

EGU2020-7937

<https://doi.org/10.5194/egusphere-egu2020-7937>

EGU General Assembly 2020

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



A Lightweight, Microservice-Based Research Data Management Architecture for Large Scale Environmental Datasets

Alexander Götz, Johannes Munke, Mohamad Hayek, Hai Nguyen, Tobias Weber, Stephan Hachinger, and **Jens Weismüller**

Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities, Boltzmannstraße 1, 85748 Garching bei München

LTDS ("Let the Data Sing") is a lightweight, microservice-based Research Data Management (RDM) architecture which augments previously isolated data stores ("data silos") with FAIR research data repositories. The core components of LTDS include a metadata store as well as dissemination services such as a landing page generator and an OAI-PMH server. As these core components were designed to be independent from one another, a central control system has been implemented, which handles data flows between components. LTDS is developed at LRZ (Leibniz Supercomputing Centre, Garching, Germany), with the aim of allowing researchers to make massive amounts of data (e.g. HPC simulation results) on different storage backends FAIR. Such data can often, owing to their size, not easily be transferred into conventional repositories. As a result, they remain "hidden", while only e.g. final results are published - a massive problem for reproducibility of simulation-based science. The LTDS architecture uses open-source and standardized components and follows best practices in FAIR data (and metadata) handling. We present our experience with our first three use cases: the Alpine Environmental Data Analysis Centre (AlpEnDAC) platform, the ClimEx dataset with 400TB of climate ensemble simulation data, and the Virtual Water Value (ViWA) hydrological model ensemble.