High-performance computing in Earth System Sciences - striving for FAIRness
— Proposal for an NFDI4Earth interest group —

Stephan Frickenhaus[1], Stephan Hachinger[2], Wolfgang Kurtz[2],
Ralph Müller-Pfefferkorn[3], Hannes Thiemann[4]

[1]Alfred-Wegener-Institute (AWI)
[2]Leibniz Supercomputing Centre (LRZ)
[3]Technical University Dresden (TUD)
[4]German Climate Computing Center (DKRZ)

November 11, 2020

Motivation

- Increasing demand for HPC for geoscientific simulations (e.g. climate) and data analytics (e.g. remote sensing)

- HPC topics mentioned throughout the NFDI4Earth proposal

- Making data from Scientific Computing FAIR is a 'hot topic'
    - Technical challenges: Output data often too large to be moved to repositories.
    - Methodical challenges: Reproducibility – exact simulation/analytics output can depend on compiler switches, libraries, etc.

$\Rightarrow$ Idea: forum of HPC centres & 'power users' for 'FAIR ESS HPC'

Scope (1)

What do we want to achieve?

- automatically amend data produced on HPC systems with a sufficient/standardised set of metadata,

- ensure compatibility of (meta-)data with NFDI4Earth standards,

- make steps towards federated 'in place' data-analysis facilities,

- increase and simplify re-use of HPC data.

Scope (2)

… and how? With a strategy appreciating our context:

- Work with other HPC-centric groups in NFDI context (NFDI4Ing archetype 'Doris', NFDIxCS…)

- Adoption of ESS-specific and HPC-specific Research Data Management standards

… and a work plan.

Initial workplan - **Topics & Tasks**

- Focus 'Interoperability and Reproducibility'/
  Metadata for HPC (/HTC/HPDA) results: clarify...
  - what metadata must be collected, and
  - how they can be automatically collected and stored,
  - in a way acceptable for HPC centres, and
  - drive adoption.
- Focus 'Findability and Accessibility'/
  Management and data-analysis infrastructure: help devising...
  - how HPC data in NFDI4Earth can be accessed cross-site,
  - how systems for distributing/ analysing 'Big Data' sets could federate the HPC centres (also applies to remote sensing or climate data – interface to other IGs?), and, in that context,
  - how 'processing in place' can work: compute access (HPC, Cloud) at data centres via federated identities or workflow orchestration.

Mode of work

- Organisation: LRZ starts (elections etc. decided on later)

- Jour Fixe: initially every month

- First deliverables:
    - Survey of 'FAIR HPC' strategies at participating centres
    - Proposal of technical metadata-enrichment solution, taking into account relevant metadata standards
    - Federated data-analysis infrastructure: requirements and possibilities in context of German academic computing centres
    - Implementation of metadata-enrichment solution

- Lifetime of the IG: as NFDI4Earth

Looking for more participants!

We need YOU to join us ;)

- experienced HPC users
- computing centres

Current Team:

- Stephan Frickenhaus (AWI)
- Ralph Müller-Pfefferkorn (TU Dresden)
- Hannes Thiemann (DKRZ)
- Stephan Hachinger, Wolfgang Kurtz (LRZ)

Write us: fair-earth-hpc@lrz.de (preliminary, but kept open)