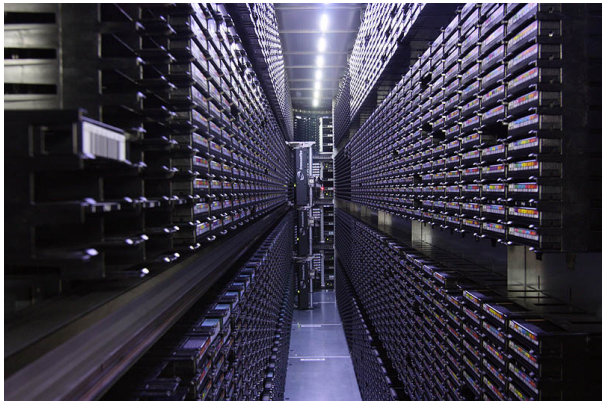
Improving the Findability of Digital Objects in Climate Science by adopting the FDO Concept

Marco Kulüke, Karsten Peters-von Gehlen, Ivonne Anders

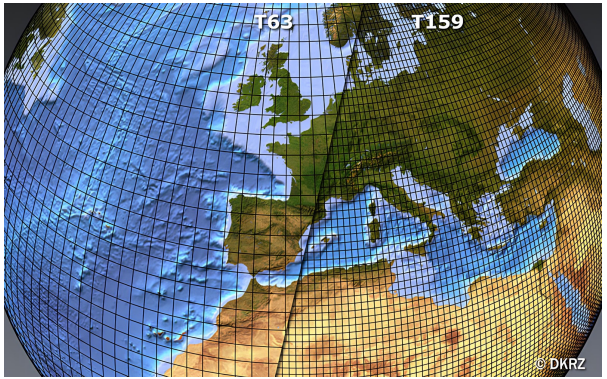
Deutsches Klimarechenzentrum (DKRZ)

Status Quo

- Large-volume climate model data
- Interdisciplinary data (re)use
- Machine to machine interaction



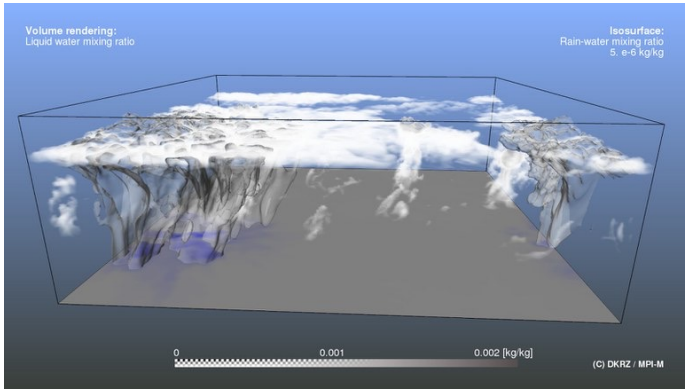
[dkrz.de]

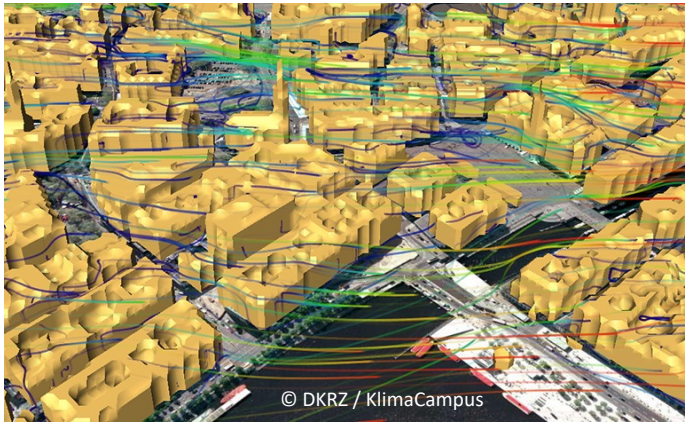


[dkrz.de]

