

Improving FAIRness with containers

Dr. Ana Trisovic ¹ and Dr. Merce Crosas ¹

¹ Harvard University, US

Software

- [SORSE](#) ↗
- [Event Website](#) ↗

Category: talks

Published: August 18, 2020

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

The FAIR guiding principles state that published research objects should be made Findable, Accessible, Interoperable, and Reusable for other researchers. Data repositories provide research dissemination following FAIR principles while also developing standards and tools to facilitate them. However, increased use of advanced research methods, such as virtual containers, supercomputers and GPUs, is introducing new challenges for research sharing. There is no standardized way of describing and disseminating such research outputs in data repositories. Furthermore, dissemination of data within virtual containers like Docker may hinder some of the commonly supported principles, such as findability and accessibility.

Improvements in documenting and sharing of advanced research materials in a FAIR-compliant way has recently been natively enabled in new container-based tools. This talk will present insights from community discussions on improving FAIRness with encapsulation in the context of the Dataverse data repository platform. Dataverse has a long-term commitment to preserving research artifacts to enable research transparency, reproducibility, and reuse. We will provide a review of a number of projects based on virtual containers to determine their potentials for enabling FAIRness of published research. In particular, we will review the following projects: Open Container Initiative, Singularity, ReproZip, and SciUnit, and their efforts on documenting and describing researcher's computing environments. Some of the new concepts, such as data containers and Binder Boxes will also be considered. We evaluate each of these approaches against established FAIR principles in the context of dissemination in data repositories.