

High Data Rate Initiative in the Helmholtz Association (HDRI)

Status Report on Workpackage 1 (Data Management)

R. Gehrke, DESY, Hamburg



High Data Rate Processing and Analysis Initiative (HDRI)

Helmholtz PNI Centres

DESY Hamburg	FZ Jülich
FZ Karlsruhe	HZG Geesthacht (former GKSS)
GSI Darmstadt	HZB Berlin

Work Packages

WP1: Data Management (DESY, HZB)

- Standardisation and Data Formats
- Data Access Strategies
- Data Lifetime Management and Archiving

WP2: Real-time Data Processing (GSI, KIT)

- Real-Time Data Assessment with Parallel Computing
- Analysis Methods and Applications
- Data Processing with Dedicated Hardware

WP3: Data Analysis, Modeling, and Simulation (FZJ)

Close co-operation with PanData



Standard Data Format

Commitment: NeXus/HDF5

1. Review of the NeXus classes (preparation)

2 .New NeXus API C++ strictly object oriented
 structural improvements
 no performance degradation compared to HDF
see talk E. Wintersberger first stable version fixed
and H. Pasic documentation under preparation
 Python still some issues to solve

3. Data Collector Software as link between Instruments and NeXus files

see talk J.-U. Hoffmann

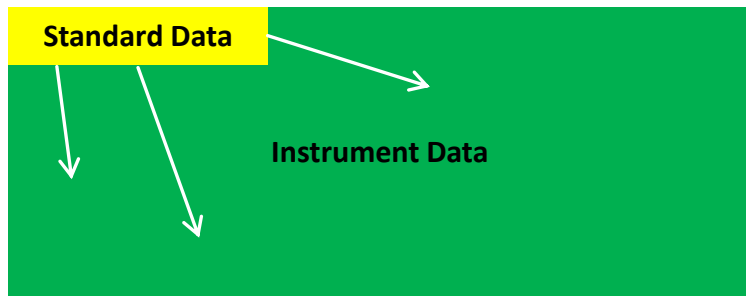
4. Application Definitions to provide method specific view on the data
 Scattering, Tomography, Macromol. Crystallography,
 X-ray Absorption Spectroscopy, X-ray Fluorescence

