Aligning Data Management Plans with Community Standards using FAIR Implementation Profiles*

1st Navroop K. Singh
Department of Computer Science
Vrije Universiteit Amsterdam
Amsterdam, the Netherlands
n.k2.singh@student.vu.nl
ORCID: 0000-0001-9131-3528

2nd Shuai Wang
Department of Computer Science
Vrije Universiteit Amsterdam
Amsterdam, the Netherlands
shuai.wang@vu.nl
ORCID: 0000-0002-1261-9930

3rd Angelica Maineri
Erasmus School of Social and Behavioral Sciences,
ODISSEI & Erasmus University Rotterdam
Rotterdam, the Netherlands
angelica@odissei-data.nl
ORCID: 0000-0002-6978-5278

4th Tycho Hofstra
University Library
Vrije Universiteit Amsterdam
Amsterdam, the Netherlands
t.m.hofstra@vu.nl
ORCID: 0009-0009-7320-864X

5th Mark Bruyneel
University Library
Vrije Universiteit Amsterdam
Amsterdam, the Netherlands
m.bruyneel@vu.nl
ORCID: 0000-0002-9426-3439

6th Stephanie van de Sandt
University Library
Vrije Universiteit Amsterdam
Amsterdam, the Netherlands
s.van.de.sandt@vu.nl
ORCID: 0000-0002-9576-1974

7th Ronald Siebes
Department of Computer Science
Vrije Universiteit Amsterdam
Amsterdam, the Netherlands
r.m.siebes@vu.nl
ORCID: 0000-0001-8772-7904

8th Jacco van Ossenbruggen
Department of Computer Science
Vrije Universiteit Amsterdam
Amsterdam, the Netherlands
jacco.van.ossenbruggen@vu.nl
ORCID: 0000-0002-7748-4715

9th Tobias Kuhn
Department of Computer Science
Vrije Universiteit Amsterdam
Amsterdam, the Netherlands
t.kuhn@vu.nl
ORCID: 0000-0002-1267-0234

***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 for each question in the chosen DMP template, we study whether participants can take such suggestions into account when writing DMPs. Subsequently, a survey was conducted for each participant, highlighted some specific DMP questions where FIPs can be effectively used as suggestions. The survey result suggests the potential of FIPs as a valuable resource to harmonize research data with community standards.**

***Index Terms*—FAIR Implementation Profile, data management plan, research data management, the FAIR principle**

I. 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) [1], etc.[1] Following community standards is a requirement of the FAIR principles (principle R1.3).[2] 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

The project is funded by the PDI-SSH (Platform Digitale Infrastructuur - Social Science and Humanities).

[1] https://www.go-fair.org/fair-principles/

[2] More details on the R1.3 principle are available at https://www.go-fair.org/fair-principles/.

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 [2]. FIPs serve as structured templates about decisions and guidelines by experts and members of communities of practice [2]. Moreover, FIP comes with related tools and metrics that make comparison of community standards and statistical analysis easy [2]. 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 from each other. Some can have multiple versions based on the faculty and funding agency. Determining which questions in the DMP template could align with specific community standards can be ambiguous. Moreover, the datasets could be of interest to multiple communities, which further complicates researchers' efforts when selected for reuse. In this paper, we explore a convenient, reliable, and easy-to-use means for researchers to align with community standards using FIPs.

The idea of using FIPs as suggestions for DMPs was initially proposed by K. Hettne et al. [3]. 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 evaluate its efficacy with a survey. We aim to answer 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 on the FAIR Wizard platform[3] 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.[4]

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.
5) A user study aimed to understand how researchers can effectively use FIPs as suggestions while creating DMPs.

The paper is organized as follows: Section II provides an overview of related work. Section III gives an overview of the workflow. Section IV provides more details about DMPs and FIPs and identifies which questions in the chosen DMP template can be addressed using FIP suggestions. Section V delves into the creation of an interface on the FAIR Wizard platform using a knowledge model that includes suggestions from chosen FIPs. Section VI provides details of the user study. The evaluation of the resulting DMPs and the survey is presented in Section VII followed by some discussion in Section VIII. Finally, Section IX presents the conclusion and outlines future work.

