

The FAIR hourglass: A framework for FAIR implementation

Erik Schultes

FAIR Implementation Lead, GO FAIR Foundation, Poortgebouw Noord, Rijnsburgerweg 10, 2333 AA Leiden, The Netherlands

E-mail: eriks@gofair.foundation; Tel.: +310642448027; ORCID: <https://orcid.org/0000-0001-8888-635X>

Editor: Barbara Magagna (<https://orcid.org/0000-0003-2195-3997>)

Solicited reviews: Louis Ter Meer (<https://orcid.org/0000-0003-2427-0614>); Mirjam van Reisen (<https://orcid.org/0000-0003-0627-8014>)

Received 30 November 2022

Accepted 16 December 2022

Abstract. The FAIR Hourglass provides a framework to organize two general phases of FAIR implementation: FAIRification (top) and FAIR Orchestration (bottom). The center of the hourglass represents the use of widely agreed-upon open, minimal standards ensuring machine-actionability.

Keywords: Data management, data stewardship, FAIR principle, knowledge of theories underlying fair implementation, fair evaluation of repositories for data deposition

Since the FAIR Guiding Principles were published in 2016 [17], there has been a diverse deployment of methods, workflows and technologies towards their instantiation: Bring Your Own Data events, the FAIR Cookbook, the RDM Kit, RO-Crates, workflows based on Jupyter Notebooks, FAIR Digital Objects of various sorts and incremental upgrades to existing repositories. Regardless of method and technology, FAIR implementation has a number of universal features. Taking inspiration from the “hourglass” architecture of the Internet [18] these common implementation activities can be organized into a “FAIR hourglass”, providing a framework making approaches to FAIR implementation more systematic, reproducible, and scalable.

Before the FAIR Hourglass can be described however, it is necessary to understand better the FAIR Principles. Broadly speaking, the FAIR Principles can be partitioned into two classes: those that address the Information Technology (IT), and those that address the data content. For simplicity we refer to the FAIR Principles focusing on IT and infrastructure as the Red principles, and those on the domain-relevant content as the Blue principles (Fig. 1) [4,8]. The Red principles tend to focus on the capabilities of the data infrastructures and are mostly agnostic to the actual content and specialized domain requirements. Conversely, the Blue principles represent implementation choices that must be made by practicing domain experts (e.g., the scholar, the research scientist, industry associations), and include standards and technologies around data formats.

Furthermore, FAIR implementation involves a number of distinct, yet interrelated activities on the part of the researcher, data steward, and infrastructure providers. These activities belong to two distinct phases of FAIR implementation that are often conflated, leading to confused expectations around FAIR implementation. Separating these two phases can help to clarify priorities and provide a more productive division of labor among stakeholders. The first phase addresses the “FAIRification” of research data and other content. By FAIRification, we mean a process by which data captured using localized or domain-specific practices are transformed into formats that follow open

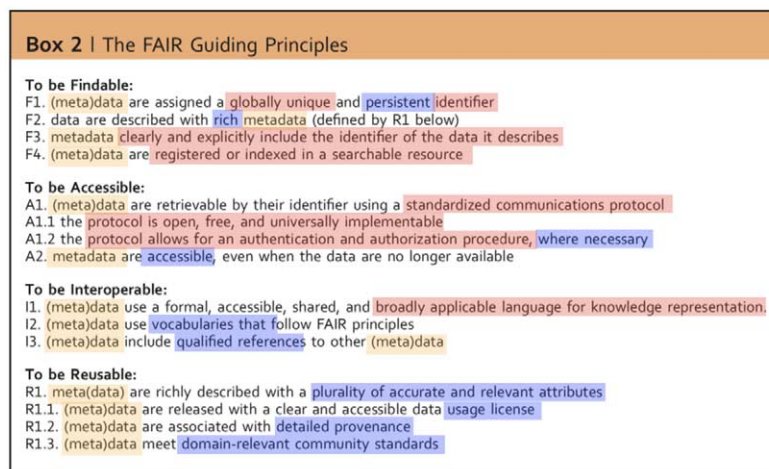


Fig. 1. The FAIR Guiding Principles as they were originally published in 2016, partitioned here into those associated primarily with technical implementation (highlighted in Red) and those associated with content-related, domain-relevant standards and practices (highlighted in Blue). Yellow highlights indicate the pervasive role of metadata throughout the FAIR Principles.

standards for interoperability. For example, blood pressure measurements captured as “free-text” in an Excel spreadsheet can be FAIRified by reformatting them into RDF triples, following a community endorsed schema and using controlled vocabularies. This transformation makes the data FAIR-ready and opens the door to machine-processing.

