International FAIR Digital Objects Implementation Summit 2024

Extended Abstracts

https://doi.org/10.52825/ocp.v5i.1045

Published: 18 Mar. 2025

FAIR Digital Objects for Seamless Research Data Management for Researchers and Higher Education Institutions

Markus Stocker[1,2], Björn Grüning[3], Tomasz Miksa[4], Claudia Biniossek[1,3,5], and Dirk Betz[1,3,5,*]

[1]TIB - Leibniz Information Centre for Science and Technology, Hannover, Germany

[2]Knowledge-based Systems Department, Institute of Data Science, Leibniz University Hannover, Germany

[3]Bioinformatics Group, Institute for Computer Science, Albert-Ludwigs-University of Freiburg, Germany

[4]TU Wien & SBA Research, Vienna, Austria

[5]Center for Empirical Research in Economics and Behavioral Sciences, University of Erfurt, Germany

*Correspondence: Dirk Betz, dirk.betz@tib.eu

Abstract. *Seamless Research Data Management for Researchers* aims to cover a complete scientific workflow from planning a research project to registration and publication of results in repositories by connecting existing components, services, and tools using FDOs. This approach combines widely used components, so large data volumes can increasingly be FAIR-ified automatically. Machine-actionable Data Management Plans (maDMP) that comprehensively document the respective research project in a machine-actionable format form the entry point. The familiar Galaxy environment, which already enables RO-Crate implementation, forms the backbone to incorporate a growing number of services and tools. Galaxy orchestrates and executes the workflow components resulting from maDMPs and data analysis. The research results and comprehensive documentation become published in a repository of the researchers' choice (e.g., Zenodo). From there, the research results can be integrated into a knowledge graph (e.g., ORKG).

Keywords: Fair Digital Objects (FDO), Research Object Crates (RO-Crates), Machine Actionable Data Management Plans (maDMP), Research Data Management

1. Integration of Digital Object Architecture (DOA) and RO-Crates

Currently, several alternative frameworks leverage the potential of FAIR Digital Objects (FDOs). We focus on the Research Object Crates (RO-Crates) approach [1], but ensure compatibility with the other approaches by using the minimal unifying mechanism of the Digital Object Interface Protocol (DOIP), which supports the cross-approach interaction of FDOs. The widespread use of RO-Crates and the benefits provided have not yet been fully exploited. The FDO based seamless research data management (RDM) approach will address the interplay of DOA/DOIP based and "webby" FDOs with regard to the convergence of the service ecosystem (use of registries for types, operations, etc.) and the flexibility of multi-primary administrators and registration agencies regarding the required changes to the PID metadata profile. The project "FDOs for seamless RDM for researchers and higher education institutions" is described below and intends to address these shortcomings.

2. Seamless Research Data Management

2.1 Canonical Workflow Framework for Research

The *FDOs for seamless RDM* project aims to implement the FDO-CWFR (Canonical Workflow Framework for Research) [2] to enable seamless RDM for the scientific community [3].

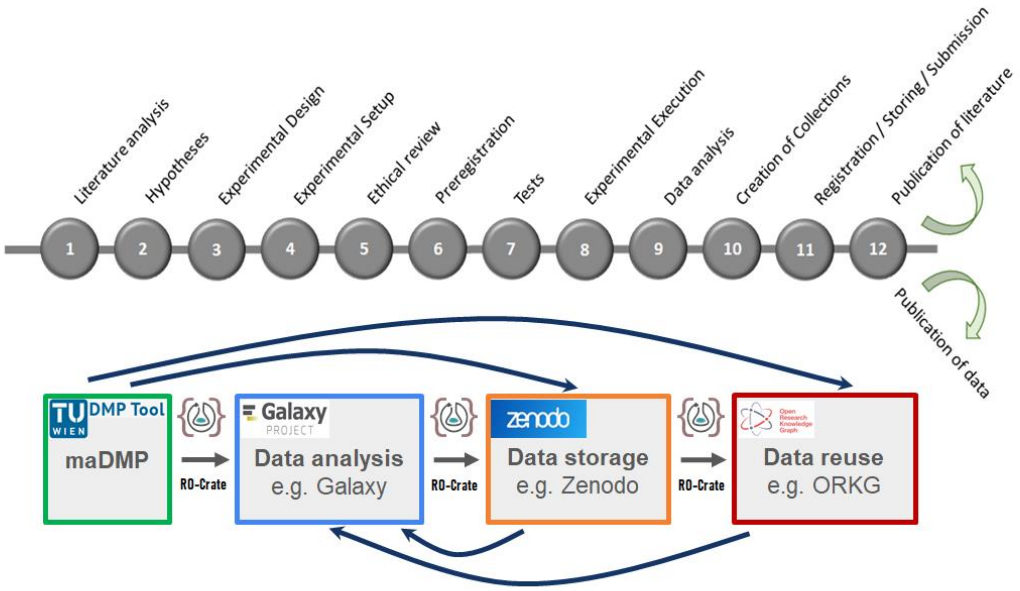

Figure 1. *RO-Crate implementation within a typical experimental research workflow (adapted and modified from [4])*

The approach will enable the composability of generic tools and services, which are often already "market leaders" in their respective fields (e.g. TU Vienna maDMP Tool, Galaxy, Zenodo, Invenio, Open Research Knowledge Graph [ORKG]), by exchanging the data produced and consumed along a workflow in FDO-compliant syntax using the DOIP and making it accessible in repositories. The approach envisages using the RO-Crate implementation of the FDO concept after it has been extended so that it interacts compatible with the basic FDO infrastructure of the overall ecosystem. However, the integration of Digital Object Architecture (DOA) and "webby FDOs" is challenging as it often requires the inclusion of new practices such as FAIR Signposting to implement the binding of digital objects based on web linking. The solution presented in this paper strikes a balance between web-based and DOA-based concepts, as it builds on DOIP for cross-approach interaction of FDOs, and additionally tests the resolving of the FDO record to an RO-crate.