II. 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. [4] 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 PIs (Principal Investigators) who wrote them [5]. It was found that the more technical parts of the DMPs (including questions on FAIR) were the least detailed, and that PIs 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 [6]. Finally, in a paper outlining how to use the Open Science Framework (OSF) platform for DMPs, Sullivan et al. [7] 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. However, our research shows that this has not been explored systematically.

To our knowledge, the only attempt that explicitly link the DMPs with the FAIR community standards declared in FIPs used the DMP template of the Leiden University. The authors identified seven questions in their DMP template that could be linked to the FIP questions [3].[5] They proposed to develop a knowledge model of the Leiden University DMP template

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

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

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

and import answers from a FIP as 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 VI.

III. Workflow

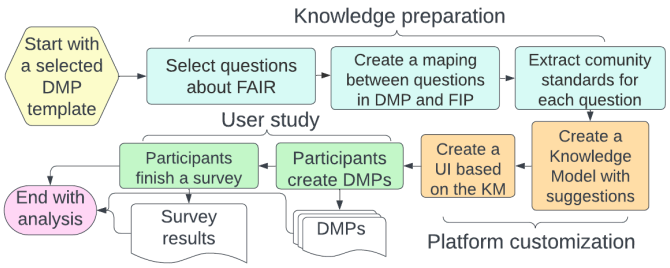

Fig. 1. A visualization of the workflow with different colors for the three stages: knowledge preparation (blue), platform customization (orange), and user study (green).

In this section, we provide the details of the workflow of our approach. Figure 1 is a visual representation of this workflow. 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.[6] The extracted information can be structured as a table of the communities and their resources for each question. More details are given in Section IV.

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 structured file consisting of questions in chapters with additional information that explains the questions, 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 V.

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 VI for details.

IV. Knowledge Preparation: Connecting DMPs with FIPs

Next, we explain the details of DMPs (Section IV-A) and assign the FAIR principles to the questions in DMPs (Section IV-B). Following that, we introduce the FIPs in detail (Section IV-C) and presents the mapping between DMP questions and FIP questions (Section IV-D).

A. 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 [9]. 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[7], 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. Researchers can

[6]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 studied in the future when there are enough FIPs in social science.

[7]https://dmponline.vu.nl/

conviniently fill in their answers and download the completed DMP in the PDF and JSON format.

DMPOnline's strength lies in the convenience for researchers to start filling in their data management plan and request feedback. The data stewards, on the other hand, can easily provide comments and guiding article on the side line. This is a fairly manual interaction, though, in which hardly any automation is present. DMPonline automatically provides the data management plan templates of all major funders of Europe, including NWO, ERC and ZonMW. These templates differ in accessibility for beginning researchers, with most of the templates focusing specifically on the FAIR needs of the funder. The NWO template, for example, covers 18 questions with multiple sub-questions bundled together. It directly addresses FAIR principles, asking researchers how they will ensure data FAIRness. The template provided by ERC consists of 8 questions, emphasizing FAIRness by asking researchers to describe how they will make their data findable, accessible, interoperable, and reusable. However, such broad questions make it difficult to associate with a specific FAIR principle, which may hinder machine-actionability. Therefore, they are not to be considered for this research. Years of iteration and refinement have made the current VU DMP template to cover all the requirements by major funders, as well as guide the researcher in planning the research data management required to streamline their research at the VU. The result is a template with roughly 53 questions, depending whether the research will contain personal or sensitive data. Guiding instructions are provided along the step-by-step questions. The current DMP template covers the two formerly parallel editions which are made for research with and without personal data, incorporating questions omitted based on researchers' choices in the document along with guiding text. 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 is identified as '*1 - VU DMP template 2021 (NWO & ZonMW certified) v1.3*'. This version consists of 53 questions[8] and is developed and maintained by the data stewards of the University Library. The VU DMP template is used by all faculties except the Medical Faculty, which maintains its own template and employs a separate data management team. In this study, we focus on researchers in the Faculty of Social Sciences. The questions in VU's DMP template cover a wide range of topics including authorship, legal and privacy ethics, funding number, etc.

