IG High-Performance Computing in Earth System Sciences (HPC in ESS)

NFDI4Earth Kick-Off Dresden 9/10 June 2022

For the IG: Stephan Hachinger, Leibniz Supercomputing Centre (LRZ, Garching b.M.)

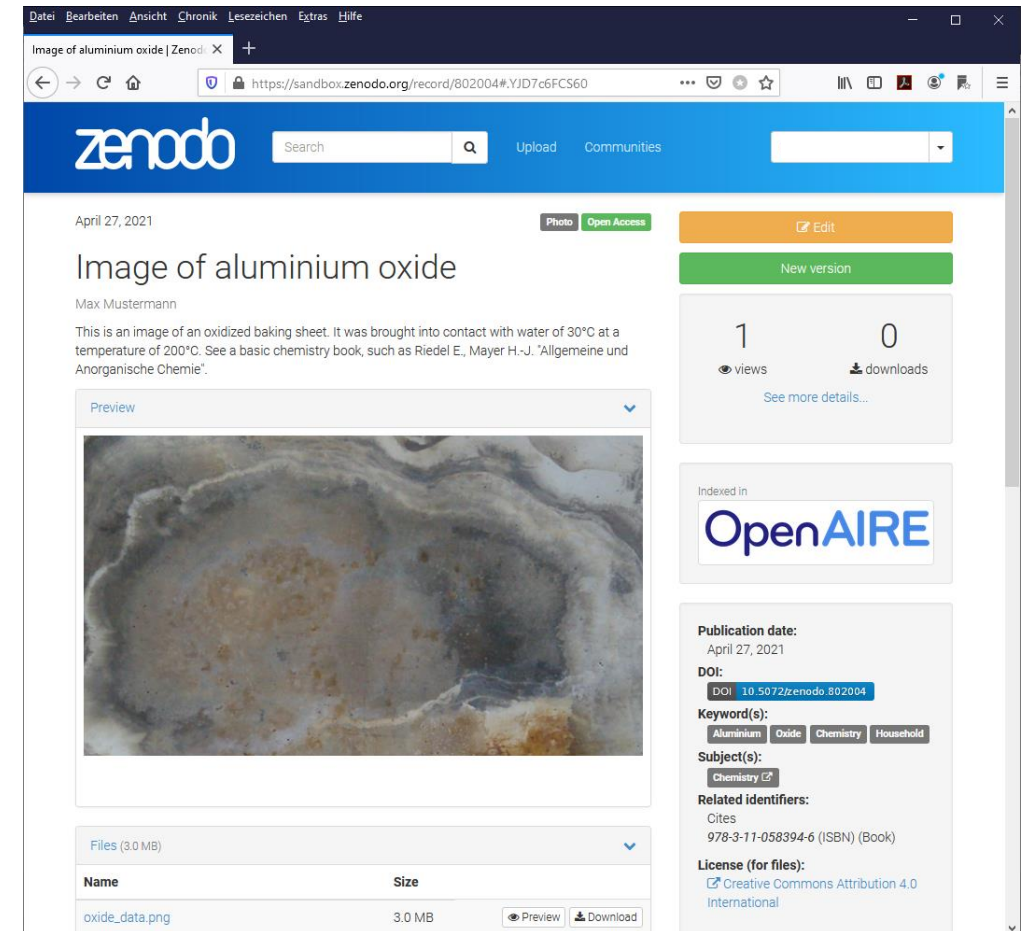# IG High-Performance Computing in ESS: Motivation

In HPC, FAIR methodology is a challenge!

Consider the role of repositories:

- offer space to deposit data and descriptive (& other) metadata

- make data findable, get you DOIs

- fulfill requirements of funding agencies

Problem: maybe 50 GB of storage is not quite enough for your multi-Terabyte HPC data.
→ Standard repositories are out!

IG HPC in ESS: Motivation

FAIR „vs." HPC in ESS: Lots of challenges!

- **Technical**:   Datasets produced too large for repositories – satellite data & global/large-scale high-resolution datasets

- **Methodical:**   Reproducibility – simulation/analytics output depends on compiler switches, libraries, etc. "Wild" community software usage: netCDF, eccodes, …

- **Administrative:**   Computing needs grants and accounting, but researchers want to quickly work where their data are.

- **Practical:**   "Everyone talks while I have 1PB to handle!"

We have collected challenges on a „living poster" on this conference!
Tell us what is important to you & give your opinions!

