

### **Helmholtz Data Federation (HDF)**



Achim Streit Karlsruhe Institute of Technology (KIT), HDF Coordinator





## Helmholtz Data Federation (HDF)

- The Helmholtz Association is developing a federated research data infrastructure in Germany
  - Based on scientific use cases from communities across the German science system
  - For the benefit of and open to the full German science system
- High strategic relevance in the context "Information & Data Science" of the further development of the Helmholtz Association
- Building block for
  - A national research data infrastructure (NFDI)
  - The European Open Science Cloud (EOSC)

Disclaimer: HDF is not a federation of data per se (as the name might imply) – it is a federation of data management systems





## Helmholtz Data Federation (HDF)

- Comprising three elements:
  - 1) Innovative **software** technologies for data management
  - 2) Excellent user support and joint R&D
  - 3) Leading-edge storage and analysis hardware
- Addressing…
  - Scientific research data where ownership remains with the scientists
  - Community-specific requirements
  - Long-term preservation, curation, availability and (re-)usability of data
  - Finding, accessing, linking, analysing of multi-disciplinary data
  - Secure federation of existing data centres with uplinks to EU/intl. activities





### **Initial Use Cases**







# Strong connection to the Helmholtz <sup>(#)</sup> Program "Supercomputing & Big Data"

 Enabling scientists and engineers to solve grand challenge problems of highest complexity for science, engineering, and society by building, providing, and supporting collaborative infrastructures of scalable supercomputing and data-science facilities.





# GridKa @ KIT

- German Tier-1 High Energy Physics and Astroparticle Data & Analysis Centre at KIT
  - Supports all four LHC experiments
  - Belle II, Pierre-Auger, several small communities
- **Close interaction** with users through experiment representatives and governance
- Joint R&D with CS towards HL-LHC
- Resources
  - Computing: ~ 29k physical cores
  - Disk: 25 PB (netto), Tape: 35 PB (used)
  - 100 Gbit/s connectivity to LHCOPN, LHCONE
- Among the largest and best performing T1s
- Annual international GridKa School
  - > 1600 participants in 16 years!

Carlsruhe Institute of Technolog











## **HDF Status**

- Preliminary remark: some use cases already have long-established data federations in Germany (crossing all science organizations) and worldwide
- Establishing the federation of data management systems
  - AAI infrastructure compliant with DFN-AAI and local/regional solutions
  - High level data transfer service for large volumes of data
  - Simple data replication service
  - Federated cloud with virtual machines for data processing (moving computing to data)
  - Helmholtz Network Backbone
- Hardware investments 2017 (year 1)
  - In total ~40 PB storage, data analysis systems (CPU+GPU), cloud systems (object storage, server infrastructure), network equipment



#### **Research Data Management (RDM) @ KIT**





#### **RDM @ KIT – Servicelandkarte**