When comparing the VU DMP template with the current Leiden version,[9] there are notable differences. The Leiden DMP consists of 48 questions (compared to 53 questions of the VU template), mostly in multiple-choice format, in contrast to that of the VU, which are entirely text-based answers. Both templates share common ground, addressing aspects such as findability, accessibility, and reusability. That of Leiden places more emphasis on privacy concerns and security risks.

B. 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 [4]. 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.[10] 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. More discussion is included in Section VIII.

C. 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 [2]. The community decision in a FIP is collected using a questionnaire[11] 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 [2], [11].[12] Examples include the REST API, IISG Dataverse[13], CC-BY-NC 2.0[14], etc. Each question corresponds to a FAIR principle. Decisions made by communities can significantly impact researchers, influencing choices related to data repositories, licenses, and metadata standards. FIPs have been conceptualized to foster convergence of FAIR implementation efforts across communities and domains [2]. FIPs are filled in using the FIP Wizard[15]. The FIP Wizard provides an easy-to-use interface and allows one to generate human-readable PDF reports and publish the FIPs in a machine-actionable format as nanopublications.

There are already over a hundred FAIR Implementation Communities covering various domains. For this study, we

[8]It was the latest version at the time of project, consisting of the most amount of questions following the division of metadata questions into questions that follow a more manageable line of thinking.

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

[10]Details are given in the supplementary material.

[11]https://bit.ly/yourFIP

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

[13]https://iisg.amsterdam/nl/data/datasets

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

[15]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.

TABLE I
A COMPARISON OF DECISIONS ON THE PRACTICE OF THE FAIR PRINCIPLE ('F' STANDS FOR FUTURE USE)

FAIR Principle	FIP question	SEH	MCAL	LGBTQVoC	SSSR	ESS	AUSSI-ESS
F1 Data	What globally unique, persistent, resolvable identifiers do you use for datasets?	DOI (F), Handle	DOI	DOI	-	DOI, DDI URN	DataCite DOI resolution service
F2	Which metadata schemas do you use for findability?	MARC21, EAD3, DDI-Codebook, DCAT2	-	MARC21, BIBFRAME	DDI-Codebook	DDI-Lifecycle 3.3	DDI-Codebook, DataCite 3.1
F4 Data	In which search engines are your datasets indexed?	Dataverse, DANS SSH data station (F)	DANS SSH data station, Dataverse, OSF, Figshare, TriplyDB (F)	-	GESIS Search	ESS Data Portal, EOSC Portal	ADA Dataverse
R1.1 Data	Which usage license do you use for your datasets?	CC-BY-SA, CC-BY-NC	-	CC-BY-NC-ND 4.0	GESIS Usage Regulations 2018	CC-BY-NC-SA 4.0	-
R1.2 Data	Which metadata schemas do you use for describing the provenance of your datasets?	DCAT2	-	-	-	DDI-Lifecycle 3.3, DDI-Codebook, DDI-CDI (future)	DDI-Codebook

TABLE II
MAPPING OF DMP QUESTIONS TO FIP QUESTIONS VIA THE FAIR PRINCIPLES.

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?

focus on the six FIPs that pertain to communities in the social sciences [11].[16] The social sciences constitute an interesting use case due to their longstanding tradition of data sharing, showed, for instance, by the abundant availability of large-scale survey data, however, 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. Due to the page limit, details about these communities, the creation of the FIPs and their details, and how they can be used for analysis of convergence and comparative studies are omitted but can be found in [11].

Table I compares the resources chosen by the communities regarding each question in the FIP. These FIP questions are relevant for DMPs and are in the mapping we create in Section IV-D. The row for 'F1 Data' highlights convergence among social science communities, indicating a preference for DOI. Regarding 'R1.1 Data', three out of six communities indicate the use of different versions of CC-BY licenses for their data. This information could indicate the popularity of certain resources across/within selected communities.

