

How does software fit into the FDO landscape?

Carlos Martinez-Ortiz[‡], Carole Goble[§], Daniel S Katz[‡], Tom Honeyman[¶], Paula Andrea Martinez[#], Michelle Barker[»], Leyla Jael Castro[«], Neil Chue Hong[»], Morane Gruenpeter[^], Jennifer Harrow[˘], Anna-Lena Lamprecht[‡], Fotis E Psomopoulos[˚]

[‡] Netherlands eScience Center, Amsterdam, Netherlands

[§] School of Computer Science, University of Manchester, Manchester, United Kingdom

[|] University of Illinois, Urbana, Illinois, United States of America

[¶] Australian Research Data Commons, Sydney, Australia

[#] National Image Facility, Brisbane, Australia

[▫] Research Software Alliance, Cairns, Australia

[«] ZB MED Information Centre for Life Sciences, Cologne, Germany

[»] University of Edinburgh, Edinburgh, United Kingdom

[^] Software Heritage, Paris, France

[˘] ELIXIR Hub, Cambridge, United Kingdom

[‡] Utrecht University, Utrecht, Netherlands

[˚] Centre for Research and Technology Hellas, Thessaloniki, Greece

Corresponding author: Carlos Martinez-Ortiz (carlos.martinezortiz@gmail.com)

Abstract

In academic research virtually every field has increased its use of digital and computational technology, leading to new scientific discoveries, and this trend is likely to continue. Reliable and efficient scholarly research requires researchers to be able to validate and extend previously generated research results. In the digital era, this implies that digital objects Kahn and Wilensky 2006 used in research should be Findable, Accessible, Interoperable and Reusable (FAIR). These objects include (but are not limited to) data, software, models (for example, machine learning), representations of physical objects, virtual research environments, workflows, etc. Leaving any of these digital objects out of the FAIR process may result in a loss of academic rigor and may have severe consequences in the long term for the field, such as a reproducibility crisis. In this extended abstract, we focus on research software as a FAIR digital object (FDO).

The FDO framework De Smedt et al. 2020 describes FDOs as being actionable units of knowledge, which can be aggregated, analyzed, and processed by different types of algorithms. Such algorithms must be implemented by software in one form or another. The framework also describes large software stacks supporting FDOs enabling responsible data science and increasing reproducibility. This implies that software is a key ingredient of the FDO framework, and should adhere to the FAIR principles. Software plays multiple roles: it is a DO itself, it is responsible for creating new FDOs (e.g., data) and it helps to make them available to the public (e.g., via repositories and registries). However there is a need to specify in more detail how non-data DOs, in particular software, fit in this framework.

Different classes of digital objects have different intrinsic properties and ways to relate to other DOs. This means that while they, in principle, are subject to the high-level FAIR principles, there are also differences depending on their type and properties, requiring an adaptation so FAIR implementations are more aligned to the digital object itself. This holds true in particular to software. Software has intrinsic properties (executability, composite nature, development practices, continuous evolution and versioning, and packaging and distribution) and specific needs that must be considered by the FDO framework. For example, open source software is typically developed in the open on social coding platforms, where releases are distributed through package management systems, unlike data that is typically published in archival repositories. These social coding platforms do not provide long term archiving, permanent identifiers, or metadata, and package management systems, while somewhat better, similarly do not make a commitment to long term archiving, do not use identifiers that fit the scholarly publication system well, and provide metadata that may be missing key elements. The FAIR for research software (FAIR4RS, Chue Hong et al. 2021) working group has dedicated significant effort in building a community consensus around developing FAIR principles that are customized for research software, providing methods for researchers to understand and address these gaps.

In this presentation we will highlight the importance of software for the FAIR landscape and why different (but related) FAIR principles are needed for software (vs those originally developed for data). Our goal here is to contribute to building an FDO landscape together, where we consider all different types of digital objects that are essential in today's research, and we are enthusiastic about contributing our expertise on research software in helping shape this landscape.

Keywords

research software, FAIR software, FAIR4RS

Presenting author

Carlos Martinez-Ortiz

Presented at

First International Conference on FAIR Digital Objects, presentation

Conflicts of interest

References

- Chue Hong NP, Katz DS, Barker M, Lamprecht A-, Martinez C, Psomopoulos FE, Harrow J, Castro LJ, Gruenpeter M, Martinez PA, Honeyman T (2021) FAIR Principles for Research Software (FAIR4RS Principles. Research Data Alliance <https://doi.org/10.15497/RDA00065>
- De Smedt K, Koureas D, Wittenburg P (2020) FAIR Digital Objects for Science: From Data Pieces to Actionable Knowledge Units. Publications 8: 21. <https://doi.org/10.3390/publications8020021>
- Kahn R, Wilensky R (2006) A framework for distributed digital object services. Int J Digit Libr 6: 115-123. <https://doi.org/10.1007/s00799-005-0128-x>