IG High-Performance Computing in ESS: Context
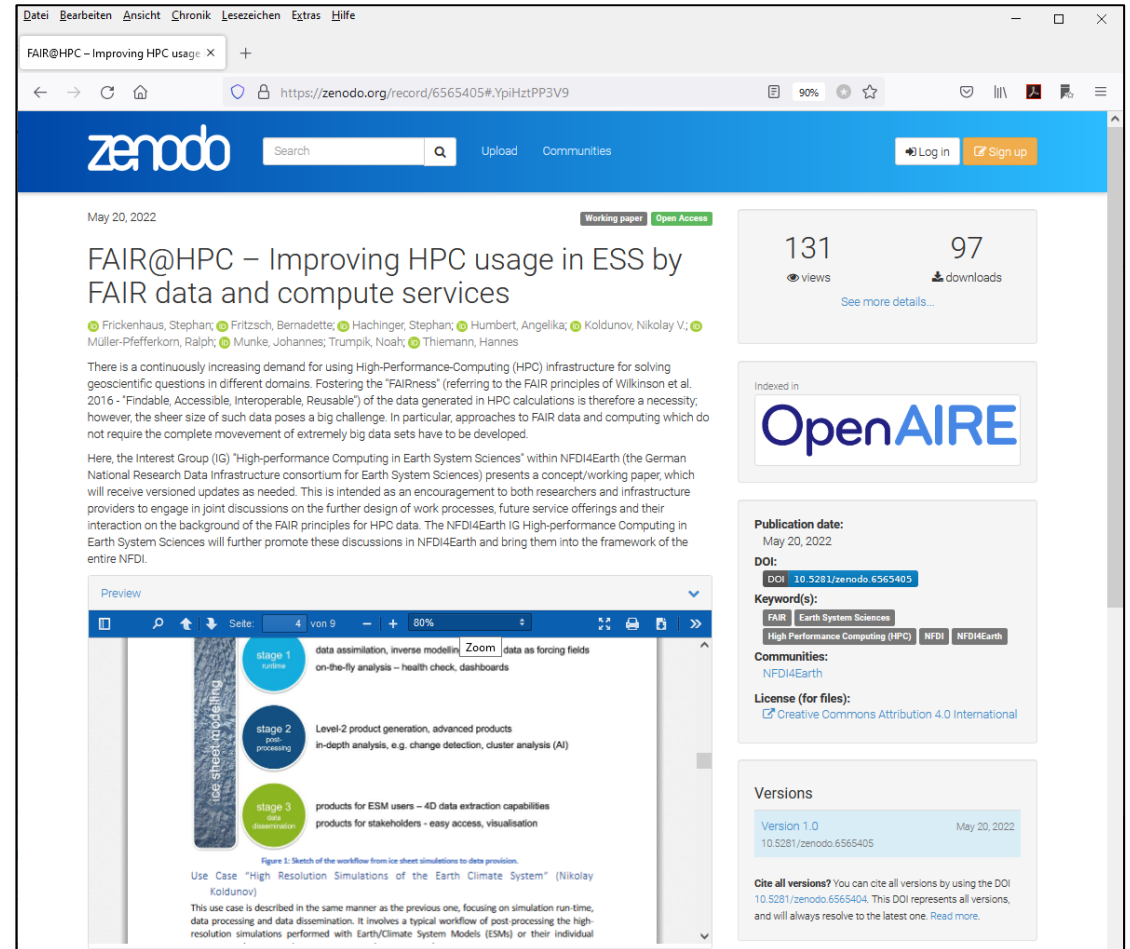## Scientists + Computing/Data Centres: NHR, Gauss-Allianz, GCS, …

IG HPC in ESS: Format / Aims

- Every 2nd Monday in month: meet 9:00 am: https://meet.lrz.de/ig-hpc-nfdi4earth

- Some goals:
  - semi-automatically <u>enrich results from HPC systems with metadata</u>
  - produce <u>(meta-)data according to NFDI4Earth standards</u>
  - steps towards federated <u>'in place' data-analysis facilities</u>
  - foster and simplify <u>re-use</u> of HPC data

- Current „projects"
  - concept paper "Use Cases → Challenges → Recommendations for Development"
  - workshop with computing centres (NHR, Gauss-Allianz, GCS)
  - contact with NFDI sections (Metadata, Common Infra)
  - tech research projects: e.g. on latencies in distributed data MGMT

IG HPC in ESS: First Result – Concept Paper

- IG activities – two „pillars":
  - (Meta-)data formats, interoperability and reproducibility (→ reusability)
  - Federated access, findability and accessibility

- Two use cases & their needs examined
  - Ice sheet simulations
  - High-resolution climate simulations

- Challenges identified
  - In HPC/system usage
  - In FAIRness, sharing, re-use

- Recommendations (monitor every 1.5 yrs)
  - Infrastructure registry & unified access
  - Homogeneous computing environments
  - Standard formats (data/metadata)
  - Standardised (meta)data access / portal solutions

Concept paper:
https://doi.org/10.5281/zenodo.6565405

Thank You !

https://www.nfdi4earth.de/

Want to get active with us ? – Join by mail to hachinger [@] lrz.de !