The second phase of FAIR implementation is to put the FAIR-ready data into action by exposing them to software applications and services that can perform operations on them. Although specialized technologies exist that can demonstrate automatic F, A, I and R under limited conditions (for example SPARQL queries on triples stores hosting suitable endpoints) general purpose solutions and agreements on standards supporting the operations themselves are only now beginning to emerge. By analogy to the more general concepts of “system orchestration” [11] and “data orchestration” [16], the automatic processing of FAIRified data can be referred to as FAIR Orchestration and includes a diverse range of operations associated with search and indexing, ontology-based access control of restricted data [2], disambiguation of semantic content, merging appropriate datasets and eventuality running analyses leading to new insights.

The Red/Blue distinction between the FAIR Principles, and the FAIRification/FAIR Orchestration phases of FAIR implementation cycles are themselves related. The activities of FAIRification are dominated by the Blue principles, heavily involving the domain experts and their established practices, while the processes of FAIR Orchestration involve largely the Red principles and the technical implementations supporting automated F, A, I and R services.

In general, FAIR implementation is a linear process beginning with data capture following the established practices of a given research domain, which in general must then be FAIRified (Blue) according to domain-relevant standards. FAIRification makes data machine-actionable, and when exposed to the Internet, susceptible to generic services leading to their FAIR Orchestration (Red). The FAIR Hourglass (Fig. 2) represents this linear process vertically, with FAIRification at the top of the hourglass, and FAIR Orchestration as the bottom of the hourglass. The waist, or center, of the hourglass is represented by the FAIR-ready data which is both the result of FAIRification and the input for FAIR Orchestration. As such the data models at the center of the hourglass act as bridges linking the two phases of FAIR Implementation.

The overall hourglass shape indicates the ‘freedom to operate’, meaning the discretion one has when choosing implementation standards and technologies. Hence, the freedom to operate is maximized at the top and bottom where end-users are either creating data (top) or analyzing data (bottom). At the top and bottom, the users (and the applications running there) are protected from the details of the FAIR infrastructure (both FAIRification or FAIR Orchestration layers), leaving them free to perform their work following practices that are best fit to, and serving of that domain community.

The freedom to operate is most constrained at the center of the hourglass by requirements for machine-readability (i.e., the subsequent interoperability) of the data/metadata. As the overall goal of the hourglass model is to facilitate

