

FAIR Supporting Resource (FSR) Description Article

Data Stewardship Wizard

Rob Hoofst^a, Marek Suchánek^{b,*} and Robert Pergl^c

^a *Dutch Techcentre for Life Sciences, Health-RI, ELIXIR, The Netherlands*

ORCID: <https://orcid.org/0000-0001-6825-9439>

^b *Faculty of Information Technology, CTU, Prague, Czech Republic*

ORCID: <https://orcid.org/0000-0001-7525-9218>

^c *Faculty of Information Technology, CTU, Prague, Czech Republic*

ORCID: <https://orcid.org/0000-0003-2980-4400>

Editor: Barbara Magagna (<https://orcid.org/0000-0003-2195-3997>)

Solicited reviews: Ellen Carbo (<https://orcid.org/0000-0002-7125-728X>); Angelica Maineri (<https://orcid.org/0000-0002-6978-5278>)

Received 22 January 2023

Accepted 24 January 2023

Abstract. Data Stewardship Wizard is a tool that facilitates the process of Data Management Planning / Data Stewardship Planning with a focus on the benefit to the research at hand.

Keywords: Questionnaire, expert system, data management, FAIRness assessment, project management activities

Data Stewardship Wizard (DSW) is a tool that is meant to use Data Management Planning to bring the benefit of a wide variety of proven data management processes to researchers and their research projects, and not just to satisfy the request of research funders. This feature sets it apart from tools which are meant to make data management plans (DMPs) [6]. DSW is available at <https://ds-wizard.org/>.

Important features of DSW for FAIR Connect are:

- It helps connect researchers and data stewards with relevant RDM services and resources or tested practices that they may not have known about.
- It guides its users towards FAIR data early, so that they can benefit from FAIR data during their projects.
- It guides its users towards techniques that result in more FAIR resources after the project.
- In the end it can summarize the FAIRness of the selected answers.

The result of a data stewardship planning process in DSW can be formed into a DMP following several different funder templates, as well as a document following the RDA common standard for machine-actionable DMPs [9]. Data Stewardship Wizard and its philosophy are described in an earlier paper [8]. A screenshot of DSW in action is

*Corresponding author: Marek Suchánek, FIT CTU in Prague, Thákurova 9, Prague 6, Czech Republic. E-mail: marek.suchanek@fit.cvut.cz; Tel.: +420 732 647 645.

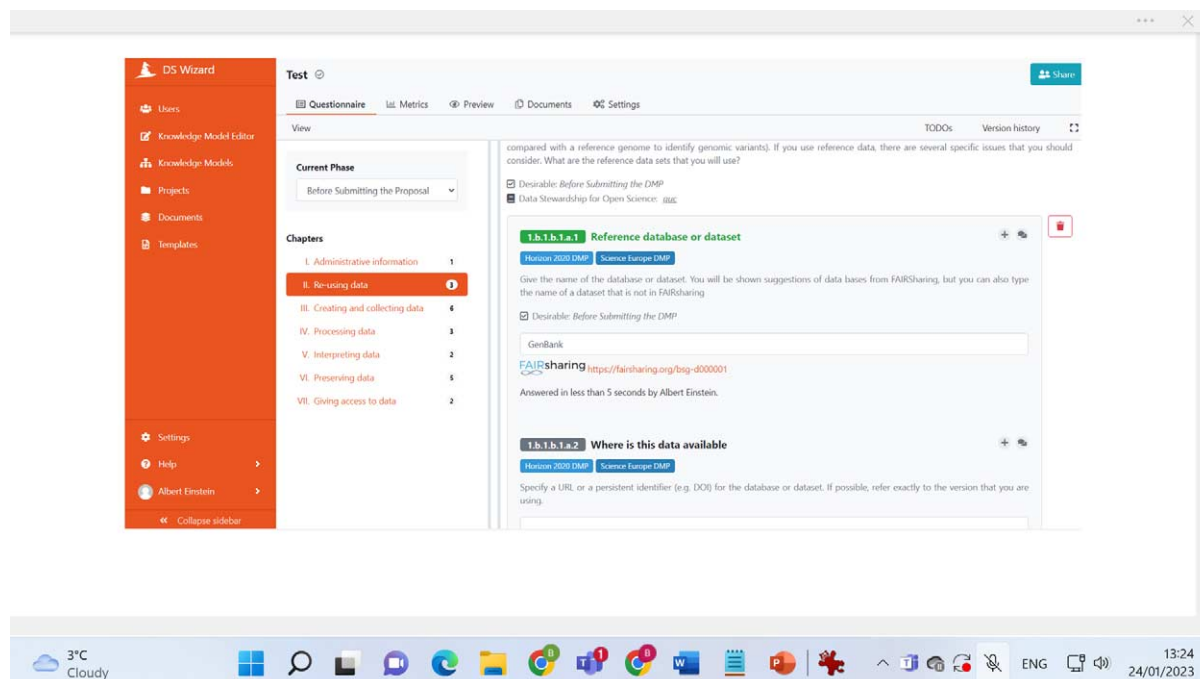


Fig. 1. Screenshot of Data Stewardship Wizard questionnaire, showing a question that gets the possible answers using an “integration” with another service.

shown in Fig. 1. DSW is bi-directionally connected to ELIXIR’s RDMkit [12]: DSW can help users of RDMkit to make a selection between different options, whereas RDMkit can give users of DSW an overview of considerations.