However, the examination of FERs regarding other principles shows the differences. Many different FERs are mentioned under 'F2': for instance, BIBFRAME is highly specific to LGBTQVoC, whereas DDI-Codebook is mostly mentioned by communities that deal to some degree with survey data. The FERs indicated under 'F4 Data' show fragmentation, with each FIC indicating different data repositories. Since these FIPs are all distinct despite being in social science domain, it is important that the researchers carefully consider which community aligns the most to their research project's objectives to maximize findability and reusability of their research data by other researchers in the same community. Finally, the blank answers could make the suggestion less informative.

As a new means to capture knowledge, FIP is not free from mistakes. The answer for F1 Data for the AUSSI-ESS community is DataCite DOI resolution service. However, this could be a mistake in this FIP. The DataCite DOI resolution service should be taken as the answer for the question corresponding to the F3 principle. For this cell, it should be DOI.

[16]These were the only social science FIPs available as far as the authors were aware by the time the project commerce in spring 2023.

D. Assigning FIPs to DMP Questions

Next, we use the mapping in Section IV-B and find the correspondence between questions in FIPs and DMPs. We manually examined questions that correspond to the same FAIR principles and identified eight DMP questions that may be successfully connected to FIPs. However, one question (Question 6.3) was excluded due to its ambiguity.[17] Table II illustrates the correspondence between these DMP questions to the FIP questions after manual examination. Questions 4.6, 4.8, 4.9, and 4.13 have a clear one-to-one mapping to the corresponding FIP questions. The case of the next section is more complicated with more knowledge required by users. Question 5.2 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. The next section shows how answers in the FIPs are included in the customized interface.

V. Platform Customization: Building an Interface with the FAIR Wizard

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 migrate 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[18] (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.[19] 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 IV-D. 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.

Figure 2 illustrates that, below the question and its description, a two-column table that specifies the FICs and the

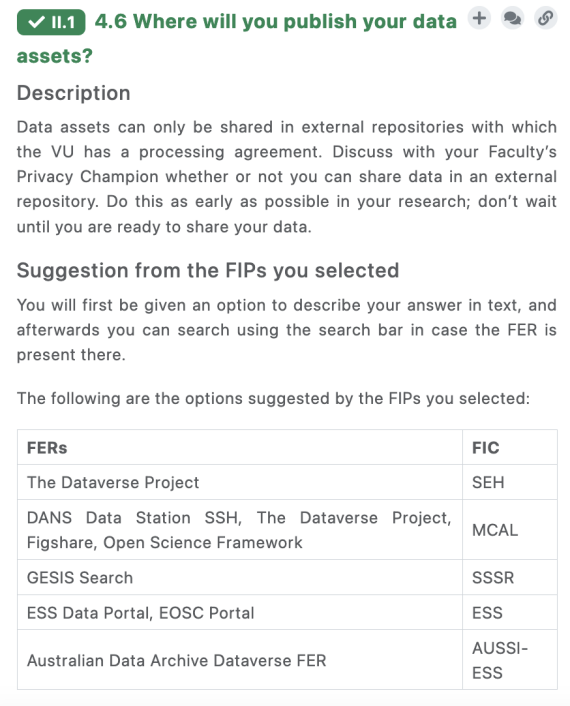

Fig. 2. FERs and corresponding communities based on user-selected FIPs in FAIR Wizard