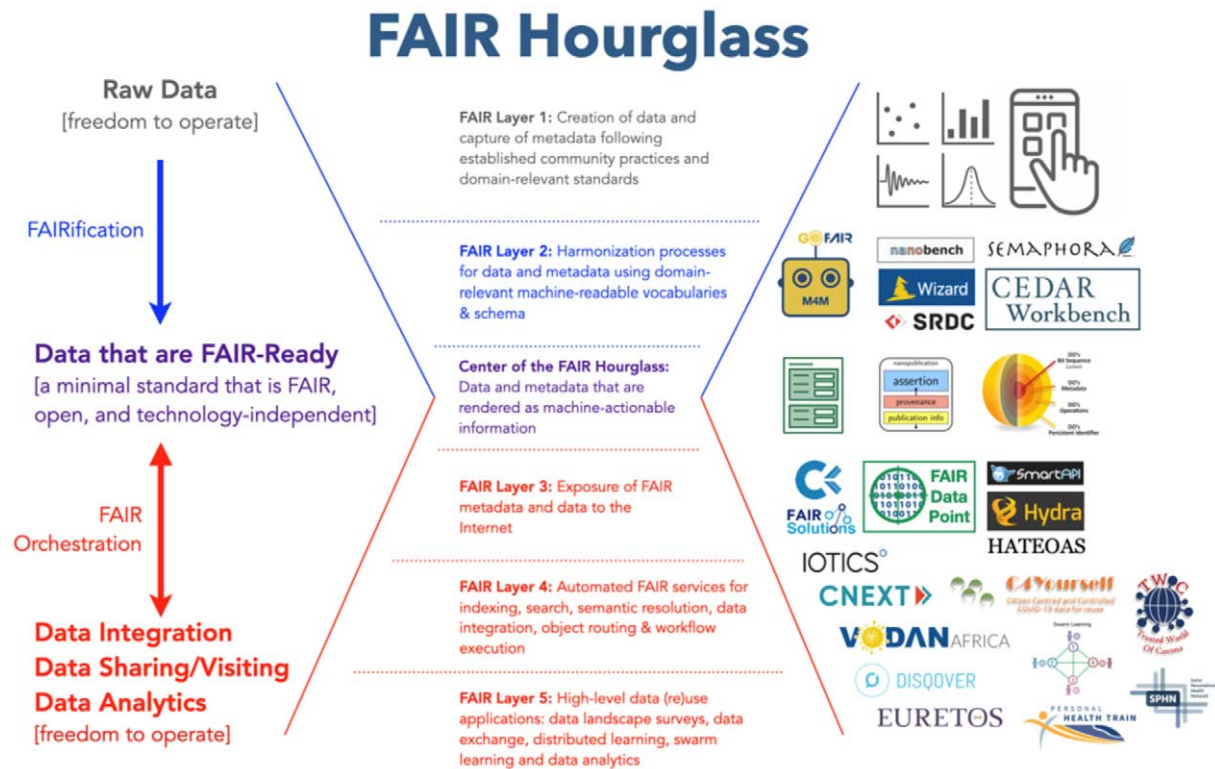


Fig. 2. The FAIR Hourglass as a general framework for FAIR implementation. Methodologies and technologies (logos, on right side of the figure) supporting specialized FAIR implementation functions at the different layers are provided as examples but do not constitute an exhaustive list of emerging solutions.

FAIR while maximizing the freedom to operate in any particular use case, the specification at the center of the hourglass aims to be as open, minimal, and least burdensome as possible to implement. Although any consistently structured data model would in principle suffice at the center, typical examples of FAIR-ready representations include RDF models making use of controlled vocabularies or high-order ontologies. However, there are candidates for more formalized specifications at the center of the FAIR hourglass, including frameworks around nanopublications [9], Signposting [13], and FAIR Digital Objects [5]. These candidate specifications reflect a in every case minimal, open, FAIR-ready standard that behaves as machine-actionable containers (MACs) of the objects they represent. Following the First International Conference on FAIR Digital Objects (October 2022) [6] there will be a concerted effort among a critical mass of implementers to road-test and refine a single, minimal specification for MACs.

Between the top and bottom layers of the hourglass, are intermediate layers of methods, tooling, services and protocols that facilitate the actual FAIRification and the FAIR Orchestration functions. These intermediate layers impose restrictions on data representations and the permitted operations that ensure adherence to the FAIR principles such that, the closer to the center of the hourglass, the more restrictions apply.

Beginning from the top, where data are generated by essentially any means (in the worst case, as hand-written scribbles on sticky notes), researchers and data stewards can use specialized tools and methods to create the necessary schema and vocabularies engendering FAIR attributes. For example, in the Metadata for Machines (M4M) workshops [10], domain experts craft metadata templates using tools like the Sheet2RDF pipeline, CEDAR Workbench and BioPortal registry. The resulting metadata templates become FAIR Supporting Resources that can be subsequently reused by any other member of that community. The M4M workshop and its associated tooling apply increasing restrictions on the otherwise unconstrained content of the community, acting like a funnel, progressively guiding metadata descriptions into more machine actionable representations at the center.

The FAIR-ready data/metadata at the center of the hourglass then opens the door to their FAIR Orchestration, that is, to the services rendering FAIR operations. The first step is to expose (publish) the FAIR-ready data/metadata where they can be made available to machine agents. Just below the center of the hourglass are services such as FAIR Data Points [1,3,7], Smart APIs [14] and other applications that provide to the data and metadata, gateways to applications that can process them. These gateway applications in turn, along with the machine-readable metadata, opens the door to a broader range of FAIR services that automatically perform operations such as indexing, search, accessibility (via ontology-based controls that, for example, implement the regulations set by the data creators in adherence to applicable jurisdictions), semantic resolution, data integration, object routing and workflow execution. With these essential FAIR functions in place, domain users can enter processes launching queries, dispatching data trains (data visiting scenarios) [12,15] and performing data analyses as they see fit. The results of these data analyses will be the generation of new data, which, if not born FAIR, enter the FAIRification phase, initiating a new cycle of the two-phase FAIR implementation flow.

