International FAIR Digital Objects Implementation Summit

Extended Abstracts

https://doi.org/10.52825/ocp.v5i.1175

Published: 18 Mar. 2025

Improving the Findability of Digital Objects in Climate Science by Adopting the FDO Concept

Marco Kulüke[1,*], Karsten Peters-von Gehlen[1], and Ivonne Anders[1]

[1]German Climate Computing Center (DKRZ), Germany

*Correspondence: Marco Kulüke, kulueke@dkrz.de

Abstract. Climate science relies heavily on the effective creation, management, sharing, and analysis of massive and diverse datasets. As the digital landscape evolves, there is a growing need to establish a framework that ensures FAIRness in handling climate science digital objects. Especially, the machine-to-machine actionability of digital objects will be a crucial step towards future AI assisted workflows. Motivated by a use case, this contribution proposes the adoption of the Fair Digital Object (FDO) concept to address the challenges associated with the emerging spread in interdisciplinary reuse scenarios of climate model simulation output.

The aim of this work is to commit to an FDO standard in climate science that enables standardized and therefore automated data analysis workflows and facilitates the extraction and analysis of relevant weather and climate data by all stakeholders involved. The current work expands on the efforts made to enable broad reuse of CMIP6 [1, 2] climate model data and focuses on requirements identified to enable automated processing of climate simulation output and their possible implementation strategies. The exemplary use case of an automated, prototypical climate model data analysis workflow will showcase the obstacles occurring when analyzing currently available climate model data. In particular, the findability of digital objects required for a particular research question in climate science or a related field shows to be challenging. In order to mitigate this issue, certain strategies are proposed: (a) Enriching the PID profiles of climate model data in accordance with the FDO concept and taking into account the needs of the climate science community will lead to improved findability of digital objects, especially for machines. (b) Defining a standardized, unique association between climate model variables and their meaningful long names will increase the findability of climate model data, especially for researchers in other disciplines. (c) Furthermore, combining the FDO concept with existing data management solutions, such as the intake-esm catalogs [3], can lead to improved data handling in line with prevailing community practices.

Eventually, implementing an FDO standard will benefit the climate science community in several ways: The reusability of the data will facilitate the cost-effective use of existing computationally expensive climate model data. Improved data citation practices will promote data sharing, and ultimately, high transparency will increase the reproducibility of research workflows and consolidate scientific results.

Keywords: Climate Science, Data Management, Machine-to-Machine Actionability, FAIR Digital Objects

Data availability statement

The work presented here focuses on digital objects within the CMIP6 database [2], which already have a high level of standardization and comply with the majority of the FAIR data principles. These data are used to establish a baseline representing the current state of FAIRness of climate simulation output. Note that only those CMIP6 data are taken into consideration which are made available directly via the Earth System Grid Federation (ESGF) [4].

Underlying and related material

Jupyter Notebooks which are closely related to the contribution are deposited on a repository: https://gitlab.dkrz.de/data-infrastructure-services/fdo

Author contributions

According to the CreDIT guidelines all authors (M.K., K.PvG., and I.A.) conceptualized this contribution and developed a methodology. M.K. wrote software and the original draft, which was reviewed and edited by K.PvG and I.A.

Competing interests

The authors declare that they have no competing interests.

Funding

Funded by the European Union. This work has received funding from the European High Performance Computing Joint Undertaking (JU) under grant agreement No 101093054.

Acknowledgement

This work used resources of the German Climate Computing Center (DKRZ).

References

[1] Coupled Model Intercomparison Project. https://wcrp-cmip.org/ (accessed 29.02.2024)
[2] Eyring, V., G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor. 2016. “Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization.” *Geoscientific Model Development* 9:1937--1958. https://doi.org/10.5194/gmd-9-1937-2016.
[3] Intake-ESM data cataloging. https://intake-esm.readthedocs.io/ (accessed 29.02.2024)
[4] Earth System Grid Federation. https://esgf-data.dkrz.de/projects/esgf-dkrz/ (accessed 29.02.2024)