Emerging FAIR Ecosystem(s): A Practical Perspective

Erik Anthony Schultes[‡,§], Arofan Gregory[|], Barbara Magagna[‡]

‡ GO FAIR Foundation, Leiden, Netherlands
§ Leiden Center for Data Science, Leiden, Netherlands
| Standards Expert, CODATA, Paris, France

Corresponding author: Barbara Magagna (barbara@gofair.foundation)

Abstract

There is broad acceptance that FAIR data (Wilkinson et al. 2016) reuse is desirable, with considerable interest and energy being devoted to its realization, but many questions remain on the part of prospective implementers. Fundamental to explaining how best to implement FAIR is an overview of how the many FAIR models (and the technologies that support them) fit together into a coherent "FAIR ecosystem". How do FAIR Implementation Profiles (FIPs) (Schultes et al. 2020) relate to FAIR Data Points (FDPs) (Bonino da Silva Santo et al. 2022), and how do these relate to the concept of FAIR Digital Objects (FDOs) (Hudson et al. 2018)? What is their relationship to other diverse FAIR resources and digital assets (metadata, datasets, repositories, and the complex web of services that run them)? How are these novel and legacy systems intended to interoperate? These questions are often encountered by those involved in the growing number of projects looking at FAIR implementation (ENVRI-FAIR 2019, CODATA 2022, SciDataCon 2022, Seoul Korea 2022b , Health-Holland Project 2021, Health-Holland Project 2022, Swiss Personalized Health Network 2022, Maxwell 2021, VODAN 2020).

Such an overview could also inform the development of specifications for the different models involved in a FAIR ecosystem, such as FIPs, FDPs, and the description of digital resources (data and services) at various levels. With an agreed picture of the FAIR-reuse ecosystem, the points of contact and "hand-off" would be easier to describe and coordinate.

This presentation looks at questions from FAIR implementation across various settings, and proposes a view of the overall ecosystem which could be agreed and communicated to prospective implementers. It suggests the relationship between various artefacts being discussed in the FAIR community today (FIPs, FDPs, FDOs, and other digital assets) and looks at how these can be connected to the business layer to support the development of services and applications within the envisioned FAIR ecosystem. Notably, this includes how the Cross-Domain Interoperability Framework (CDIF) being developed through the

WorldFAIR project can connect to the underlying FAIR ecosystem in practical terms (Weise et al. 2022).

The presentation will address high-level considerations around the major technology components of a FAIR ecosystem, their roles within a range of common user scenarios (often having unavoidable legacy technology), and their relationship to each other and to the set of models needed to provide practical services for FAIR interoperation at the business level.

Three basic scenarios are examined, in order to understand the practical requirements of different communities. The first scenario is one which has received a good deal of attention during initial efforts to implement the FAIR principles, a domain or user community without a strong pre-existing culture of data sharing and resume, wishes to become FAIR. The second scenario is one where a community with a strong existing culture of data sharing and reuse is looking to integrate its current approaches with those advocated by the FAIR community. The third scenario looks at FAIR from the perspective of the implementer of FAIR services from an “industrial” perspective: how does FAIR provide the kind of market which is needed to support full-scale services and application development?

Each of these scenarios provide valid, but different views of what it will take to implement the FAIR in practical terms. In order to understand them, we can draw parallels with other large-scale data-sharing efforts in other communities - the Internet and the Web itself can be understood as a useful example of how vision, standards, and implementation combine to provide successful infrastructure at this scale. Indeed, the concept of Digital Objects (and now FAIR Digital Objects) has its roots in this analogy (Kahn and Wilensky 2006). Other, smaller examples also exist, which focus more specifically on the exchange of data and metadata: for example the Statistical Data and Metadata Exchange (SDMX) Initiative (SDMX community 2022) or the emerging Cross-Domain Interoperability Framework (SciDataCon 2022, Seoul Korea 2022a). Although implemented within targeted communities, these efforts exchange a wide range of data and metadata not entirely dissimilar to what is envisioned in FAIR. There is currently no single exact parallel for FAIR ecosystems, but there are examples from which we can learn in terms of making large-scale data reuse a practical reality.

