

FDO Implementation Summit 2024

Looking Back

Christine Kirkpatrick¹, Dimitris Koureas², George Strawn³, and Peter Wittenburg⁴

¹University of California San Diego San Diego Supercomputer Center: La Jolla, CA, USA

²Naturalis Biodiversity Center, Leiden, Netherlands

³National Academies of Sciences, Engineering, and Medicine, Washington, USA

⁴Gesellschaft für wissenschaftliche Datenverarbeitung (GWDG) Göttingen, Germany

This was the second conference with the focus on FAIR Digital Objects organized by the FDO Forum. The Leiden meeting in 2022 was special in so far that everything was new and many participants had not heard about FDO details beforehand. The working groups wanted to have their sessions to finalize their basic specification work which some newcomers found difficult to follow. In Leiden we heard a lot about the promises of FDOs and the thoughts about which role FDOs could play when developing the global integrated dataspace. During the Berlin meeting we could get an impression on the many ongoing implementation steps towards reality.

The meeting started with a training day where experts were asked to present basics about FDOs as well as about some advanced work. These sessions were a success for two reasons: (1) After the first 40 summit registrations all places were booked showing the great interest in getting concrete information about FDO related work. (2) The amount of interactions during the sessions created an excellent atmosphere. One of the lessons learned from the Leiden meeting was to organize a newcomer session. Sara elGebali and Anne Fouilloux managed to also create a great atmosphere where small groups were asked to interact about open questions. A few FDO experts were invited to answer questions, but were not allowed to intervene at the beginning. The top floor of the DIN building which allowed the participants to look across the roofs of Berlin completed the positive impressions.

During the summit a remarkable number of concrete contributions were presented which were selected from twice as many very good submissions. Altogether, these submissions showed that FDOs are reality and are already being implemented in a variety of labs along two dimensions: most are working on vertical scientific applications, but some are also addressing horizontal core FDO technologies to drive infrastructure building. A highlight was certainly the presentation of the lightning talks where colleagues managed to present major messages within three minutes.

The presentations showed that there are different “flavours” of FDO implementations raising a few questions: (1) Are these flavours all compliant with the FDO specifications as defined by the FDO Forum? (2) Does everyone agree on basic principles such as machine actionability, suitable level of abstraction, importance of typing, etc.? One major distinction between implementations can be described by the terms “webby” and “non-webby”, i.e. some fully rely on the URL/HTTP/HTML technology stack, while others rely on Handles/DOIs and structured information. Most differences can be found in packaging the semantics used in the

structure resulting from the resolutions of either the URLs (landing page), the Handles (attribute-value structures) or even in complex metadata schemas as developed in some scientific or industrial domains. These differences confirm the FDOF view that only minimal mandatory specifications should be made mandatory by the FDOF. It is obvious that the further development needs to be based on clear specifications, but on the other hand shows enough flexibility to allow for different implementations. There was evidence that some colleagues were even not aware that there are FDO specifications and that many implementations are being developed in separation. But there is also an indication of coalescence driven by the insights due to all these concrete implementations. While until now most FDO work is being done embedded in projects, it is welcomed that in the FDO One project the FDO is the core focus, developing a basic infrastructure, initiating an international FDO testbed, working on bridge-building with other dataspace that rely on industrial type of technologies. There is some optimism that FDO One may be kicking-off a standardization process.

Of great value were also the three panels with distinguished experts. From the “Essential Factors for enabling FDOs” panel FDOF can draw the following conclusions: (1) need to increase the measures for international capacity building and education, (2) strengthen the community engagement, (3) enhance interoperability and use open standards where possible, (4) promote policy advocacy, and (5) ensure equity and inclusion, i.e. addressing in particular the international dimension. For an organization which is based on voluntary work this is a tall order. FDOF needs to work out how efforts can be intensified.