Challenges

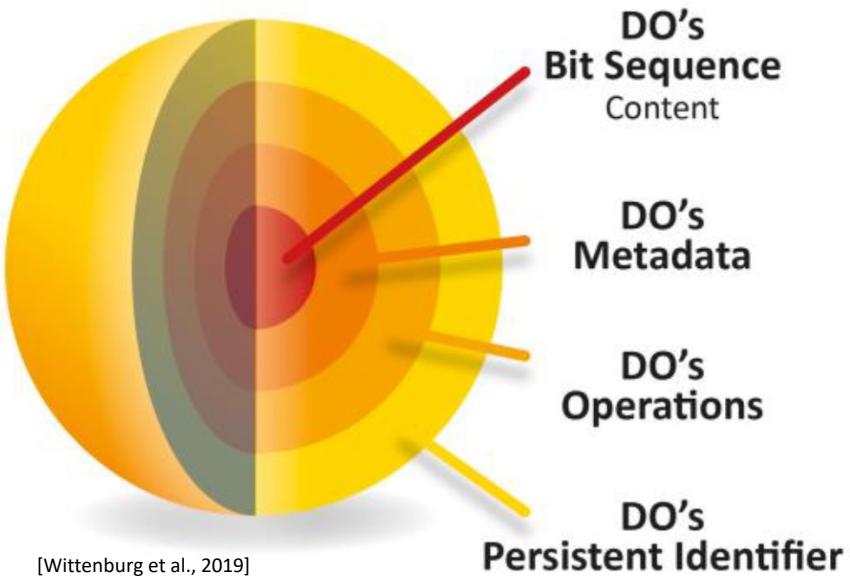
- Hard to find the right data
- Increasing workflow complexity
- Increasing data amounts





What are FAIR Digital Objects?

- Digital Objects (DOs) optimized for M2M actionability
- Includes data, metadata, documents, software
- *Concept of specifications* driven by international & interdisciplinary working groups



[Wittenburg et al., 2019]

FAIR DIGITAL OBJECTS FORUM

https://fairdo.org/

Imagine FDO as DO with a “Passport”

Standardized format

Machine and human readable

Provenance information

Access rights

Author attribution

Location

Verifiable

Unique persistent identifier

Metadata about content

License

FDO-readiness of CMIP6 Data*

Systematic review from researcher perspective:

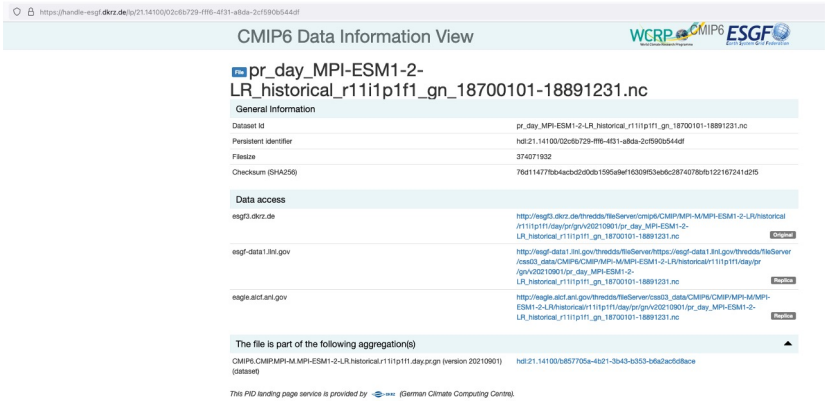
- M2M actionable?
- Encourage Cross-disciplinary research?

[Kulüke et al., 2023]

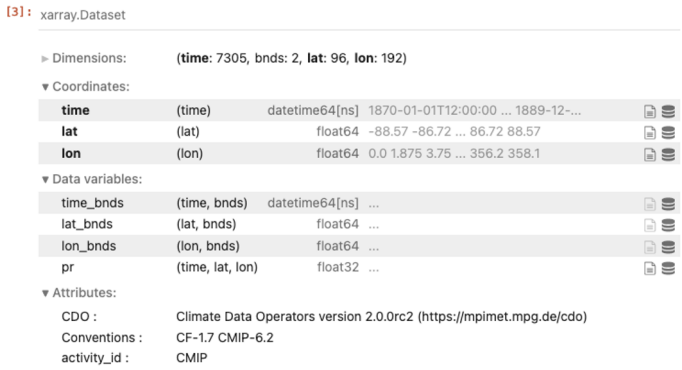
* 6th Phase of the Coupled Model Intercomparison Project
[Eyring et al., 2016; Petrie et al., 2021]

Findable

- **Obstacle 1:** no mapping between meaningful names and model variable names
- **Obstacle 2:** PID profile does not contain variable info