Core to these is a vision of all of the component pieces, and how they can act in concert to provide a scalable infrastructure which will address the needs of the many different communities of users. Such a common vision may be implicitly agreed among those working on FAIR implementation today, but in the interests of clear communication, it is time to document it - and in keeping with FAIR, this documentation should be itself machine-actionable. As we move toward the specification of the many components of the FAIR ecosystem, it seems only common sense to have an agreed roadmap.

Keywords

FAIR Digital Objects (FDOs), FAIR Implementation Profiles (FIPs), FAIR Data Points (FDPs)

Presenting author

Erik Anthony Schultes

Presented at

First International Conference on FAIR Digital Objects, presentation

Conflicts of interest

References

- Bonino da Silva Santo LO, Burger K, Kaliyaperumal R (2022) FAIR Data Point Specification . https://specs.fairdatapoint.org. Accessed on: 2022-7-09.
- CODATA (2022) WorldFAIR: Global cooperation on FAIR data policy and practice. https://codata.org/initiatives/decadal-programme2/worldfair/. Accessed on: 2022-7-09.
- ENVRI-FAIR (2019) ENVRI-FAIR: Advancing the findability, accessibility, interoperability, and reusability (FAIRness) of the data and services offered by the ENVRI Cluster research infrastructures and to connect them to the emerging European Open Science Cloud. https://envri.eu/home-envri-fair/. Accessed on: 2022-7-09.
- Health-Holland Project (2021) Trusted World of Cornona. http://www.twoc.eu. Accessed on: 2022-7-09.
- Health-Holland Project (2022) Citizen Centred and Controlled COVID-19 data for reuse. https://www.health-holland.com/project/2021/2021/citizen-centred-and-controlled-covid-19-data-reuse. Accessed on: 2022-7-09.
- Hudson S, Jones S, Collins S, Genova F, Harrower N, Laaksonen L, Mietchen D, Petrauskaité R, Wittenburg P (2018) Turning FAIR data into reality: interim report from the European Commission Expert Group on FAIR data. Zenodo. https://doi.org/10.5281/zenodo.1285272
- Kahn R, Wilensky R (2006) A framework for distributed digital object services. International Journal on Digital Libraries 6 (2): 115-123. https://doi.org/10.1007/s00799-005-0128-x
- Maxwell L (2021) FAIR Data for European COVID-19 Response Fall 2021 Workshop Series. https://figshare.com/collections/FAIR_Data_for_European_COVID-19_Response_Fall_2021_Workshop_Series/5740430/1. Accessed on: 2022-7-09.

- Schultes E, Magagna B, Hettne KM, Pergl R, Suchánek M, Kuhn T (2020) Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence. Lecture Notes in Computer Science138-147. https://doi.org/10.1007/978-3-030-65847-2_13
- SciDataCon 2022, Seoul Korea (2022a) Session 438: Practical Cross-Domain Data Sharing: Envisioning a Cross-Domain Interoperability Framework. https://www.scidatacon.org/IDW-2022/sessions/438/. Accessed on: 2022-7-09.
- SciDataCon 2022, Seoul Korea (2022b) Session 439: WorldFAIR - Global cooperation on FAIR data policy and practice. https://www.scidatacon.org/IDW-2022/sessions/439/. Accessed on: 2022-7-09.
- SDMX community (2022) Statistical Data and Metadata eXchange. https://sdmx.org. Accessed on: 2022-7-09.
- Swiss Personalized Health Network (2022) SPHN Data Ecosystem for FAIR Data. https://www.youtube.com/watch?v=pqV0qp4oisM. Accessed on: 2022-7-09.
- VODAN (2020) Virus Outbreak Data Network - Africa and Asia . https://www.vodan-totafrica.info. Accessed on: 2022-7-09.
- Weise M, Kovacevic F, Popper N, Rauber A (2022) OSSDIP: Open Source Secure Data Infrastructure and Processes Supporting Data Visiting. Data Science Journal 21 https://doi.org/10.5334/dsj-2022-004
- 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