FIP2DMP: Linking data management plans with FAIR implementation profiles

Kristina Maria Hettne [a,*], Barbara Magagna [b], Alessa An Gambardella [c], Marek Suchánek [d], Fieke Schoots [e] and Erik Schultes [f]

[a] *Centre for Digital Scholarship, Leiden University Libraries, The Netherlands*
ORCID: https://orcid.org/0000-0002-4182-7560
[b] *GO FAIR Foundation, Leiden, The Netherlands*
ORCID: https://orcid.org/0000-0003-2195-3997
[c] *Faculty of Science, Leiden University, The Netherlands*
ORCID: https://orcid.org/0000-0002-4930-2662
[d] *Data Stewardship Wizard/CTU in Prague, Czech Republic*
ORCID: https://orcid.org/0000-0001-7525-9218
[e] *Health-RI, The Netherlands*
ORCID: https://orcid.org/0000-0002-4385-9312
[f] *GO FAIR Foundation, Leiden, The Netherlands*
ORCID: https://orcid.org/0000-0001-8888-635X

Abstract. FIP2DMP is a mapping resource between a FAIR Implementation Profile (FIP) and a Data Management Plan (DMP) template implemented in the Data Stewardship Wizard. The main benefit of this resource is to facilitate easier adoption of community choices from the FIP by a researcher in their DMP.

Keywords: Data Management Plan, FAIR implementation profiles, FAIR enabling resource, metadata standards, machine actionable DMPs

1. Introduction

With ever increasing volumes and complexity of data, data management planning has fast become a necessity at many universities. The planning process usually results in a Data Management Plan (DMP). By providing information about how data is handled during and after a project, a DMP serves both as useful support for researchers in

*Corresponding author: Kristina Maria Hettne, Ph.D., Universiteit Leiden Universitaire Bibliotheken Leiden, Leiden, Zuid-Holland, The Netherlands. E-mail: k.m.hettne@library.leidenuniv.nl; Tel.: +31 71 527 7281.

carrying out their research and as an informative document for those responsible for the sustainable operation of data management facilities at a university. In addition, a DMP is currently demanded by many funders (e.g. European Research Council, Wellcome trust, Dutch Research Council, just to name a few), often with an approved template and with particular focus on ensuring FAIR choices are being made by the researcher. Researchers are usually aided in preparing a DMP – and, in turn, their data management choices for their research project – by explanations in the DMP template itself, external Tips & Tricks documents [13] or workshops especially designed for filling out a DMP template (for example, at Leiden University researchers can follow a tailor-made workshop at their faculty or a general training which is given every 6 weeks).

Applying the FAIR principles to the management of research data requires researchers to choose resources and/or tools supporting the implementation of FAIR data. However, such choices are in practice most often done at the community level. Reflecting this, a FAIR Implementation Profile (FIP) is a list of technology choices, so-called FAIR Enabling Resources, declared by a community to implement each of the FAIR principles [12]. Taking this a step further, the FIP Wizard [6], a customized instance of the Data Stewardship Wizard (DSW) [1], makes it possible to create a machine-readable output of internal agreements represented as a nanopublication [9] containing an assertion linked to provenance and publication information for each declaration made by members of this community. As such, a community FIP itself is a FAIR enabling resource able to be reused by other communities or individual researchers within a community.

A DMP template similarly contains questions about how a researcher intends to manage their data, via different tools and resources. However, while many DMP templates aim to make sure that researchers produce FAIR data, there is to our knowledge no assessment that the resources mentioned in the resulting DMP which are used to make the data FAIR are actually themselves FAIR according to the FAIR principles [5]. In contrast, a FIP questionnaire is specifically designed to address each FAIR principle using FAIR Enabling Resources of principle-specific types. In addition, a DMP template is usually filled out by researchers that need to manage their data while FIP questionnaires are filled out by data stewards representing and acting in the name of their communities. Further, a DMP also has definitively a larger scope than a FIP, as it is about everything related to data management for an individual researcher or research project (including for instance secure storage or handling sensitive data).

Despite these differences, there is a clear link between DMPs and FIPs, where a FIP can inform the answers in a DMP. For example, one of the most common questions a researcher has when filling out a DMP template is "which metadata standard should I use?". This question corresponds to the question "Declaration F2: What metadata schema do you use for findability?" in the FIP questionnaire. Such overlapping questions implies that declarations made in a FIP could be used to automatically fill in the corresponding question in the DMP template. As a result, the researcher would not need to go through the process of finding or defining a standard themselves but rather can select a trusted, previously published FIP where community-chosen standards have already been defined.