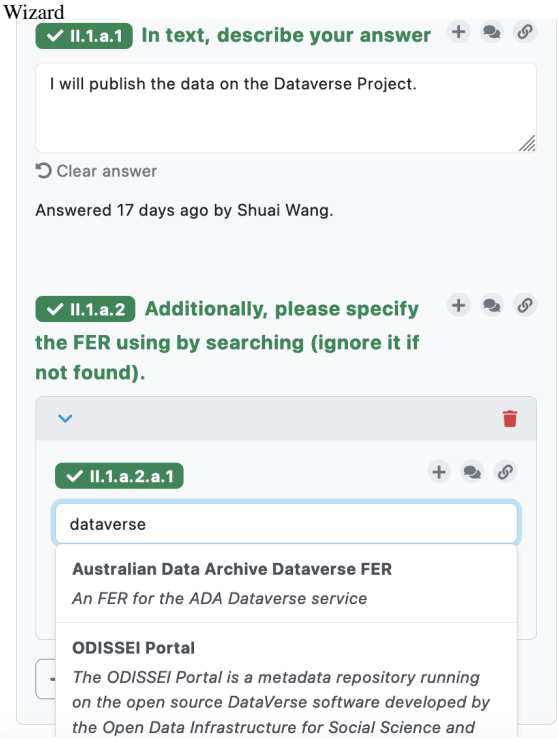

Fig. 3. User describing answer in text and selecting FER from a list of FERs using search engine in FAIR Wizard

FERs is provided as suggestions. The left column of the table presents all the decisions made by the communities, while the right column shows the corresponding research communities that made the decisions. For example, the table shows that SEH

[17] 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.

[18] https://palletsprojects.com/p/jinja/

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

uses the Dataverse Project to publish data assets. When the user agrees with the suggestion from a community, they can search for the corresponding FER as shown in Figure 3. In this case, after typing 'dataverse', related resources are displayed. A video demo of the interface is included in the supplementary material.

VI. User Study

After reaching out to almost all the researchers in the Faculty of Social Sciences of the VU[20], a total of six researchers agreed to participate in this study. These participants have written some DMPs or have sufficient knowledge of DMPs. Participants first pick one DMP that is closest to their research among four mock DMPs. These mock DMPs were inspired by some ideas of existing DMPs by university researchers in the past. Appendix A includes more details about these mock DMPs. The participants were then asked to complete the DMPs with selected questions by using the interface described in Section V. Finally, participants were asked to complete a survey that is designed for the understanding of participants' experiences and the effectiveness of the suggestions.

The survey consisted of twelve survey questions (SQs) in two parts (see Appendix B for the complete list of questions). In Part A (SQ1-SQ7), participants assessed the relevance and usefulness of the suggestions. More specifically, this section evaluated the relevance of communities for DMPs, assessed decision alignment with communities, and evaluated the effectiveness of suggestions for specific DMP questions, along with the ease of locating FERs. In addition, Part B (SQ8-SQ11) focused on the background and experience of the participants. They should indicate how long they have been in their academic career and how much experience they have 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 in comparison with other stakeholders (e.g. university requirements, journals). The user study was completed in October 2023.

VII. Evaluation

To address RQ3, in this section, we summarize the users' background in Section VII-A, the efficacy and significance of research communities in Section VII-B, and associating FERs with DMPs in Section VII-C.

A. Participants' Background and Experience

For our analysis, we used the answers of five participants[21]. 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 III). Some indicated the lack of knowledge of FIP and FER.

SQ ID	Survey question	Mean	Median	Std
SQ1	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
SQ2	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
SQ3	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
SQ7	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
SQ10	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 III
Survey questions and their results together the range, mean, median, and standard deviation.

B. The Efficacy of Suggesting Community Standards in DMPs

Five key survey questions and an analysis of the corresponding results are included in Table III. 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.

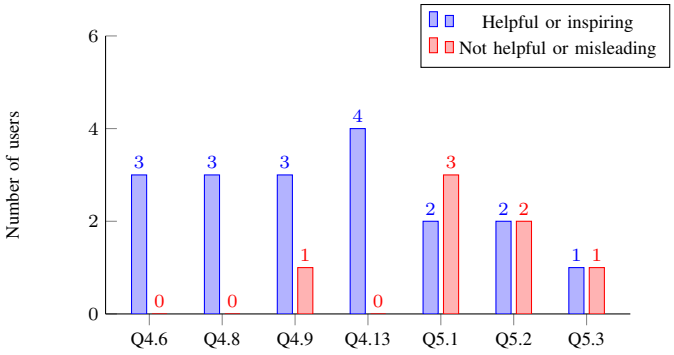

Fig. 4. Comparing the impact of suggestions on DMP questions: helpful or inspiring vs. not helpful or misleading, corresponding to SQ4 and SQ5 respectively.

