

# PID-Driven Global Access to Flagship km-scale Climate Simulation Data

Karsten Peters-von Gehlen<sup>1,\*</sup>, Kameswar Rao Modali<sup>1</sup>, Florian Ziemer<sup>1</sup>, Martin Bergemann<sup>1</sup>, Christopher Kadow<sup>1</sup>, Karl-Hermann Wieners<sup>2</sup>, Siddhant Tibrewal<sup>2</sup>, Ivonne Anders<sup>1</sup>, Katharina Berger<sup>1</sup>, Tobias Kölling<sup>2</sup>, Lukas Kluff<sup>2</sup>, Marco Kulüke<sup>1</sup> and Fabian Wachsmann<sup>1</sup>

<sup>1</sup> Deutsches Klimarechenzentrum (DKRZ), Hamburg, Germany; <sup>2</sup> Max-Planck-Institut für Meteorologie (MPI-M), Hamburg, Germany



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



Max-Planck-Institut für Meteorologie

## What is the issue?

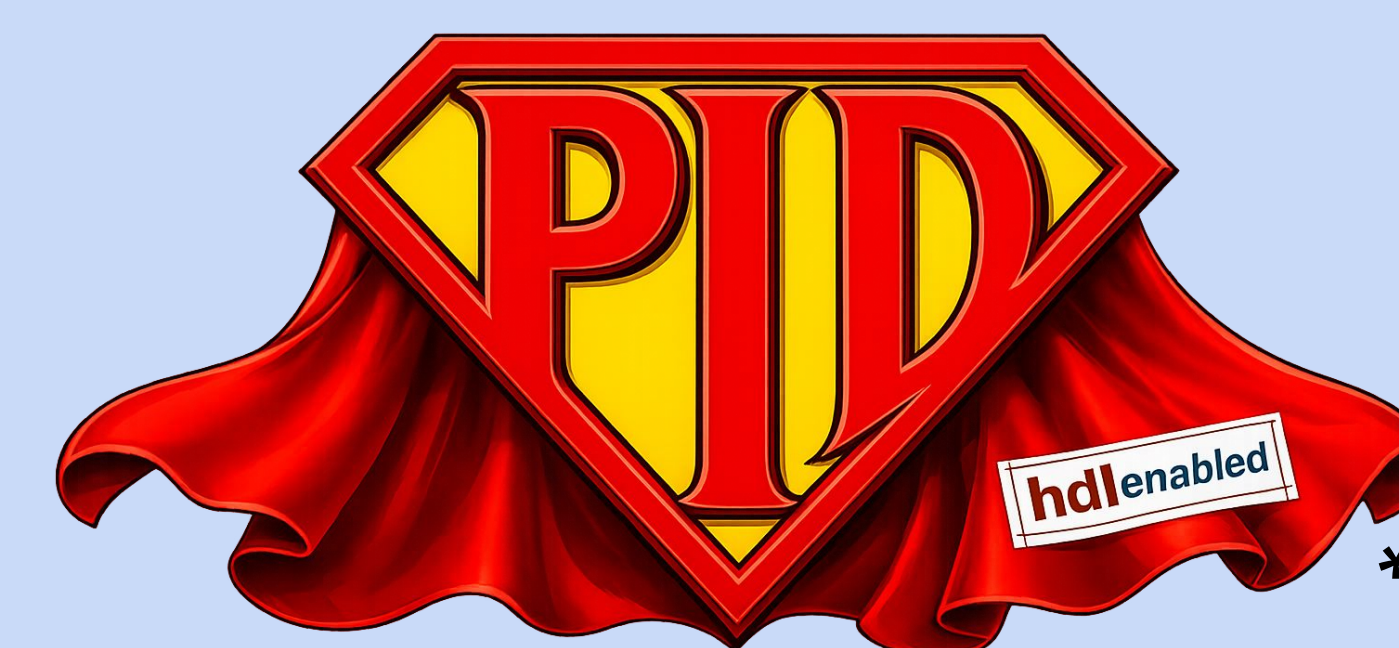
km-scale ESM simulations are often performed in loosely coordinated frameworks or are part of single flagship projects