In this paper, we present our first draft of the FIP2DMP mapping resource implemented in the Data Stewardship Wizard, which links the FIP questionnaire to the Leiden University DMP template. The main benefit of such a linking would be to facilitate easier adoption of community choices – from the FIP – by a researcher – in their DMP.

2. Implementation

As a pilot, we implemented the Leiden University DMP template [11] as a new knowledge model [3] (Fig. 1) in an instance [2] of the DSW kindly provided by the DSW team. Then a conceptual mapping [7] was made between the questions in the Leiden University DMP template and the questions in the FIP Wizard questionnaire [8] (Fig. 2). Where possible, questions in the DMP template were added or adjusted to correspond to questions in the FIP Wizard using integration questions that support the integrated service nanopublication API search. The Knowledge Model (KM) [10] (Fig. 3) of the optimized DMP template provides FAIR enabling resources implemented as a nanopublication as answer options. For example, Dublin Core as FAIR Enabling Resource (FER) has been defined as a nanopublication [4]. See Fig. 4 for an example of how a DMP question with a FER integration would appear. By substituting multiple choice questions into list questions in the Knowledge Model, other options that are not representable as nanopublications can be provided through controlled vocabularies making the resulting DMP more machine-readable overall.
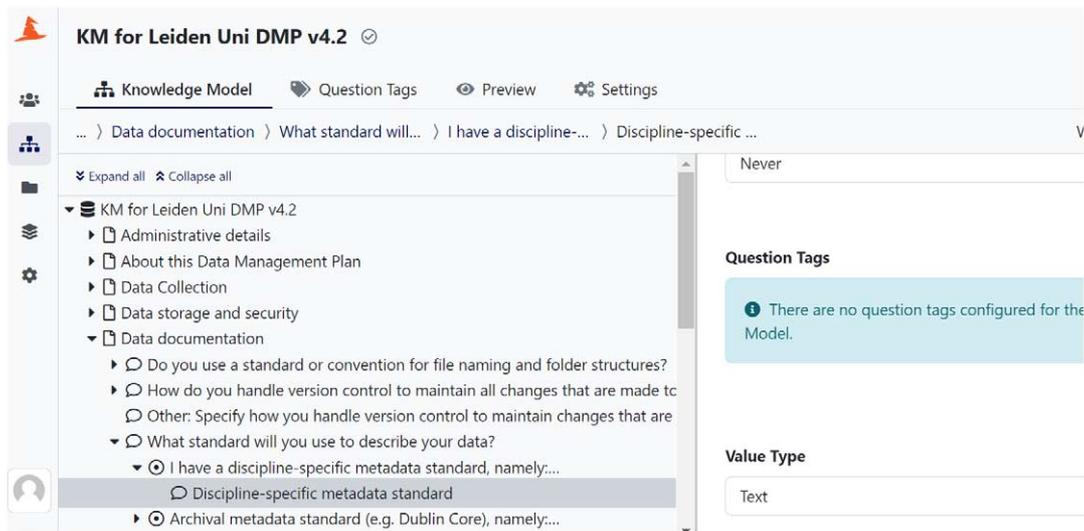
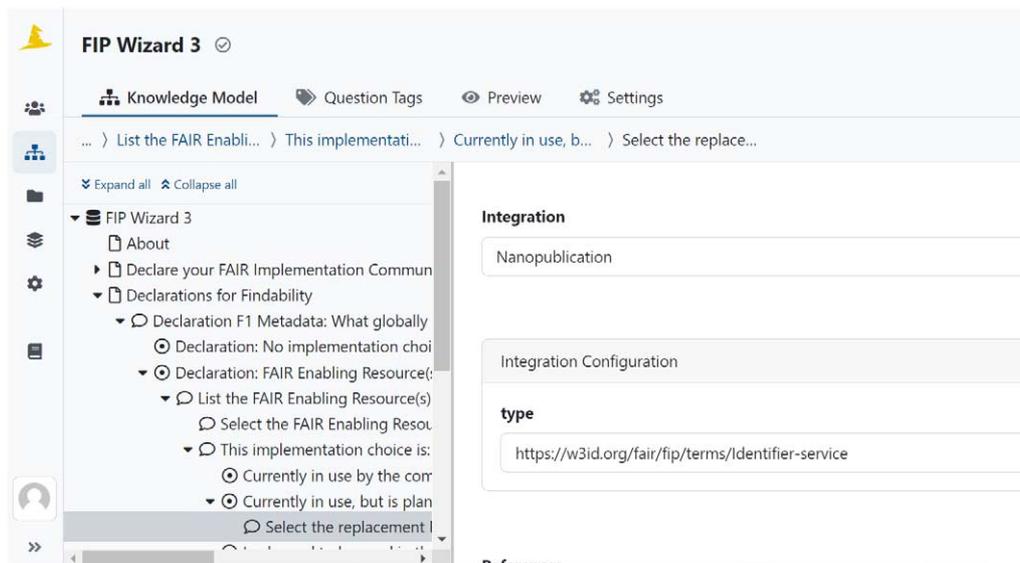Fig. 1. Screenshot of Leiden university DMP knowledge model as implemented in the DS wizard.