Next, we evaluate the effectiveness of the suggestions for each DMP question (corresponding to SQ4 and SQ5). Figure 4 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

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

[21]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

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 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.

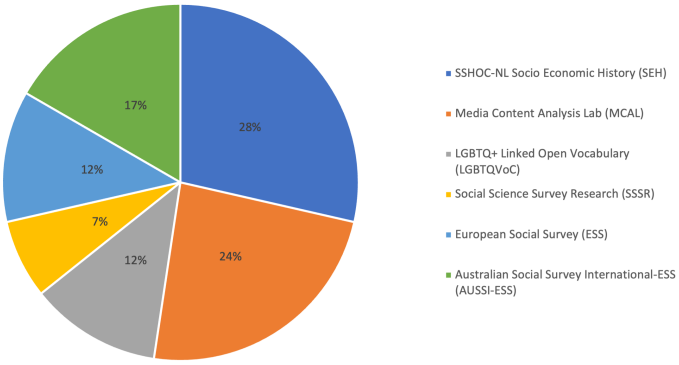

Fig. 5. Distribution of FERs among research communities chosen by participants.

In addition, Figure 5 shows that suggestions from communities can vary regarding their popularity: many FERs from SEH and MCAL communities were selected by participants while few FERs specific to the SSSR community were chosen. No conclusions could be drawn about the popularity of the FIPs given the small sample in FIPs and mock DMPs provided. This could be affected by the limited number of mock DMPs and their topics. The figure indicates the essence of the diversity of FIPs.

C. Associating FAIR Enabling Resources with DMPs

Next, we focus on participants' experiences with FERs. Only one participant reported successfully finding all the FERs they intended to specify (SQ6). The others reported missing FERs. However, when asked to rate how easy it was to find FERs using the search bar, the mean rating was 3.6 out of 5. This suggests that although most participants cannot find all the FERs desired through the search bar, they find the search bar easy to use to find resources. 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. Conversely, for Questions 4.9 and 5.1, some users struggled to locate the desired FERs through the search bar. This is consistent with the survey results in Section VII-B, as the participants find the suggestions for Question 5.1 not helpful or misleading. This was also reflected in SQ12: some are confused about the terminology; some are not familiar with metadata standards.

VIII. Discussion

The pilot study, along with the analysis presented earlier, provokes discussion on the advantages and limitations of this approach. The mapping between the questions in the VU DMP template and the FIP questions is not unique and can depend on the view of interpretation. This is because the chosen template did not take the FAIR principles into account by design. As a result, the suggestion provided can lead to confusion. Our analysis shows that out of the seven identified DMP questions, three questions (5.1, 5.2, and 5.3) face the problem of providing unhelpful or misleading suggestions. This is because the Questions 5.1 and 5.3 cover multiple issues and are only partially linked to the FIP questions, which require additional explanation. This calls for a review of some questions, especially 5.1 and 5.3. Moreover, both correspond to R1.2 where three FIPs lack resources to contribute to the suggestion. Question 4.9 assumes that the researchers will be registering their datasets on the PURE system (required by the university)[22]. However, the formation of the question excludes PURE as an answer, which could lead to a missing FER and can be a problem for automated analysis at a later stage. 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. Thus, a unified vocabulary for DMP and FIP is essential. The way FIPs interpret 'Accessibility' focuses mostly on machine accessibility. In contrast, DMPs focus more on human accessibility for repositories to make machine accessibility possible. This can lead to confusion and result in some missing pairs in the mapping. Take Question 5.3 for example, the word "access" in the DMP can be interpreted by the researcher as the software necessary to open the data (to view/edit) given the publishing format (e.g. *.dta files can be opened and used in STATA, .sav files in SPSS, etc.). However, for FAIR experts, it could mean the communication protocol (A1.1 Data). While the questions in FIP are very specific, it can be confusing when mapping DMP questions that are too general. The questions in DMPs are sequential and thus introduce context (e.g., Questions 4.6 and 4.9) for the interpretation of the latter question.

