

Agile Research Data Management with FDOs using LinkAhead

Timm Fitschen[‡], Henrik tom Wörden[‡], Alexander Schlemmer^{§‡}, Florian Spreckelsen[‡], Daniel Hornung[‡]

[‡] IndiScale GmbH, Göttingen, Germany

[§] Max-Planck-Institut für Dynamik und Selbstorganisation, Göttingen, Germany

Corresponding author: Henrik tom Wörden (h.tomwoerden@indiscale.com)

Abstract

One essential question with regard to the implementation of FAIR (Wilkinson et al. 2016) Digital Objects (FDOs) in everyday research is the following: How is data that is acquired in some way transformed into FDOs? Creating FDOs from data is a two-fold problem: "FAIR principles are policies, whereas the digital objects are technical abstractions" (Schwardmann 2020). Regarding the technical side, in order to become FDOs, raw data stored in files and databases have to be bundled with their metadata and PIDs have to be assigned. With good tools at hand, sharing data as an FDO with others might only be a matter of a few mouse clicks -- if the metadata is readily available.

However, the process of collecting metadata comes with significant challenges of its own.

While sometimes necessary, the manual annotation with metadata is error-prone and time-consuming. Due to resource constraints and time pressure, researchers might skip this task whenever it does not have any direct benefits for their work in the time frame of their current project. The experience with existing data repositories tells us that adding metadata at a late stage in the research data life cycle (for example just before publication) delays the problem in the best case. In the worst case, important information has already been lost at that stage. Furthermore, there are the FAIR principles which, for researchers, mean more rules to follow and thus more time spent on data management.

Research data should be enriched with FAIR metadata as early as possible to ensure that the research data is FDO-ready when needed. In order to do this, researchers need tools that assist them with the task of making their data FDO-ready and those tools must not hinder the research process but in the best case even promote it. This means that the drawbacks of making data FDO-ready need to be mitigated and compensated by direct benefits to researchers.

In this contribution, we present how early-on FDO-readiness can be achieved with the open source research data management toolkit [LinkAhead](#) and how researchers profit

from the FDO-readiness directly in their work. LinkAhead, a [CaosDB](#) (Fitschen et al. 2019) distribution, assists its users from the very first steps of data acquisition to the completion of FDOs and data publication by means of a semantic data modal, metadata annotation to raw data and a powerful search capabilities.

Why would researchers do what is necessary to make their data FOD-ready, early-on?

With LinkAhead, the FDO-readiness is a welcome side-effect for users. Even though LinkAhead cannot magically generate all relevant metadata and make data FAIR, LinkAheads allows the automation of the process where possible and assists users elsewhere. The inevitable additional work for researchers is reduced as well as compensated with new possibilities for users to work with their data. Thus users are nudged into storing their data in clear and understandable structures and into annotating their data with high-quality metadata. We will highlight in the following, how users benefit from early FDO-readiness in their daily work due to those characteristics of LinkAhead. LinkAhead adapts to the changing needs of the researchers. It thus allows research data management to be an agile process and ensures that researchers can efficiently conduct their daily work. At the same time, it supports the development, documentation and observance of standards which is vital for the commensurability, reusability, and reproducibility of research findings. LinkAhead is designed to be the first tool after data acquisition and the last tool before the publication of data. It can be fed with data from LIMS, ELNs, simulation and analysis software, helps with automation of workflows, and manages raw data in files.

Which direct benefit can LinkAhead offer to its users if they do what is necessary to make FDOs from data?

When searching for data in general or FDOs in particular researchers can employ metadata and the connections among data in order to find what they are looking for. Thereby, browsing the data for example in the LinkAhead web interface can be very targeted. Additionally, these search capabilities can be used within analysis workflows in order to create the correct basis of data using FDOs directly for the question at hand. Client libraries, like the Python client, allow to include this into automated analyses. Since manual data insertion is inefficient in many research environments, LinkAhead does not only offer the insertion of data via web forms, but encourages the usage of automatic processes like the LinkAhead Crawler. While metadata should be added already during this insertion step if possible, LinkAhead assists in completing metadata after the initial insertion in order to strike a balance between interrupting the research workflow and running into the above mentioned challenges when adding metadata too late. This automatic data insertion process is highly customizable and allows to complement data as soon as possible such that FDOs are constituted.

The semantic data model of LinkAhead allows researchers to use ontologies of their domain but also to extend those where necessary for the work at hand. This allows an agile adaption to changed requirements or new challenges and LinkAhead assures compatibility with old data if possible. The data model can capture both relations within

an FDO, among metadata, data, and possibly files and references to other FDOs directly within LinkAhead or FDOs stored elsewhere using PIDs. The semantic data model and additional constraints facilitate the creation and validation of FDOs.

LinkAhead allows to seamlessly integrate the FDO concept into the workflows of LinkAhead users. Thereby collecting information necessary for FDOs is not a burden for the researcher but the information can be directly used. The search capabilities of LinkAhead can employ the metadata of FDOs and the connections to other FDOs and their metadata. Components of the LinkAhead toolkit like the web interface allow users to access FDOs in an intuitive way while the LinkAhead API allows the direct use of data (and FDOs) in automatic processing and analysis. Thus, LinkAhead is a tool which does not only assist in the technical process of creating digital objects, it also creates incentives for its target users to adhere to FAIR guiding principles. It brings the benefits of FDOs to the people who have to do the extra work. Strategically, this is of utmost importance if the FDO initiative as a whole is to succeed.

Keywords

research data management, CaosDB, FAIR, metadata, open science, open source, FOSS

Presenting author

Henrik tom Wörden

Presented at

First International Conference on FAIR Digital Objects, presentation

Conflicts of interest

References

- Fitschen T, Schlemmer A, Hornung D, tom Wörden H, Parlitz U, Luther S (2019) CaosDB —Research Data Management for Complex, Changing, and Automated Research Workflows. *Data* 4 (2). <https://doi.org/10.3390/data4020083>
- Schwarzmann U (2020) Digital Objects – FAIR Digital Objects: Which Services Are Required? *Data Science Journal* 19 <https://doi.org/10.5334/dsj-2020-015>
- Wilkinson M, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J, da Silva Santos LB, Bourne P, Bouwman J, Brookes A, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo C, Finkers R, Gonzalez-Beltran A, Gray AG, Groth P, Goble C, Grethe J, Heringa J, 't Hoen PC, Hooft R, Kuhn T, Kok R, Kok J, Lusher S,

Martone M, Mons A, Packer A, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S, Schultes E, Sengstag T, Slater T, Strawn G, Swertz M, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3 (1). <https://doi.org/10.1038/sdata.2016.18>