Fig. 2. Screenshot of the FIP wizard knowledge model as implemented in the DS wizard.

The mapping between questions of the DMP template and those of the FIP questionnaire then serves as a pre-scription for how to transform the decision from a FIP into a DMP. The DSW has a "project importer" feature that can be used to realize such a transformation. Basically, a user creates a DMP project in the DSW and decides to pre-fill it using an existing FIP. Then the user logs into the FIP Wizard to select the FIP from there during the import procedure. After providing a FIP, the importer goes through the FAIR choices in there and, according to the mapping, adds replies to the DMP project. Such an import can be done with either an empty DMP project or with already existing and (partly) filled DMP templates as well. Although an import may overwrite some of the previous replies, all changes are recorded in the version history in the DSW and can be easily reverted when needed.
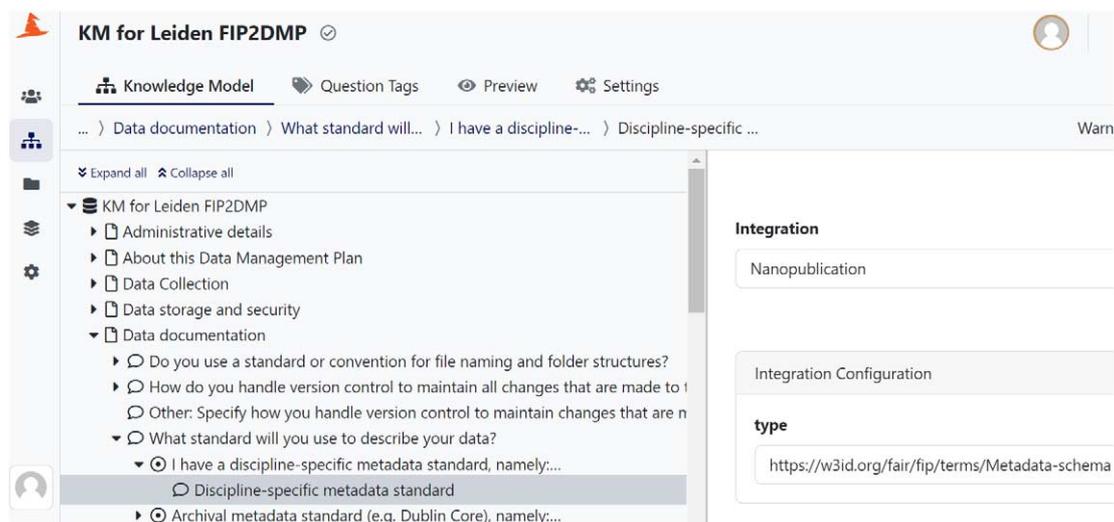
Fig. 3. Screenshot of the optimized knowledge model for Leiden FIP2DMP as implemented in the DS wizard showing the nanopublication service integration.
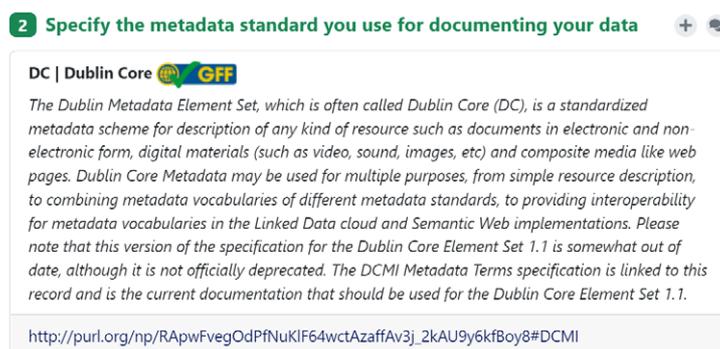