Despite the detailed introduction with used terminology explained on the first page of the user interface, some participants expressed in feedback that they found it difficult to comprehend certain terminologies and lack 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. However, in this study, participants cannot consult data stewards.

Furthermore, the current KM on the FAIR Wizard platform generates a fixed project template and lacks the functionality to customize the selection of FIPs. Thus, users are presented

[22]https://vu.nl/en/employee/research-data-support/research-portal-pure

with suggestions uniformly based on all available FIPs. This process could be improved, by which we mean, if users can choose their relevant research community beforehand, the suggestions could be clearer and more domain-specific.

Inspired by [12], we included in the survey a question (SQ11) about which stakeholders have the most 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. Community standards are important, but maybe not among the most influential 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.

When comparing our study to the research using the Leiden University's DMP template [3], we encountered the same challenge in addressing interoperability. Despite both studies identifying questions in their respective DMP templates that can be linked to FIP questions, we observed that our study, based on the VU DMP template, included a slightly larger subset of questions (7 compared to Leiden's 6). They proposed to use a "project importer" feature to pre-fill answers in the DMP using existing FIPs, which is different from our approach. Given the above-mentioned factors and multiple FIPs relevant for a DMP, pre-filling an answer could cause confusion (which could indicate mistakenly the best answer suggested by a chosen FIP). Both studies address several issues with respect to the DMP templates used.

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.

IX. Conclusion and Future Work

In conclusion, 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 seven DMP questions that could be effectively addressed using community standards captured by FIPs. 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. Therefore, some discussion was provided. 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.

In future research, we plan to enhance our KM with customization features, allowing users to pre-select the FIPs from which they wish to follow the decisions. As a proof-of-concept, our primary focus is 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. In future studies, mock DMPs could be replaced by researchers' own (past) DMPs. Initially, the authors were concerned that introducing FER and a new interface may result in a learning overhead. However, given that no participant complained about the interface or the search bar for finding FERs, this design needs not to be altered in future research.

Researchers' answers are often embedded in text in most existing DMP platforms. Essentially, the introduction of FER to DMP as answers is a step towards bringing DMPs to the ecosystem of linked data. Although the RDA DMP Common Standard has taken a significant step towards creating machine-actionable Data Management Plans (maDMPs),[23] there is still some work until DMPs become FAIR digital objects in the web. When fully integrated, the DMPs could then take suggestions from not only FIPs but also other types of resources, which in turn, can be used to inform the creation of new FIPs. Finally, a detailed review of the DMPs is essential for future research connecting DMPs with other resources. When compared with other DMPs [4], more questions could be added in the future to the VU DMP template, especially those concerning metadata, accessibility, and interoperability. A revised mapping between the updated VU DMP templates and FAIR principles and FIP questions could be published as it could not only be useful for taking FIPs as suggestions but also other digital assets in the research data management eco-system.

Appendix

A. Mock DMPs

For this user study, we newly developed four mock DMPs based on relevant existing DMPs, covering various research topics and facilitate potentials for the use of FERs and simulate real-world scenarios in research data management. To focus on the aspects under evaluation, irrelevant text was omitted in these mock DMPs. Participants could select a mock DMP on a research topic they are familiar with. Questions 4.6 and 4.8 have some pre-filled answers based on our understanding of the research topics to provide participants a starting point. That of 4.9, 5.1, 5.2, and 5.3 are not provided with suggestions but Question 4.13 has been pre-filled using the VU's default licensing option, which is CC-BY 4.0. These mock DMPs serve solely as a reference for the study. Participants have been instructed to make their own independent decisions in response to these questions.