In Summary, the FAIR Hourglass is not a particular method of FAIR implementation, but a framework that structures the decisions and activities common to any FAIRification or FAIR Orchestration effort. The overall hourglass shape, and the restricted waist at the center, alludes to a strategy towards FAIR data reuse inspired by the distributed, technical infrastructure composing the Internet. In FAIR data networks however, this infrastructure must be augmented with domain-relevant community standards declared by communities of practice. The FAIR Hourglass is therefore not a purely technical architecture, but a socio-technical framework.

Funding

The authors report no funding.

Disclosures and conflict of interest

The author reports no conflict of interest. The author was not involved in the peer-review process.

References

- [1] O.M. Benhamed, K. Burger, R. Kaliyaperumal, L.O.B. da Silva Santos, M. Suchánek, J. Slifka et al., The FAIR data point: Interfaces and tooling, *Data Intell.* **28** (2022), 1–18. doi:10.1162/dint_a_00161.
- [2] C. Brewster, B. Nouwt, S. Raaijmakers and J. Verhoosel, Ontology-based access control for FAIR data, *Data Intell.* **2**(1–2) (2020), 66–77. doi:10.1162/dint_a_00029.
- [3] L.O.B. da Silva Santos, K. Burger, R. Kaliyaperumal and M.D. Wilkinson, FAIR data point: A FAIR-oriented approach for metadata publication, *Data Intell.* **28** (2022), 1–21.
- [4] Detailed interpretation by the GO FAIR Foundation of the FAIR Principles [Internet]. Available from: <https://www.gofair.foundation/interpretation> [cited 2022 Nov 15].
- [5] FAIR Digital Object Forum [Internet], available from: <https://fairdo.org> [cited 2022 Nov 15].
- [6] First international conference on FAIR digital objects, FDO Forum [Internet], Available from: <https://www.fdo2022.org> [cited 2022 Nov 15].
- [7] R. Hooft, FAIR Data Point [Internet], available from: <https://www.fairdatapoint.org> [cited 2022 Nov 15].
- [8] A. Jacobsen, R.M. Azevedo, N. Juty, D. Batista, S. Coles, R. Cornet et al., Fair principles: Interpretations and implementation considerations, *Data Intell.* **2**(1–2) (2020).
- [9] T. Kuhn, Tutorial videos for Nanobench [Internet], available from: <https://github.com/peta-pico/nanobench#tutorials> [cited 2022 Nov 15].
- [10] Metadata for Machines resource page [Internet], available from: <https://www.gofairfoundation.org/m4m/> [cited 2022 Nov 15].
- [11] Orchestration (computing) Wikipedia [Internet]. Available from: [https://en.wikipedia.org/wiki/Orchestration_\(computing\)](https://en.wikipedia.org/wiki/Orchestration_(computing)) [cited 2022 Nov 15].
- [12] S. Österle, See the Swiss Personal Health Network (SPHN) Data Ecosystem for FAIR Data. Available from <https://www.youtube.com/watch?v=pqV0qp4oisM> [cited 2022 Nov 15].
- [13] Signposting the Scholarly Web [Internet]. Available from: <https://signposting.org> [cited 2022 Nov 15].
- [14] SmartAPI, Building a connected network of fair APIS [Internet], available from: <https://smart-api.info> [cited 2022 Nov 15].

- [15] S. Wamat-Herresthal, H. Schultze, K.L. Shastry, S. Manamohan, S. Mukherjee, V. Garg et al., Swarm learning for decentralized and confidential clinical machine learning, *Nature*. **594**(7862) (2021), 265–270. doi:10.1038/s41586-021-03583-3.
- [16] What is Data Orchestration & Why It's Essential for Analysis [Internet], available from: <https://segment.com/resources/data-strategy/what-is-data-orchestration/> [cited 2022 Nov 15].
- [17] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak et al., Comment: The FAIR guiding principles for scientific data management and stewardship, *Sci Data*. **3** (2016).
- [18] P. Wittenburg and G. Strawn, Common Patterns in Revolutionary Infrastructures and Data [Internet], 2018 Mar. Available from: <https://b2share.eudat.eu/records/4e8ac36c0dd343da81fd9e83e72805a0> [cited 2022 Nov 15].