Fig. 4. A DMP question about metadata standards answered by the FAIR enabling resource Dublin core, corresponding to the FIP question "declaration F2: what metadata schema do you use for findability?". The name and the description of the resource are displayed and the persistent identifier of the resource is provided as a nanopublication at the bottom.

3. Discussion

There is a considerable overlap between DMPs and FIPs, which has encouraged us to build a more explicit link between the two question lists. The technical implementation of this linkage depends on the system wherein the FIP and DMP is created. Also, DMP templates cover roughly the same topics but are different in their layout and accuracy of specific questions. Starting with the Leiden University DMP template in the DSW tool, we have implemented a technical link between DMPs and FIPs by mapping questions used in the two respective DSW instances. In addition, we have modified the Knowledge Model of the DMP template by including nanopublication integration questions for FERs as answer options. Currently, mapping the questions in a DMP template to the questions in the FIP questionnaire is a manual process which requires knowledge about DMPs and knowledge about FIPs. We are implementing the mapping technically by using the "project importer" feature to facilitate this process but are continuing to explore different mapping options between DMPs and FIPs. This will enable the Wizard tool to more generally and automatically pre-fill those overlapping questions by selecting an existing FIP. Future work will extend the mapping to the core knowledge model and other templates implemented in the DSW but also within other tools like DMP Online.

It should be noted that the Leiden University DMP template in the DSW is only intended for this pilot project on linking DMPs and FIPs for now. Researchers of Leiden University that need to prepare their DMP should do that by filling out the conventional Microsoft Word DMP template available online [11].

Funding

The authors report no funding.

Disclosures and conflict of interest

All authors report no conflict of interest. Barbara Magagna and Erik Schultes were not involved in the peer-review process.

References

[1] Data Stewardship Wizard, available from: https://ds-wizard.org/ (accessed: 03.12.2022).
[2] Data Stewardship Wizard instance of the University of Leiden, available from: https://leiden.ds-wizard.org/ (accessed: 03.12.2022).
[3] DSW Knowledge Model for Leiden University DMP, available from: https://osf.io/q73gy (accessed: 03.12.2022).
[4] Dublin Core Metadata Set represented as a nanopublication, available from: http://purl.org/np/RAS-6QtTBMbe3fPNvx0IB4l3qrNDhJTY duN57IR_BS59M (accessed: 03.12.2022).
[5] FAIR principles, available from: https://www.go-fair.org/fair-principles/ (accessed: 03.12.2022).
[6] FIP Wizard, Available from: https://fip-wizard.ds-wizard.org/ (accessed: 03.12.2022).
[7] FIP2DMP Leiden_conceptual mapping. available from: https://osf.io/5jsfp (accessed: 03.12.2022).
[8] Knowledge Model for FIP Wizard V3, available from: https://osf.io/5py8a (accessed: 03.12.2022).
[9] Nanopublications, available from: https://nanopub.org/ (accessed: 03.12.2022).
[10] Optimised DSW Knowledge Model for Leiden University DMP, available from: https://osf.io/jmdk9 (accessed: 03.12.2022).
[11] F. Schoots, M. van den Berk, K. Hettne and J. Yeomans, Data Management Plan template Leiden University (4.2), Zenodo, 2021. doi:10.5281/zenodo.4423065.
[12] E. Schultes, B. Magagna, K.M. Hettne, R. Pergl, M. Suchánek and T. Kuhn, Reusable FAIR implementation profiles as accelerators of FAIR convergence, in: *Advances in Conceptual Modeling. ER 2020*, G. Grossmann and S. Ram, eds, Lecture Notes in Computer Science, Vol. 12584, Springer, Cham, 2020. doi:10.1007/978-3-030-65847-2_13.
[13] J. Yeomans, M. van den Berk, F. Schoots and K. Hettne, Tips and tricks for writing a Data Management Plan Leiden University. Zenodo, 2021. doi:10.5281/zenodo.4312894.