DSW is a tool originally developed as a collaborative project between the ELIXIR nodes in the Czech Republic (ELIXIR-CZ) and The Netherlands (ELIXIR-NL). ELIXIR-CZ delivered the software code and the technical platform, and ELIXIR-NL contributed collected data stewardship expertise. Since 2015, the tool evolved from a simple questionnaire generated from the JSON-encoded mind map of data stewardship expertise [5] (the initial knowledge model) to a full-fledged solution for data management planning. Instances of DSW can be deployed in different ways (e.g. self-hosted by an institute or infrastructure, with or without using containers, or hosted as a service). DSW can be customized at many levels to accommodate the local needs – by adjusting the look and feel, modifying and adding document templates, addition of integrations with local systems, changing the knowledge model with localized questions, adding or modifying possible answers, and tailoring the guidance to include local choices and pointers to local experts.

Researchers or project data stewards do not have to work alone on data management planning: a group of people from a project can collaboratively explore the data stewardship knowledge, perform the actual planning for their projects, and eventually get a DMP in desired template and format. The collaboration works similar to well-known Google Documents; each contributor can have a view, comment or edit role – and the project in DSW is “live” from two points of view: First, when multiple people work on the planning at the same time, any comments and their replies are directly shown to others that collaborate. Second, the system encourages users to keep the planning information up-to-date and allows them to track the history of changes. To kickstart the planning process, there is the possibility to import information about the project from a custom source via an importer. This could for instance be the RDA maDMP importer that allows to pre-fill the questionnaire from a machine-actionable DMP encoded in JSON.

Institutional or community data stewards can similarly collaborate on building and improving the knowledge models (KMs) using the built-in KM editor. This allows them to reflect specialized domain knowledge and local needs/expertise. Once a new version of any KM is released, users can easily update and migrate their projects to the latest version in order to be able to apply the latest questions, answers and guidance to their project. This migration

process further encourages those responsible for data stewardship in a project to continue planning throughout the project. The same migration process also can be applied to localized knowledge models, to propagate all changes made in the general parent knowledge model to the localized children. Aside from the guidance in the knowledge model, data stewards responsible for any instance of the DSW may prepare project templates for researchers, where they set default document templates, prefill some answers, or provide additional guidance in comments.

The flexibility of DSW allows it to be used in various tool assemblies for data management, e.g., in the Norwegian e-Infrastructure for Life Sciences (NeLS) [7]. In addition, there are already non-DMP related use cases; for example, DSW serves as a form tool to allow FAIR communities to define their FAIR Implementation Profiles (FIP) [10] and submit them as nanopublications [4]. Other non-DMP use cases of DSW, namely Software Management Plans (SMP) [11] and Data Protection Impact Assessment (DPIA) [2], were developed during BioHackathon Europe 2021 [1].

DSW is still being developed and further improved following the community needs [3]. The development is both community- and project-driven. It is an open-source project with Apache 2.0 license, which makes it possible for anyone to make their own changes and either exclusively use the changed version locally or contribute the changes back to the project.

Acknowledgements

The authors have no acknowledgements to report.

Funding

The authors have no funding to report.

Conflict of interest

The authors have no conflict of interest to report.

References

- [1] Biohackathon Europe 2021. Available at <https://2021.biohackathon-europe.org/>. Accessed October 10, 2022.
- [2] Data Protection Impact Assessment Project. Available at <https://github.com/elixir-europe/bioHackathon-projects-2021/tree/main/projects/18>. Accessed October 10, 2022.
- [3] DSW Community. Available at <https://ideas.ds-wizard.org>. Accessed October 10, 2022.
- [4] FIP Wizard. Available at <https://fip-wizard.ds-wizard.org/>. Accessed October 10, 2022.
- [5] R.W.W. Hooft, Data Stewardship Mindmap (Version 20190725). Zenodo, 2019. doi:10.5281/zenodo.3351949.
- [6] S. Jones, R. Pergl, R. Hooft, T. Miksa, R. Samors, J. Ungvari, R.I. Davis and T. Lee, Data management planning: How requirements and solutions are beginning to converge, *Data Intelligence* 2(1–2) (2020), 208–219. doi:10.1162/dint_a_00043.
- [7] Norwegian e-Infrastructure for Life Sciences (NeLS). Available at https://rdmkit.elixir-europe.org/nels_assembly.html. Accessed October 10, 2022.
- [8] R. Pergl, R. Hooft, M. Suchánek, V. Knaisl and J. Slifka, “Data Stewardship Wizard”: A tool bringing together researchers, data stewards, and data experts around data management planning, *Data Science Journal* 18(1) (2019), 59. doi:10.5334/dsj-2019-059.
- [9] RDA DMP Common Standard. Available at <https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard>. Accessed October 10, 2022.
- [10] 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.
- [11] Software Management Plan. Available at <https://github.com/elixir-europe/bioHackathon-projects-2021/tree/main/projects/20>. Accessed October 10, 2022.
- [12] The Research Data Management toolkit for Life Sciences. Available at <https://rdmkit.elixir-europe.org/>. Accessed October 10, 2022.