2.2 Registration of FDOs as metadata

Two methods will be tested: (1) the publication of RO-Crate FDOs as files in repositories (e.g. Zenodo, Invenio, GWDG repository) and the DOI-based persistent identification of this record; and (2) the registration of RO-Crate FDOs as metadata of a PID via a PID provider (e.g. ePIC, the European Persistent Identifier Consortium). The first procedure is well established and already in use but only identifies the record (containing the RO-Crate FDO), while in the second variant the RO-Crate forms the PID metadata. In the second approach, however, the PID provider must support metadata profiles.

2.3 Modular and scalable approach

The modular approach developed ensures seamless, efficient, and effective production and use of FDOs and the data embedded in FDOs throughout the research lifecycle. Each module can receive FDOs via the DOIP and output them as FDOs. Researchers can assemble any number and type of modules as their requirements demand. The FDOs and DOIP form the glue of the open, distributed system. Only by using FDOs as the connecting component the potential of modular workflow components (CWFR) can be exploited widely among researchers and the tools they already use. At the same time, researchers and university staff are relieved from the increasing burden of data wrangling.

With this approach, which combines widely used components, large amounts of data can increasingly be FAIRified automatically. Machine-actionable Data Management Plans (maDMPs) are used as an entry point to document the respective research project comprehensively in a machine-readable format [5]. As a starting point, the familiar Galaxy environment will be the introductory tool as it already allows RO-Crate implementation. Galaxy is used to orchestrate and execute the workflow components resulting from maDMPs and data analysis. The research results are published together with the extensive documentation in a repository of the researchers' choice, to which an interface has been implemented (initially Invenio-based repositories such as Zenodo, plus the GWDG repository). From there, further steps can be taken, such as the integration of research results into a knowledge graph (e.g. ORKG) [6].

3. Conclusion

The importance of this sub-project for the scientific system is substantial, as it is designed to ensure the integration of actual users from the large national and international communities. No new services or tools are being developed. Instead, as a first step, widely used services such as Galaxy, Zenodo, and ORKG will be adapted to the basic FDO infrastructure so that they can automatically generate, exchange, and consume FDOs in the future. In further steps of the project, the possibility of connecting further (widely used) tools will be ensured gradually.

Data availability statement

Not applicable.

Underlying and related material

Not applicable.

Author contributions

Conceptualization (MS, BG, TM, CB, DB); Data curation (N/A); Formal Analysis (N/A); Funding acquisition (N/A); Investigation (N/A); Methodology (N/A); Project administration (N/A); Resources (N/A); Software (N/A); Supervision (N/A); Validation (N/A); Visualization (N/A); Writing – original draft (MS, BG, TM, CB, DB); Writing – review & editing (MS, BG, TM, CB, DB).

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Acknowledgements

We would like to thank the anonymous reviewers for their valuable comments.

References

[1] Soiland-Reyes, S., Sefton, P., Castro, L.J., Coppens F., Garijo, D., Leo, S., Portier, M., & Groth, P. (2022). Creating lightweight FAIR Digital Objects with RO-Crate. *1st International Conference on FAIR Digital Objects (FDO 2022)* (Poster). Research Ideas and Outcomes 8:e93937. https://doi.org/10.3897/rio.8.e93937. https://s11.no/2022/phd/fdo-with-ro-crate/

[2] The CWFR Group (2020). Canonical Workflow Framework for Research. CWFR - Position Paper – Version 2. Eds.: A. Hardisty & P. Wittenburg. https://orca.cardiff.ac.uk/id/eprint/142465/3/CWFR-position-paper-v3.pdf

[3] Betz, D., Biniossek, C., Blanchi, C., Henninger, F., Lauer, T., Wieder, P., Wittenburg, P., & Zünkeler, M. (2022). Canonical Workflow for Experimental Research. *Data Intelligence,* 4 (2): 155–172. https://doi.org/10.1162/dint_a_00123

[4] Jeffery, K., Wittenburg, P., Lannom, L., Strawn, G., Biniossek, C., Betz, D., & Blanchi, C. (2021). Not Ready for Convergence in Data Infrastructures. *Data Intelligence,* 3 (1): 116–135. https://doi.org/10.1162/dint_a_00084

[5] Miksa, T., Suchánek, M., Slifka, J., Knaisl, V., Ekaputra, F. J., Kovacevic, F., Ningtyas, A. M., El-Ebshihy, A. M., & Pergl, R. (2023). Towards a Toolbox for Automated Assessment of Machine-Actionable Data Management Plans. *Data Science Journal*, *22*, Article 28. https://doi.org/10.5334/dsj-2023-028

[6] Stocker, M., Oelen, A., Jaradeh, M. Y., Haris, M., Oghli, O. A., Heidari, G., Hussein, H., Lorenz, A.-L., Kabenamualu, S., Farfar, K. E., Prinz, M., Karras, O., D'Souza, J., Vogt, L., & Auer, S. (2023). FAIR scientific information with the Open Research Knowledge Graph. In B. Magagna (Ed.), FAIR Connect (Vol. 1, Issue 1, pp. 19–21). IOS Press. https://doi.org/10.3233/fc-221513