3rd FAIR Digital Objects Conference

Placeholder Session title

https://doi.org/10.52825/xxxx..... DOI placeholder (WILL BE FILLED IN BY TIB Open Publishing)

Published: (WILL BE FILLED IN BY TIB Open Publishing)

FDO and Dataspaces for the Distributed Digital Era

Sven Bingert[1,*] https://orcid.org/0000-0003-4905-1936, Jana Böhm[1] https://orcid.org/0000-0001-9547-1582 and Timm Fitschen[2] https://orcid.org/0000-0002-4022-432X

[1]GWDG, Germany

[2]Indiscale, Germany

*Correspondence: Jana Böhm@gwdg.de

Abstract. This paper describes the results of the FDO One Project which had as tasks to demonstrate that FDOs are excellent basic building blocks to bridge between highly different dataspaces such as the Mobility Dataspace based on early IDSA architecture suggestions, a dataspace defined by the use of the I4.0 Asset Administration Shell and 4 dataspaces defined by repositories of different nature including an S3 object store.

Keywords: Fair Digital Objects, Dataspaces, Global Data Domain

1. Introduction

In 2010 the EC published the report “Riding the Wave on Scientific Data” [1] where it was stated that establishing trust will be the most crucial dimension for improving data sharing and reuse. Of course, trust has many different dimensions but in this document we will limit ourselves on dimensions that can be described with the terms “data sovereignty, transparency, traceability, accountability and persistence”, i.e. other important dimension such as data quality will not be touched in this paper. The question then is whether we made real steps in having implemented a global trust framework, since data exchange happens at global level. The answer is no, although much work is going on in testing technologies that include trust building with the hope to improve the situation. The pressure for achieving real steps is increasing due to the emerging broad AI applications and the building of small and large AI models based on data from various partners. It also becomes much clearer that even for “open data” it will be important to formally state reuse conditions. AI models will be updated regularly with new data and new parameter sets. Since these AI models increasingly often include information relevant for societies, steps to implement the above-mentioned five dimensions will become crucial.

2. Relevant Technologies

We see two technology strands that in our view will become game changers: FAIR Digital Objects (FDO) [2] and Dataspace Technologies (DST). **FDOs** will overcome the current inefficiencies related with the fact that we are still living in a world where we exchange individual files, i.e. data, metadata and rights metadata. These are still treated separately although they are closely related and needed when data reuse is the goal. In our distributed data domain where we do copying, creating derivations, etc. this scattered management results in many of the great problems and inefficiencies. It is the great achievement of FDOs to bundle all information in a secure and persistent way which is necessary to process a given bit-sequence and to associate the bundle with one single PID. This bundle can even be encapsulated similarly

as it is known from object-oriented programming, i.e. the data producer can associate operations with the bundle to ensure that access is only possible in specific predefined ways [3]. Of course, this bundle needs to be protected against manipulations by unauthorised persons which can be done in various ways using PKI and signed certificates where necessary. We should add here that the concept of "digital objects" is not new. Publishing and film industry are using this concept by having introduced DOIs [4] and EIDRs [5] decades ago, and basically object-stores function internally according to similar mechanisms [6]. The international FDO Forum specified the FDOs and since they are not interpreting rights specifications and not putting them into action, the FDO specifications can easily be accepted globally.

DST is a comparatively new term and describes software architectures that allow the control of data reuse. We see a couple of approaches that have been worked out recently such as by the Mobility Dataspace [7], IDSA Architecture WG [8], the Factory X project [9], and the Ouranos project [10], for example, of which some are being implemented with huge effort. All these approaches have three components: (1) The trust building layer in which formal contracts between partners are being signed to agree on rules, technologies and semantics. (2) The control plane layer where provider rights specifications are compared with the consumer credentials and wishes. (3) The data plane where the data will actually be exchanged if the usage negotiations had a positive result. Dataspace technology approaches are therefore comprehensive.

But we may not forget the many data infrastructures that have been realised in the last decades for example within the ESFRI program [11], that implemented dataspaces characterized by agreements on community specific rules and technologies. As can be expected, there are differences between all these implementations which will also be influenced by the law systems in the various regions and countries. Standardisation processes are going on within ISO for example, but it will take much time to develop worldwide standards for some aspects.

3. FDO and DST

As indicated, FDOs bundle all relevant information about the included bit-sequences together in a secure and persistent way including the rights. Non-FDO Repositories, for example, only need to provide an FDO adapter to offer FDOs to the outside world which for well-structured repositories is easy to do. FDOs can now provide the rights information specified for example in ODRL to the Control Plane of dataspace technologies and it can offer data and metadata of all kinds to the Data Plane, all by using one single protocol interface worldwide. This bridge function is independent of the specific implementation of DST making FDOs an excellent choice for bridge building.

In the FDO One project [12] we have implemented this bridge building role of FDOs between dataspaces of different types: the Mobility Dataspace being built on early IDSA based suggestions, a dataspace defined by I4.0 Asset Administration Shell specifications [13], a native DO-supporting repository (Cordra) [14], a S3 object store and three non-FDO repositories (LinkAhead [15], Dataserve [16], B2Share [17]). All software including the adapters and a first FDO Manager were developed within 8 months.

4. Conclusions

To enhance data sharing and reuse, we will need to implement trust dimensions such as data sovereignty, transparency/traceability, accountability and persistence it will be necessary to broadly implement new ways of dealing with data. At the basis we will need to turn to (FAIR) Digital Objects that are bundling the bit-sequences and all metadata information associated with them, in particular statements about reuse conditions. This bundling will take care that all information will be kept together independent of the circumstances and the activities being undertaken, especially when encapsulation techniques are being used.

In a distributed and heterogeneous domain of data providers and data consumers which are using different kinds of dataspace technologies interpreting the rights specifications and putting

them to action, FDOs will serve as a basic building block and thus have a similar function as TCP/IP in the Internet. In both cases bridges between existing networks or dataspaces based on globally accepted standards are crucial. FDO is based on a comparatively simple specification and can easy be implemented to act as a general layer. The FDO One project has shown that this bridge building concept is feasible to be implemented in shortest times.

We will show how the combination of FDOs and DST will support the 5 trust dimensions mentioned above.

5. Data availability statement

There is no data involved.

6. Underlying and related material

All related material is openly accessible.

7. Author contributions

All authors contributes to all sections.

8. Competing interests

There are no competing interests.

9. References

[1] https://www.dariah.eu/wp-content/uploads/2017/02/hlg-sdi-report.pdf
[2] https://fairdo.org/
[3] https://arxiv.org/pdf/2504.05361
[4] https://www.doi.org/
[5] https://www.eidr.org/
[6] https://aws.amazon.com/de/s3/
[7] https://mobility-dataspace.eu/
[8] https://internationaldataspaces.org/we/working-groups/
[9] https://factory-x.org/
[10] https://www.meti.go.jp/english/policy/mono_info_service/connected_industries/ouranos.html
[11] https://www.esfri.eu/
[12] https://www.tib-op.org/ojs/index.php/ocp/article/view/1413
[13] https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/AAS-ReadingGuide_202201.pdf?__blob=publicationFile&v=1
[14] https://cordra.org/
[15] https://getlinkahead.com/
[16] https://dataserve.eu/en/
[17] https://b2share.eudat.eu/