See talk R. Gehrke

Next Steps: Implementation at specific instruments

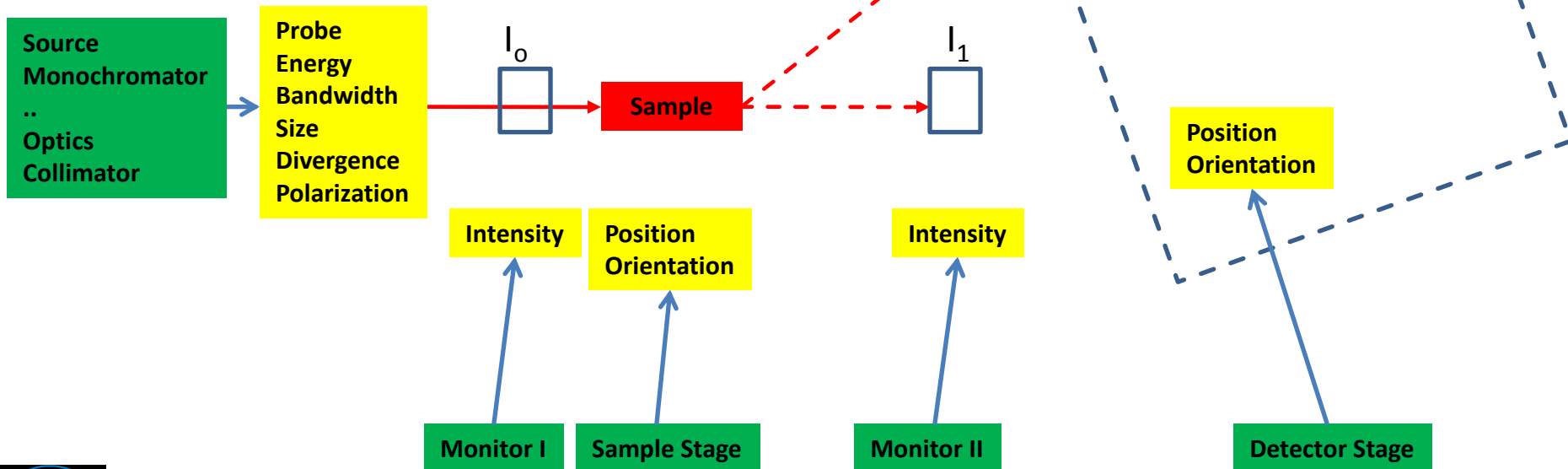


Standard Data Set for a Generic Scattering Experiment



Application Definition

Provide a unique and complete view on the stored data that allows to transform the measured intensities to reciprocal space



Standard Data Format (accompanying measures)

Common Data Model (CDM) of SOLEIL/ANSTO

Agreement DESY/SOLEIL on co-operation

Codecamp at DESY in spring 2012

CDM is presently based on Java plugins

C++ based implementation (SOLEIL)

Python based implementation (DESY)

Serves another method specific interface to the data

Negotiations with Detector Vendors

Dectris is up to adopt the HDRI Standard Data Format

Couple of NeXus extensions in close collaboration with NeXus Group

Might have signaling effect for other suppliers

Negotiations with HDF5-Group

Modification and extension of HDF5 (in co-operation with PSI)



Common Data Policy

How shall data stored at the centers be treated

- Commitment to provided infrastructure and services for storage and access
- Rules for the data lifecycle
- Rules for access to the data

Discussion on basis of PanData (ODI) Draft

Can all HDRI partners agree on this?

What are concerns and how to resolve them?

