

Starting FDO in the Cradle -- ROcrating Live Data

Guido Aben[‡], Juri Hößelbarth[§]

[‡] CS3mesh4EOSC project (<https://cs3mesh4eosc.eu/>), Utrecht, Netherlands

[§] Westfälische Wilhelms-Universität, Münster, Germany

Corresponding author: Guido Aben (guido.aben@aarnet.edu.au), Juri Hößelbarth (hoesselbarth@uni-muenster.de)

Abstract

This talk discusses the use of Fair Digital Objects (FDOs for short) for a democratised approach to FAIRness, that is, adherence to the Findable/Accessible/Interoperable/Reusable set of requirements, collectively called [FAIR](#). This capability is being built for the [CS3MESH4EOSC](#) project.

CS3MESH4EOSC is a 3-year EU-funded project in the EOSC context (we started Feb 2020) that addresses the challenges of the fragmentation of file and application services, digital sovereignty and the application of FAIR principles in the everyday practice of researchers. Initially, 7 major data services will be combined into [ScienceMesh](#) - a federated service mesh providing a frictionless collaboration platform for hundreds of thousands of users (researchers, engineers, students and staff). The service will offer easy access to data across institutional and geographical boundaries. The infrastructure will be gradually expanded and offered to the entire education and research community in Europe and beyond. The initial service will connect services in NL (SURFdrive), PL (PSNCBox), AU (CloudStor), DE (Sciebo), [CZ \(owncloud@CESNET\)](#), CH (SWITCHdrive) and DK (ScienceData), as well as domain data stores at CERN (CERNBox) and the EU's own Joint Research Centre's Copernicus (earth observation) datastore.

The CS3MESH4EOSC project is busy designing, building and deploying the necessary technology to achieve this. CS3MESH4EOSC grew out of the grassroots "CS3" community, which started out as a self-help forum of infrastructure builders and providers from the academic sector who look after rapidly growing datastores of the "synch-'n'-share" paradigm (dropbox is a commercial equivalent); this type of store is growing rapidly more popular as a basic building block for live data storage and collaboration in research.

The mission for the CS3MESH4EOSC project is to improve scientific collaboration across the entire mesh (essentially an interoperating federation of data stores), and to ensure that data sharing across this resulting mesh is done according to FAIR principles. This puts the CS3MESH4EOSC in a unique position: we need to bring FAIR tooling in front of a broad audience of research users (not just "FAIR literate" ones), and convince them that FAIRness

is relevant at the point where [live data is being collected](#), not just when data has congealed to collections. We recognise two main obstacles:

1. FAIR-aware infrastructure needs to be simply available, right in front of every user's face, and be so usable that it gets broad uptake. By rule of thumb every additional step required sheds half of the userbase you start out with.
2. Research communities need to be motivated, trained and assisted to use the FAIR infrastructure. It needs to make their lives easier. Without relevant infrastructure in place, there is no point in mounting FAIRness awareness campaigns.

Therefore, CS3MESH4EOSC's approach to FAIR uptake is to start from the Science Mesh of datastores as described in the first paragraph, already in widespread use by researchers. We add a FAIR Description Service to these stores, available for any researcher of the system to use (an instance of the "Describo" tool, <https://arkisto-platform.github.io/tools/description/describo-online/>). Thus they can create [FAIR Digital Object](#) packages from their own data (using the RO-Crate standard) and also manage the deposition process, initially targeting the open access [Zenodo](#) and [Dataverse](#) repository services and the [Open Science Framework \(OSF\)](#) science workflow portal. The resulting system of metadata annotation and user guidance wizards that facilitate the process is called "ScieboRDS" (<https://www.research-data-services.org/page/about/>).

By thus adding metadata awareness and annotation capabilities to this mesh that already has several hundreds of thousands of users and tens of Petabytes of live data on it, we end up with a democratised, low-barrier-of-entry approach to FAIR. Allowing researchers to generate FDOs from their live data (no onerous upload / collections steps) will help create more FAIR data supply. The capabilities thus far described are already available and are being deployed to users starting Q3 2022. Further development is underway that allows better capability negotiation between the live data store ("the Science Mesh") and the backend repository, such that users can rely on the relevant schema being autoprovisioned and ontologies being agreed upon before the FDO is packaged, thus improving metadata quality.

Keywords

data management, data lifecycle, born digital, FAIR Data, FDO, digital objects

Presenting author

Juri Hößelbarth

Presented at

First International Conference on FAIR Digital Objects, poster

Conflicts of interest