From the panel on “FDO Forum Future” one can also draw a few essentials: (1) FDOF efforts should be embedded in current trans-regional developments and interdisciplinary efforts; (2) FDOF needs to understand that the expectations towards the FDO and FDOF are growing due to new initiatives such as “dataspaces” and already ongoing international interactions; (3) There is an urgent need for transcontinental solutions as a result of fair interactions and by including a broad number of stakeholders; (4) The work on demonstrable use cases across borders which act as catalysts should be increased; (5) There is a clear need to work on the clarification of terminology to increase the common understanding.

From the panel on “FDO Flavours” we can also draw some essentials: (1) The discussion on flavours is necessary, but we need to accept that there are multiple ways to implement FDOs and wishes to accommodate specialty needs. Most of these needs will become apparent through ongoing use cases meeting specific purposes and demonstrating a few typical patterns; (2) A major difference can be found in “webby” and “non-webby” approaches and the ways metadata is packaged into attributes of different structures (Handle Records, Landing Pages, Metadata frameworks). Hybrid solutions should be possible; (3) A major question then is how to achieve interoperability. Looking at layers could be a promising approach and the challenge of a suitable FDO typing framework needs to be addressed. Interoperability requires a proper design; (4) Whatever FDOF will be doing, it needs to follow a pragmatic approach that is made up of connectors, protocols, or gateways

A few important actions can be drawn from these discussions:

- FDOF needs to take actions to disseminate the specifications, to check them against practices and to bring developers together.
- More analysis work needs to be carried out by the TSIG working group to analyze and compare the different flavors and tackle the typing and interoperability issue.
- Erik Schultes presented the idea of FDO Compliance Forms, which could be used for the self-assessment of candidate FDO technology, was very well received and will be worked out.
- Finally, FDOF needs to address the international dimension.

The well-known acronym FAIR can also be interpreted as data that are Fully AI Ready. During the walking dinner and social event at Berlin's famous Museum für Naturkunde, a dialogue was hosted under the Brachiosaurus on the topic "FDOs and AI". Aarne Talman, Data & AI Senior Manager at Accenture and Large Language Model expert fielded questions and elaborated on numerous cross-cutting issues: The role of FDOs in ensuring equitable data access for training LLMs; The role of FDOs and rich provenance to give credit, to assess bias or errors in AI output; Are ML applications FDOs? Should they be? How will the FDO spec impact the future developments of AI? The dialogue was informal but well attended by Summit participants.

Summarizing we can state that the Berlin summit

- indeed focussed on implementations and showed that the concept of FDOs is taking off,
- was very productive in so far that from the beginning lots of discussions and interactions took place which stimulated dialogues and work beyond the conference
- clearly identified the next steps for the FDO Forum.

Therefore it was a logical follow up that the focus of the 3rd FDO conference, which will probably happen at the AGU facilities in Washington DC in 2025, will be put on the interoperability issues. There is an interest from our African colleagues to organize the FDO conference 2026 in an African country. We as co-chairs support this idea and suggest taking measures to prepare for such a step.

As the FDO Forum we need to find ways to intensify the engagement to meet the expectations. We already took steps to implement a lightweight governance structure that will help to make our decision processes more transparent and open. This implies that we will also increase some light formalisms to understand who is a member of the FDO Forum and thus part of the decision-making processes. Currently, the FDO Forum only manages a list of registered individuals which are part of the "FDO community". FDOF also needs to review its working group activities and make suggestions to tackle the challenges.

We welcome the idea to ask all presenters and all who submitted abstracts for extended paper publications using the TIB Open publishing possibility.

As co-chairs of the FDOF we would like to thank all those colleagues who contributed to the success of this FDO Implementation Summit: the summit steering and program committees, the local organizing committee, the presenters, the panelists, the session and panel chairs. We also would like to thank the FDO One project (the FDO One project is funded within the Mission AI program of the German federal government and administered by Acatech) and COS for its funding support and DIN and MfN/Leibniz for offering their facilities.