Research Ideas and Outcomes 8: e95817
doi: 10.3897/rio.8.e95817

OPEN ACCESS

CrossMark

Conference Abstract

# Facing the Challenges in simulation-based Earth System Sciences and the Role of FAIR Digital Objects

Ivonne Anders[‡], Karsten Peters-von Gehlen[‡], Martin Bergemann[‡], Hannes Thiemann[‡]

‡ DKRZ - German Climate Computing Center, Hamburg, Germany

Corresponding author: Ivonne Anders (anders@dkrz.de), Karsten Peters-von Gehlen (peters@dkrz.de)

Received: 29 Sep 2022 | Published: 12 Oct 2022

Reviewable v 1

Citation: Anders I, Peters-von Gehlen K, Bergemann M, Thiemann H (2022) Facing the Challenges in simulation-based Earth System Sciences and the Role of FAIR Digital Objects. Research Ideas and Outcomes 8: e95817. https://doi.org/10.3897/rio.8.e95817

## Abstract

### Motivation

Results of simulations with climate models form the most important basis for research and statements about possible changes in the future global, regional and local climate. These output volumes are increasing at an exponential rate (Balaji et al. 2018, Stevens et al. 2019). Efficiently handling these amounts of data is a challenge for researchers, mainly because the development of novel data and workflow handling approaches have not proceeded at the same rate as data volume has been increasing. This problem will only become more pronounced with the ever increasing performance of High Performance Computing (HPC) - systems used to perform weather and climate simulations (Lawrence et al. 2018). For example, in the framework of the European Commission's Destination Earth program the Digital Twins (Bauer et al. 2021) are expected to produce hundreds of terabytes of model output data every day at the EuroHPC computing sites.

The described data challenge can be dissected into several aspects, two of which we will focus on in this contribution. Available data in the Earth System Sciences (ESS) are increasingly made openly accessible by various institutions, such as universities, research centres and government agencies, in addition to subject-specific repositories. Further, the exploitability of weather and climate simulation output beyond the expert community by

humans and automated agents (as described by the FAIR data principles (F-Findable, A-Accessable, I-Interoperable, R-Reusable), Wilkinson et al. 2016) is currently very limited if not impossible due to disorganized metadata or incomplete provenance information. Additionally, developments regarding globally available and FAIR workflows in the spirit of the FAIR Digital Object (FDO) framework (Schultes and Wittenburg 2019, Schwardmann 2020) are just at the beginning.

## Cultural Change

In order to address the challenges with respect to data mentioned above, current efforts at DKRZ (German Climate Computing Center) are aimed at a complete restructuring of the way research is performed in simulation-based climate research (Anders et al. 2022, Mozaffari et al. 2022, Weigel et al. 2020). DKRZ is perfectly suited for this endeavor, because researchers have the resources and services available to conduct the entire suite of their data-intensive workflows - ranging from planning and setting up of model simulations, analyzing the model output, reusing existing large-volume datasets to data publication and long-term archival. At the moment, DKRZ-users do not have the possibility to orchestrate their workflows via a central service, but rather use a plethora of different tools to piece them together.

## Framework Environment Freva

The central element of the new workflow environment at DKRZ shall be represented by the Freva (Free Evaluation System Framework) software infrastructure, which offers standardized data and tool solutions in ESS and is optimized for use on high-performance computer systems (Kadow et al. 2021). Freva is designed to be very well suited to the use of the FDO framework. The crucial aspects here are:

1. the standardisation of data objects as input for analysis and processing,
2. the already implemented remote access to data via a Persisitent Identifier (PID),
3. the currently still system-internal capture of analysis provenance and
4. the possibility of sharing results but also workflows by research groups up to large communities.

It is planned to extend the functionality of Freva so that the system automatically determines the data required for a specific analysis from a researcher's research question (provided to the system via some interface), enquires available databases (local disk or tape, cloud or federated resources) for that data and retrieves the data if possible. If data are not available (yet), Freva shall be able to automatically configure, set up and submit model simulations to the HPC-System, so that the required data is created and becomes available (cf. Fig. 1). These data will in turn be ingested into Freva's data catalog for reuse. Next, Freva shall orchestrate and document the analysis performed. Results will be provided either as numerical fields, images or animations depending on the researcher's need. As a final step, the applied workflow and/or underlying data are published in accordance with the FAIR data guiding principles.

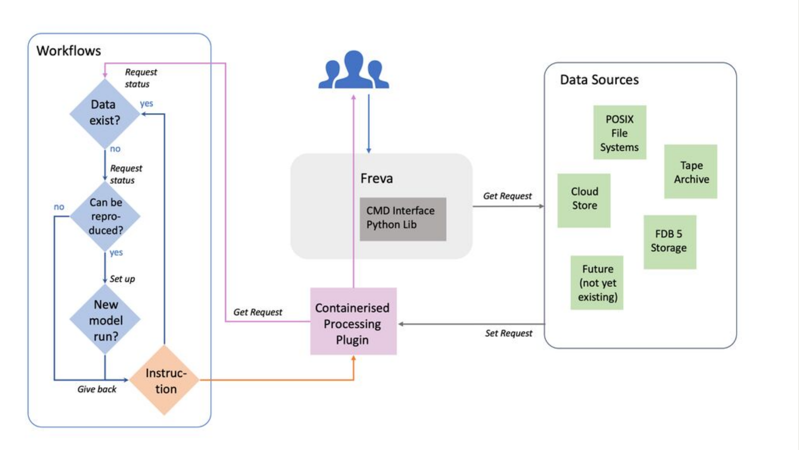

Figure 1. doi

Next development steps within Freva.