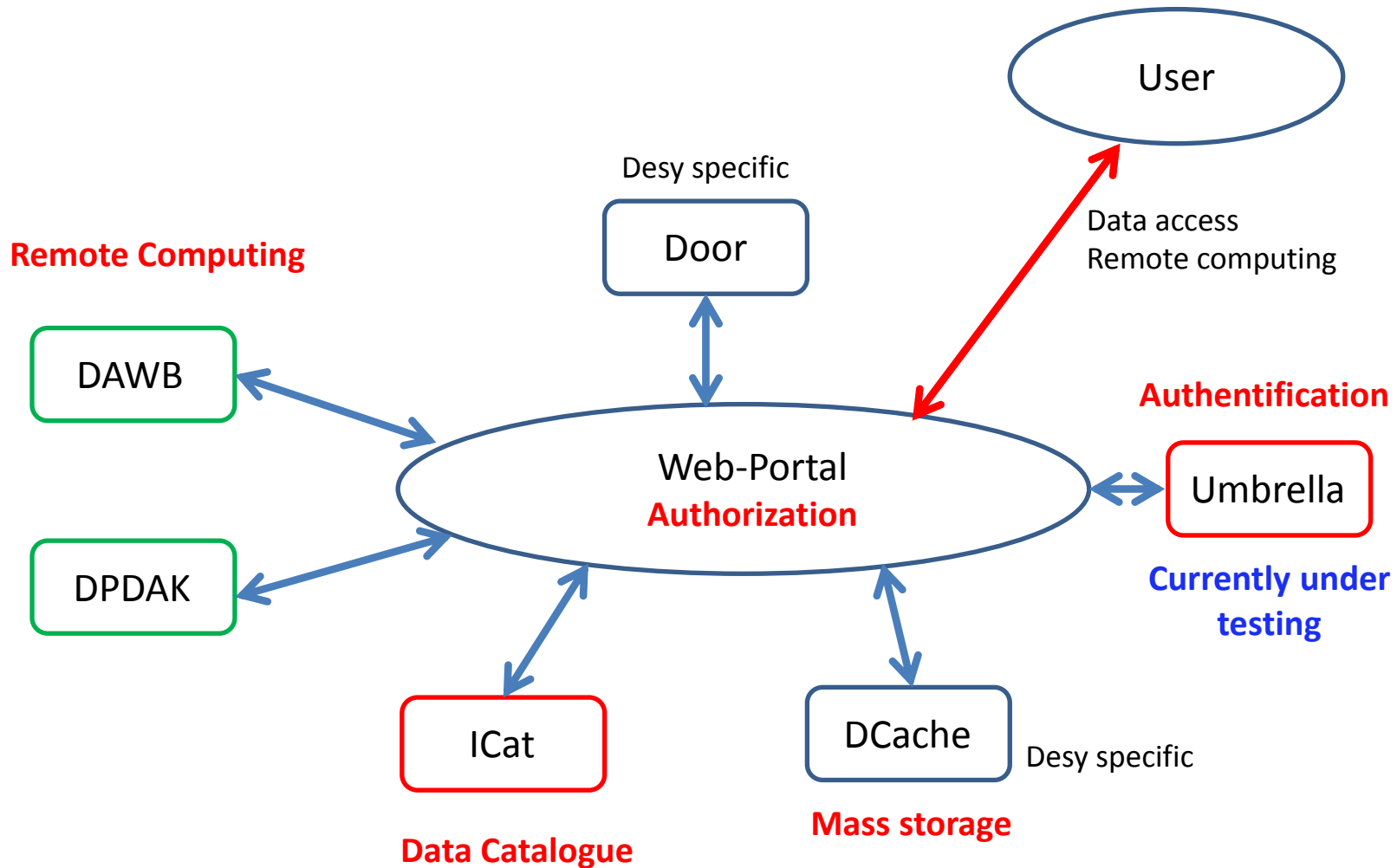
Discussion session during this Workshop

→ Official approval in 2012



Remote Data Access

In close co-operation with PanData



Workshop Agenda (First Day)



10:00 - 12:00 HDRI

- 10:00 Welcome
- 10:15 HDRI WP1 Status Report
- 10:30 HDRI WP2 Status Report
- 10:45 HDRI WP3 Status Report
- 11:00 Discussion Data Policy

- Edgar Weckert (DESY)
- Gehrke (DESY)
- Andreas Kopmann (KIT)
- Joachim Wuttke (JCNS)
- Rainer Gehrke (DESY) [Chair]

12:00 - 13:00 Lunch

13:00 - 15:20 HDRI + PaNdata

- 13:00 Dedicated Hardware in HDRI WP2
- 13:25 NeXus API
- 14:05 NeXus Application Definitions
- 14:30 NeXus at DLS
- 14:55 Data Collection Software

- Matthias Balzer (KIT)
- Eugen Wintersberger (DESY), Halil Pasic (KIT)
- Rainer Gehrke (DESY)
- Tobias Richter
- Jens-Uwe Hoffmann (HZB)

15:20 - 15:40 Coffee break

15:40 - 18:30 HDRI+PaNdata

- 15:40 Directly Programmable Data Analysis Kit
- 16:05 Data Analysis Workbench
- 16:30 The Scientific Data Analysis (SDA) Project at DLS
- 16:55 Practical Experiences with GPUs for High-Throughput Computing
- 17:20 Provision of Low level Services to Write pHDF and Planned Implementation at DLS
- 17:45 Update of Software to Exploit Attached Processors and Clusters for Tomography

- Gunthard Benecke (DESY)
- Andy Goetz (ESRF)
- Mark Basham (Diamond Light Source)
- Suren Chilingaryan (KIT)
- Ulrik Pedersen (Diamond Light Source)
- Mark Basham (Diamond Light Source)

18:10 Discussion

19:00 Dinner

