DESY Data Management Photon Science and High-Energy Physics

Mikhail Karnevskiy for Central IT
Hamburg, DESY

HELMHOLTZ

DESY.

HEP

Peculiarity: Large scale experiments, which are driven by huge collaborations (ATLAS, CMS, Belle).

Step	Tool	Community
Detector, Data-taking	Collaboration tools	Collaboration, Local infrastructure support
Start-to-End Simulation	HEP community tools: GEANT4, Pythia, DD4HEP, key4hep	HEP community
Online processing and online data-reduction	Complex high-performance trigger system to process multiple TB/sec	Collaboration
Data storage (DESY part)	dCache, NAF, NSF storage	IT
Offline analysis	CERN root, Experiment specific frameworks (e.g. CMSSW, Athena, ...), running on DESY Grid & NAF clusters	Experiment experts, incl. local experiment members, IT

HEP

Step	Tool	Community
FAIR data handling, publication and archiving	experiment specific frameworks, archiving: CTA CERN OpenData	Experiment experts, incl. local experiment experts, IT
Step-overarching: Metadata handling & ELN	Collaboration tools, HERA data preservation	Collaboration, Community, local experiment experts
Used data formats, Data flow & automatisation	common WLCG low level tools (incl. eg. Grid, RUCIO), experiment specific additions, some site policy and tools (dCache/CTA)	CERN, Experiment experts, incl. local experiment experts, IT

Photon Science

Step	Tool	Community
Proposal Management	https://door.desy.de/door/	IT & FS-EC
Data-taking / detector	Tango (device integration) & Sardana (control software) Potentially future control software: bluesky , BLISS, ? data ingest: ASAPO + NFS + SMB + ZeroMQ (HiDRA) file formats: HDF5, TIFF, CBF, ... Classical mode 'first store, then process data'	FS computing: EC/SC Data transport: Central IT (ASAPO)
Start-To-End-Simulation	Driven by the science community Different simulation tools for different steps, depending on purpose and use cases	Content: Science community / FS-EC/SC Infrastructure: Central IT

Photon Science

Step	Tool	Community
Online processing and online data reduction	Driven by user community. H5Tools for data compression Maxwell HPC infrastructure operated by central IT ASAP3 allows (for lower rates and latencies) automatic processing in the classical mode - first save then analyse. ASAPO for online analysis (worker concept used with e.g. PyFAI or XDS) Moving towards real time / on the fly processing	Science community plus FS computing groups: EC/SC Infrastructure: Central IT

Photon Science

Step	Tool	Community
Data storage	GPFS, dCache using various access protocols (all POSIX - close to) i.e. GPFS-native, NFS(3,4,pNFS), SMB, dCache-native. Custom workflows using ASAP3	Concepts: Central IT & FS-EC Infrastructure: Central IT Catalogues: Central IT Data formats: Wider user community
Offline data analysis	Software: Community specific analysis software (CrystFEL, PtyPy, XDS...) Python: Numpy, Scipy Infrastructure: Maxwell HPC (SLURM, Jupyter, SciCat),	Wider photon science community plus FS computing groups: EC/SC Infrastructure: Central IT

Photon Science

Step	Tool	Community
FAIR data handling, publication and archiving	Metadata: SciCat in discussion / n.a. + specific tools (details to be defined) Archiving: CTA Publication: DESY PubDB	Experiment experts, incl. local experiment experts, IT
Step-overarching: Metadata handling & ELN	Data management & remote access: Gamma Portal SciCat / community specific tools (details to be defined)	Mostly Central IT Needs driven from user community Integration of ELNs with instrumentation: Photon science EC/SC
Used data formats, Data flow & automatisation	Data formats: HDF5, Nexus, ... Data flow & automatisation: ASAPO	User community Infrastructure and data flow: central IT

DM: tools + infrastructure + community + service

- Community: Central IT, FS-ES, FS-SC
- Infrastructure: NAF HPC, Maxwell HPC, GRID, dCache, ...
- Tools: ASAPO, ASAP3, ...