→ km-scale model simulation output is not made available via dedicated federated infrastructure

Hence, simulation output is often stored using institutional resources, with limited accessibility to outside users

→ large interest of global community meets complicated data access protocols/methods

## PIDs to the rescue

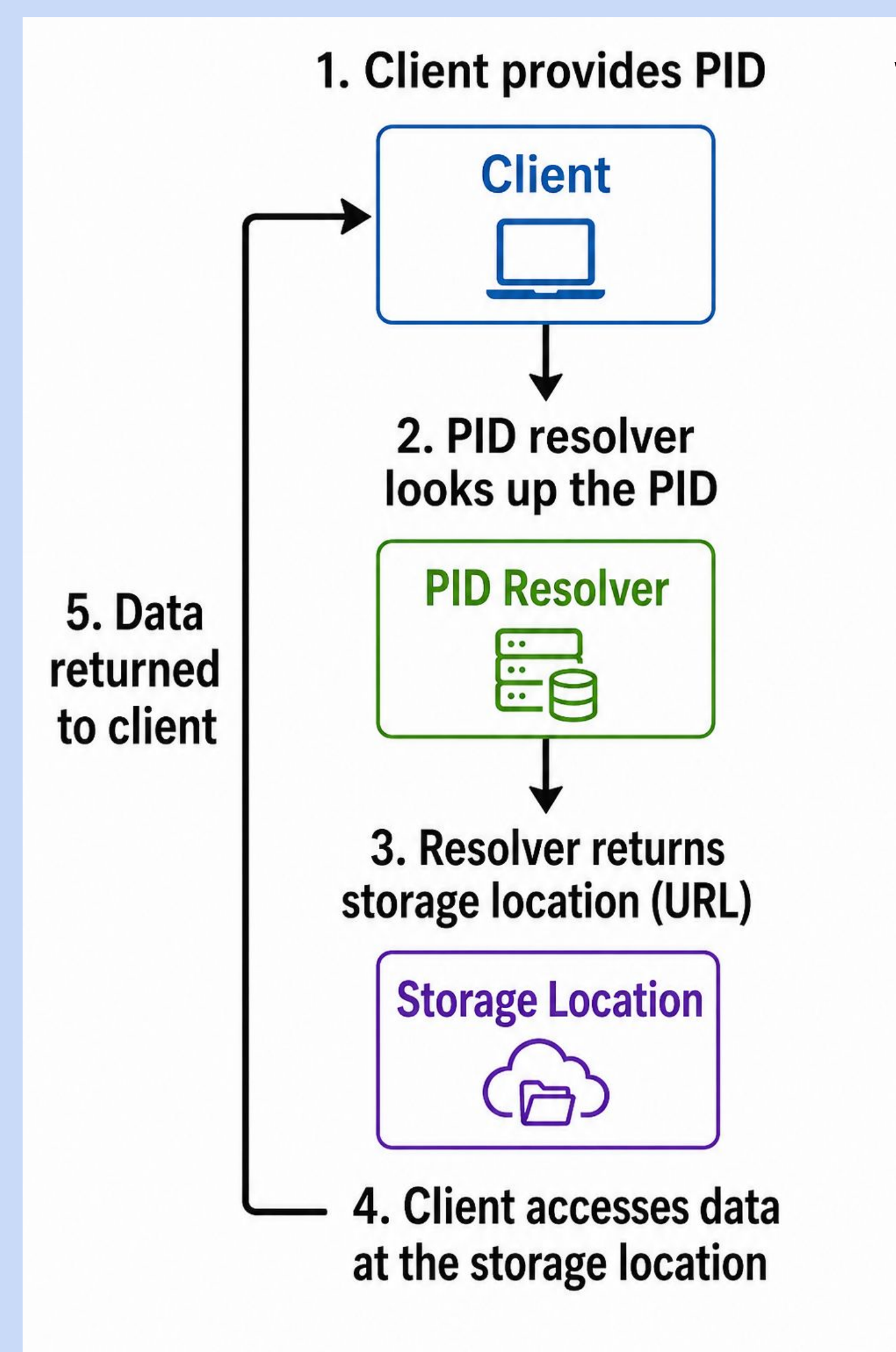


What's their superpower?