[23]https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard

B. Questions in the Survey

Table IV shows the questions asked in the survey. For SQ1, 1 indicates that no community is relevant and 5 indicates that many communities are relevant. For SQ2, 1 indicates that it is not helpful. For SQ3, 1 indicates minimal alignment and 5 indicates complete. For SQ7, 1 indicates extremely difficult and 5 indicates extremely easy. For question 10, 1 indicates not clear at all and 5 indicates very clear. The remaining auxiliary questions are included in the supplementary material.

TABLE IV
QUESTIONS IN THE SURVEY

Survey Question ID	Survey Question
SQ1	On a scale of 1 to 5, how relevant are communities for this DMP?
SQ2	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.
SQ3	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?
SQ4	On which question(s) were the suggestions helpful or inspiring?
SQ5	On which question(s) were the suggestions not helpful or misleading?
SQ6	Did you find all the FAIR-Enabling Resources you wanted to specify in the list (using the search bar)?
SQ7	On a scale of 1 to 5, how easy was it for you to find the FAIR-Enabling Resource in the search bar?
SQ8	How many years have you been in academia (counting from the start of your PhD)?
SQ9	Have you written any DMP before?
SQ10	On a scale of 1 to 5, how clear was the goal of the study to you?
SQ11	If you were provided with suggestions from the following, whose suggestions will have an impact on your decision? You can select multiple choices from the list. Department, Faculty, and University Research Data Management, University I.T. Support, Data Management Platform, Data Repository Platform, University Legal and Policy team, Faculty Ethical Committee, International or EU Policy offices, Funders.
SQ12	During the process of creating this DMP, did you face any specific difficulties or challenges? If so, kindly provide details or specify the challenges you encountered.

REFERENCES

[1] M. D. Wilkinson *et al.*, "The fair guiding principles for scientific data management and stewardship," *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016. 1

[2] E. Schultes *et al.*, "Reusable FAIR Implementation Profiles as accelerators of FAIR Convergence," in *Advances in Conceptual Modeling* (G. Grossmann and S. Ram, eds.), (Cham), pp. 138–147, Springer International Publishing, 2020. 2, 4

[3] K. Hettne *et al.*, "Fip2dmp: Linking data management plans with fair implementation profiles," *FAIR Connect*, vol. 1, pp. 23–27, 01 2023. 2, 9

[4] P. 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. 2, 4, 9

[5] S. Mannheimer, "Toward a Better Data Management Plan: The Impact of DMPs on Grant Funded Research Practices," *Journal of eScience Librarianship*, vol. 7, no. 3, 2018. 2

[6] M. Grootveld *et al.*, "OpenAIRE and FAIR Data Expert Group survey about Horizon 2020 template for Data Management Plans," 2018. 10.5281/zenodo.1120245. 2

[7] I. Sullivan *et al.*, "Open and reproducible research on open science framework," *Current Protocols Essential Laboratory Techniques*, vol. 18, 2019. 2

[8] F. Schoots *et al.*, "Data management plan template leiden university (4.2)." Zenodo, 2021. https://doi.org/10.5281/zenodo.4423065. Accessed June 2023. 3

[9] T. Miksa *et al.*, "Ten principles for machine-actionable data management plans," *PLoS Comput. Biol.*, vol. 15, no. 3, p. e1006750, 2019. 3

[10] J. H. Spaaks *et al.*, "The data stewardship wizard: A tool bringing together researchers, data stewards, and data experts," *Int. J. Digital Curation*, vol. 14, no. 1, pp. 165–175, 2019. 3

[11] S. Wang *et al.*, "FAIR Implementation Profiles for Social Science," in *Proceeding of 17th Int Conf on Metadata and Semantics Research*, Springer, 2023. 4, 5

[12] L. Kvale *et al.*, "Understanding the data management plan as a boundary object through a multi-stakeholder perspective," *Int. J. Digit. Curation*, vol. 16, p. 16, 2021. 9