reproducible research on open science framework. *Current Protocols Essential Laboratory Techniques*, 18, 2019.
12. Shuai Wang et al. FAIR Implementation Profiles for social science. In Emmanouel Garoufallou et al., editors, *Metadata and Semantic Research*, pages 284–290, Cham, 2024. Springer Nature Switzerland.
13. Mark D Wilkinson et al. The fair guiding principles for scientific data management and stewardship. *Scientific data*, 3(1):1–9, 2016.