The handle profile!

- collection of key-value-pairs shipped with the handle
- can contain (almost unlimited) useful metadata for data access and (re)use

**PIDs provide the basis for automated data findability, accessibility and processing!**



## Prototypical implementation

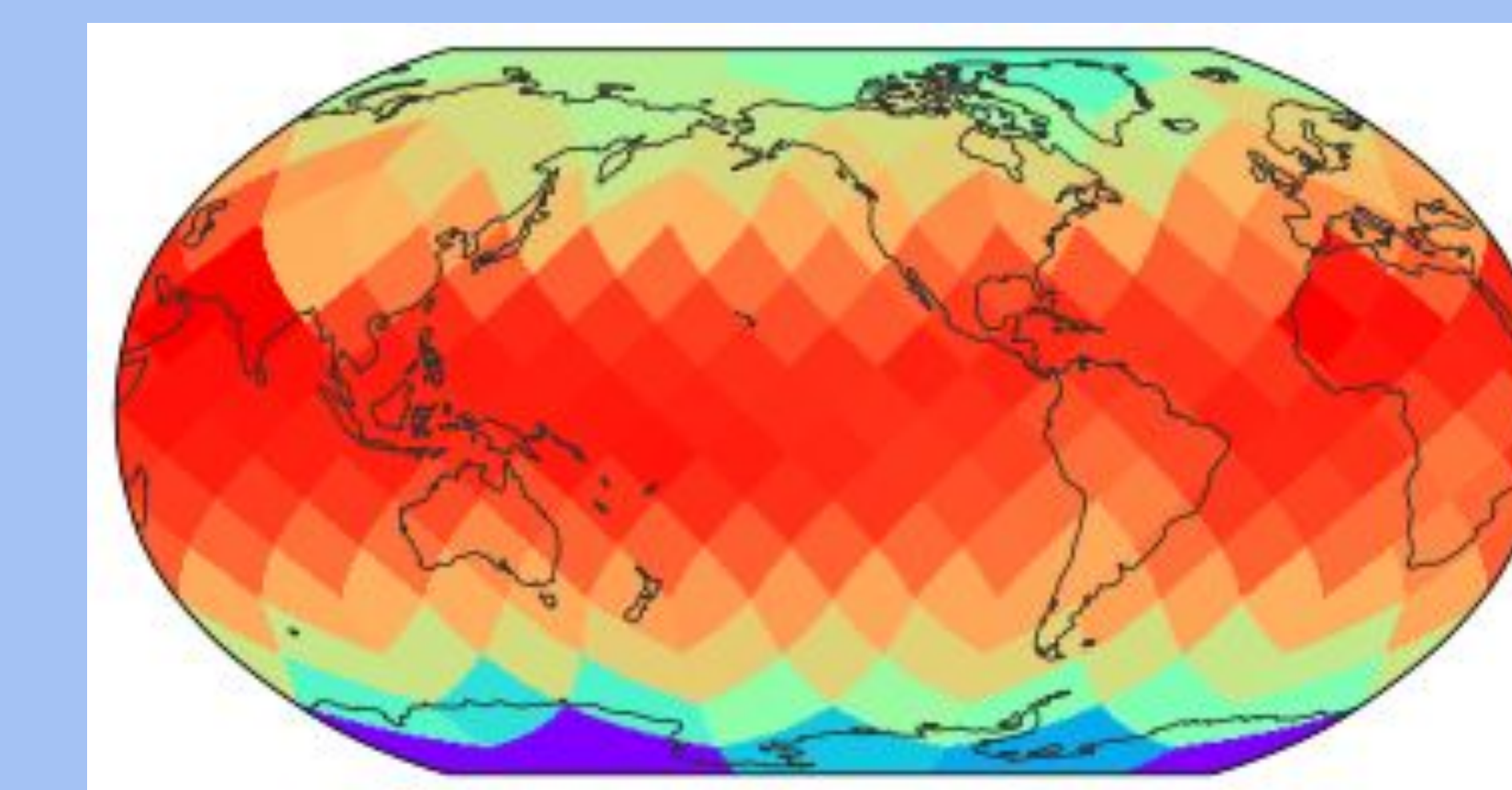
Dataset: 2.5km ICON simulation created for the WCRP km-scale Hackathon (May 2025)



