

EGU2020-10939

<https://doi.org/10.5194/egusphere-egu2020-10939>

EGU General Assembly 2020

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



ERDA: External Version-Controlled Research Data

Willi Rath, Carsten Schirnack, and Claas Faber

GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel, Kiel, Germany

This presentation will detail the design, implementation, and operation of ERDA, which is a collection of external version-controlled research datasets, of multiple synchronized deployments of the data, of a growing set of minimal examples using the datasets from various deployments, of stand-alone tools to create, maintain, and deploy new datasets, and of documentation targeting different audiences (users, maintainers, developers).

ERDA was designed with the following principles in mind: Provide clear data provenance and ensure long-term availability, minimize effort for adding data and make all contents available to all users immediately, ensure unambiguous referencing and develop transparent versioning conventions, embrace mobility of scientists and target independence from the infrastructure of specific institutions.

The talk will show how the data management is done with Git-LFS, demonstrate how data repositories are rendered from human-readable data, and give an overview of the versioning scheme that is applied.