### FDOs - towards a global integrated Data Space

To make the process sketched out above a reality, application of the FDO concept is essential (Schwardmann 2020, Schultes and Wittenburg 2019). There is a long tradition in the ESS community of global dissemination and reuse of large-volume climate data sets. Community standards like those developed and applied in the framework of internationally coordinated model intercomparison studies (CMIP) allow for low-barrier reuse of data ( Balaji et al. 2018). Globally resolvable PIDs are provided on a regular basis. Current community ESS standards and workflows are already close to being compatible with implementing FDOs, however, now we also have to work on open points in the FDO concept, which are:

1. the clear definition of community-specific FDO requirements including PID Kernel Types specifications,
2. the operation of data type registries and
3. the technical implementation requirements for global access to FDOs.

With these in place and implemented in Freva following standardized implementation recommendations, automated data queries across spatially distributed or different types of local databases become possible.

We introduce the concept of implementations in Freva and also use it to highlight the challenges we face. Using an example, we show the vision of the work of a scientist in earth system science.

## Keywords

FAIR Digital Objects, HPC infrastructures, Data services, Climate science, Freva

## Presenting author

Karsten Peters-von Gehlen

## Presented at

First International Conference on FAIR Digital Objects, presentation

## References

- Anders I, Peters-von Gehlen K, Thiemann H (2022) Canonical Workflows in Simulation-based ClimateSciences. Data Intelligence 4 (2): 212-225. https://doi.org/10.1162/dint_a_00127
- Balaji V, Taylor K, Juckes M, Lawrence B, Durack P, Lautenschlager M, Blanton C, Cinquini L, Denvil S, Elkington M, Guglielmo F, Guilyardi E, Hassell D, Kharin S, Kindermann S, Nikonov S, Radhakrishnan A, Stockhause M, Weigel T, Williams D (2018) Requirements for a global data infrastructure in support of CMIP6. Geoscientific Model Development 11 (9): 3659-3680. https://doi.org/10.5194/gmd-11-3659-2018
- Bauer P, Stevens B, Hazeleger W (2021) A digital twin of Earth for the green transition. Nature Climate Change 11 (2): 80-83. https://doi.org/10.1038/s41558-021-00986-y
- Kadow C, Illing S, Lucio-Eceiza E, Bergemann M, Ramadoss M, Sommer P, Kunst O, Schartner T, Pankatz K, Grieger J, Schuster M, Richling A, Thiemann H, Kirchner I, Rust HW, Ludwig T, Cubasch U, Ulbrich U (2021) Introduction to Freva – A Free Evaluation System Framework for EarthSystem Modeling. Journal of Open Research Software 9 (1). https://doi.org/10.5334/jors.253
- Lawrence B, Rezny M, Budich R, Bauer P, Behrens J, Carter M, Deconinck W, Ford R, Maynard C, Mullerworth S, Osuna C, Porter A, Serradell K, Valcke S, Wedi N, Wilson S (2018) Crossing the chasm: how to develop weather and climate models for next generation computers? Geoscientific Model Development 11 (5): 1799-1821. https://doi.org/10.5194/gmd-11-1799-2018
- Mozaffari A, Langguth M, Gong B, Ahring J, Campos AR, Nieters P, Escobar OJC, Wittenbrink M, Baumann P, Schultz M (2022) HPC-oriented Canonical Workflows for Machine Learning Applications in Climate and Weather Prediction. Data Intelligence 4 (2): 271-285. https://doi.org/10.1162/dint_a_00131
- Schultes E, Wittenburg P (2019) FAIR Principles and Digital Objects: Accelerating Convergence on a Data Infrastructure. Communications in Computer and Information Science3-16. https://doi.org/10.1007/978-3-030-23584-0_1
- Schwardmann U (2020) Digital Objects – FAIR Digital Objects: Which Services Are Required? Data Science Journal 19 https://doi.org/10.5334/dsj-2020-015
- Stevens B, Satoh M, Auger L, Biercamp J, Bretherton C, Chen X, Düben P, Judt F, Khairoutdinov M, Klocke D, Kodama C, Kornblueh L, Lin S, Neumann P, Putman W, Röber N, Shibuya R, Vanniere B, Vidale PL, Wedi N, Zhou L (2019) DYAMOND: the DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains. Progress in Earth and Planetary Science 6 (1). https://doi.org/10.1186/s40645-019-0304-z

- Weigel T, Schwardmann U, Klump J, Bendoukha S, Quick R (2020) Making Data and Workflows Findable for Machines. Data Intelligence 2: 40-46. https://doi.org/10.1162/dint_a_00026
- Wilkinson M, Dumontier M, Aalbersberg IJ, et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3 (160018). https://doi.org/10.1038/sdata.2016.18