PID: [https://hdl.handle.net/21.14102/EGU26\\_test](https://hdl.handle.net/21.14102/EGU26_test), to view the profile, add ?noredirect at the end



Example access/analysis code snippet: <https://tinyurl.com/PIDforESMdemo>



## One step towards FDOs!!

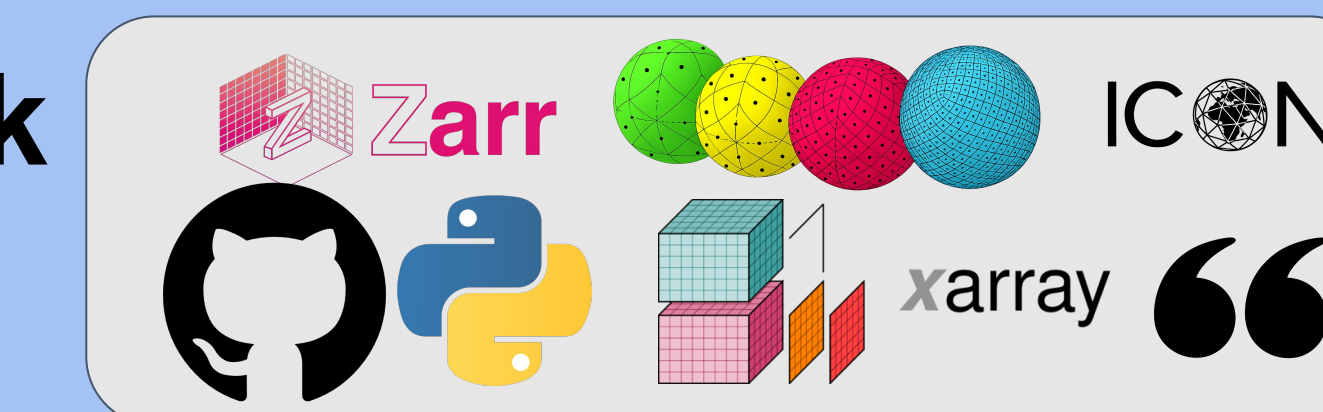
FDOs are digital assets optimized for machine-actionability

Machine-actionable, specific metadata encapsulated in the access method



What is missing to make it work seamlessly?

- automated PID generation
- consolidation of required metadata entries
- ensure compliance with FDO specs



## What would be really cool?



## What do we need for it to work?

PID-profile populated with metadata that is actually useful for access and analysis

Model simulation and output configuration, source code provenance, authors, ... sourced from model setup routine

PID assigned ideally at model runtime; own PID server

DKRZ runs their own PID server, PID creation at start of model run

Publicly accessible ESM data, e.g. on an S3 store, via URL

DKRZ S3 storage provides URL at dataset level for easy sharing with global community

Helper functions enabling hassle-free data ingestion also for non-experts

Openly accessible python functions extracting PID profile information made available via DKRZ gitlab