PID profile of CMIP dataset



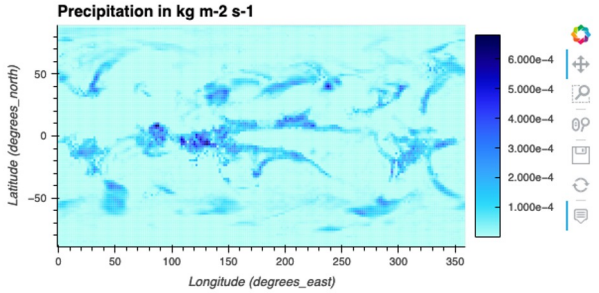
NetCDF header of same dataset

Accessible & Interoperable

hdl:21.14100/0163da97-35d1-387c-a51b-0d12fc4d3b24

jupyter

notebook



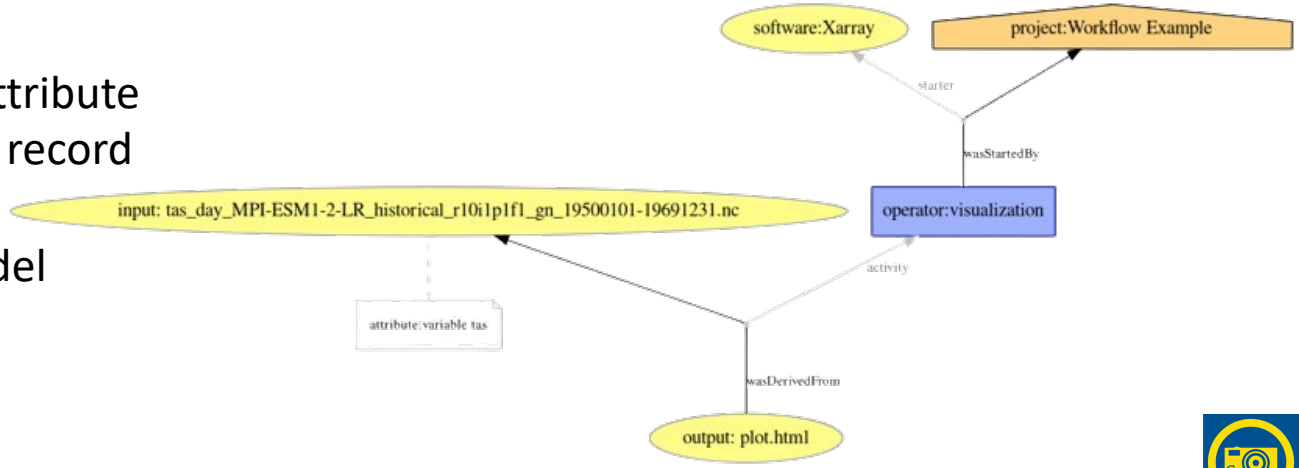
CMIP6 DOs are machine accessible
and to a large extend interoperable

Reusable

CMIP6 DOs have *history* attribute
with sporadic provenance record

→ Adopt PROV data model

[Belhajjame et al., 2013]



How can FDO Concept help?

- Enrich data PID profiles
- Data catalogs as FDOs
- Break down jargon barriers

INTAKE

[faircore4eosc.eu]

Use Case: Apply FDO Concept to ESM Data

- 1 PB high-res ESM data derived from nextGEMS project[1]
- Stored on HEALPix grid[2] → allows hierarchical output and horizontal chunking

Publish Intake / STAC catalog at WDCC[3] with persistent identifier

[1] https://nextgems-h2020.eu/
[2] https://healpix.sourceforge.io/
[3] https://www.wdc-climate.de/

Thank You!

References

Wittenburg, P.; Strawn, G.; Mons, B.; Bonino, L.; Schultes, E. Digital Objects as Drivers towards Convergence in Data Infrastructures; EUDAT: Helsinki, Finland, 2019. https://doi.org/10.23728/b2share.b605d85809ca45679b110719b6c6cb11

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

Petrie, Ruth, Sebastien Denvil, Sasha Ames, Guillaume Levavasseur, Sandro Fiore, Chris Allen, Fabrizio Antonio, et al. 2021. "Coordinating an operational data distribution network for CMIP6 data." Geoscientific Model Development 14:629–644. https://doi.org/10.5194/gmd-14-629-2021

Kulüke, M., Peters-von Gehlen, K., & Anders, I. (2023). Strategy and Requirements for FDO-Standard in Simulation-based Climate Science - Milestone M3.1. Zenodo. https://doi.org/10.5281/zenodo.10423391

Belhajjame, K., R. B'Far, J. Cheney, S. Coppens, S. Cresswell, Y. Gil, P. Groth, et al. 2013. "PROV-DM: The PROV Data Model." Edited by L. Moreau and P. Missier. W3C. https://www.w3.org/TR/2013/REC-prov-dm-20130430/

esiwace
CENTRE OF EXCELLENCE IN SIMULATION OF WEATHER AND CLIMATE IN EUROPE

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

NFDI4Earth

eosc | FAIRCORE4EOSC
Core Components Supporting a FAIR EOSC

Funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) project number: 460036893