High Data Rate Processing and Analysis Initiative (HDRI)

Helmholtz PNI Centres

DESY Hamburg	FZ Jülich
FZ Karlsruhe	HZG Geesthacht
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 ODI

WP1: Data Management

Data Format Standardisation

Design & Definition

NeXus / HDF5

Application definitions:

- **Scattering**
- **Tomography**
- **Macromolecular Crystallography**
- **X-Ray Absorption Spectroscopy**
- **X-Ray Microfluorescence**

Unique method-specific view on experiment data

Implementation

- NeXus API for high data rates (C++ and Python)
- Data Collector Software (flexible link between NeXus and the data sources)

Misc.

- Dialog with detector vendors (e.g. Dectris)
- Connection with CDM (Common Data Model, SOLEIL/ANSTO)

Remote Data Access

- Authentification
- Archiving
- Common web based data management portal
- Authorisation
- Remote computing (&Software repository)

Data Policy

Strong overlap with PanData activities

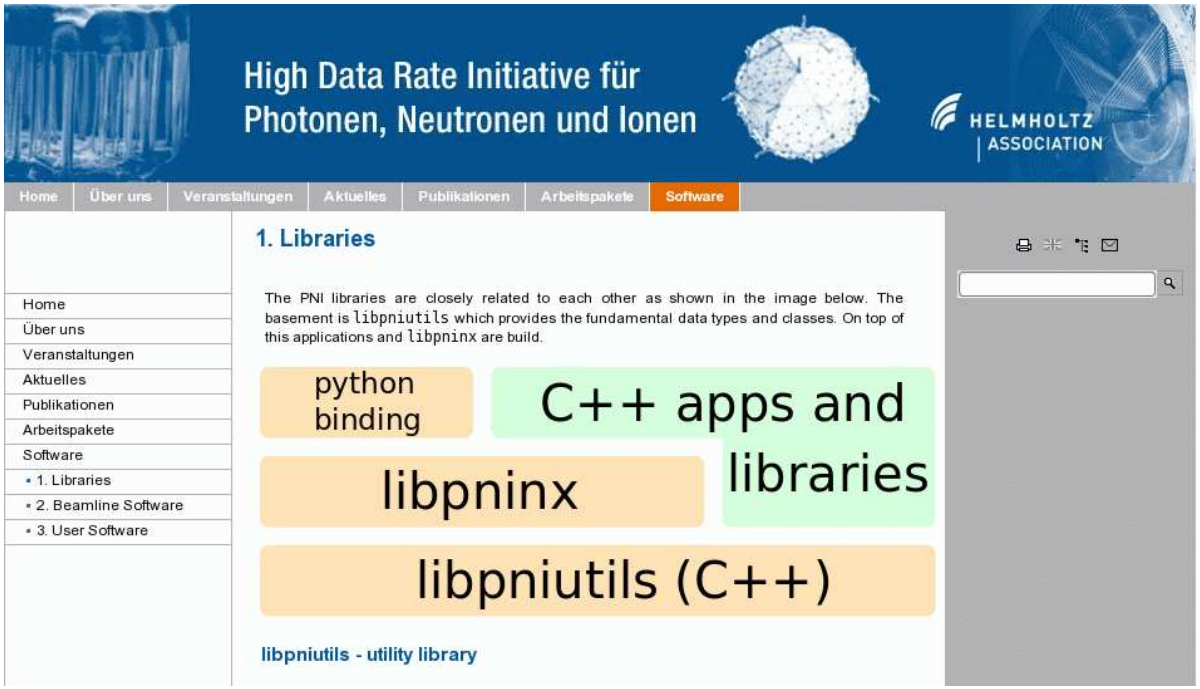
NeXus implementation – Step 1 – NeXus API

E. Wintersberger

- A new NeXus C++ library, exclusively based on HDF5
- Purely object-oriented interface
- Python bindings
- External filters support

- Coordinated with NIAC
- Code on “Google Code”

Status:
All Milestones reached



CDMA: Common data model architecture, SOLEIL, DESY, ANSTO

- Abstraction layer between files and applications
- DESY implemented Python bindings

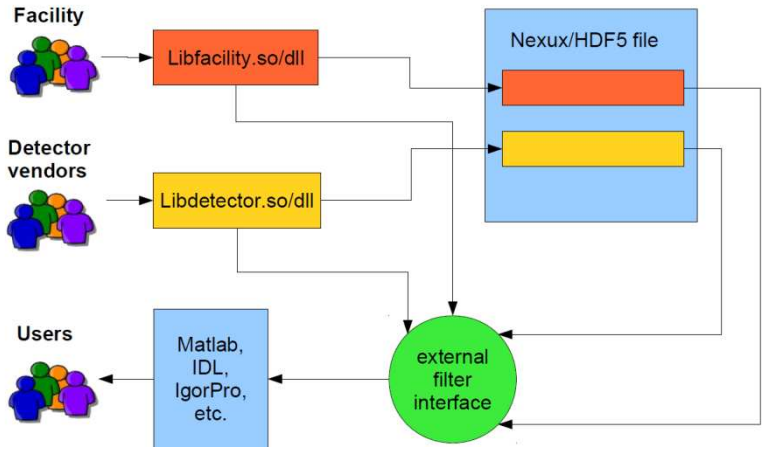
NeXus - Accompanying Measures

HDF5-Group

Framework for precompressed Data → funded by DECTRIS/PSI

External Filter Interface → funded by DESY → to be finished May 2013

- Eiger Detectors using custom compression
- New version will be used in autumn release of MATLAB



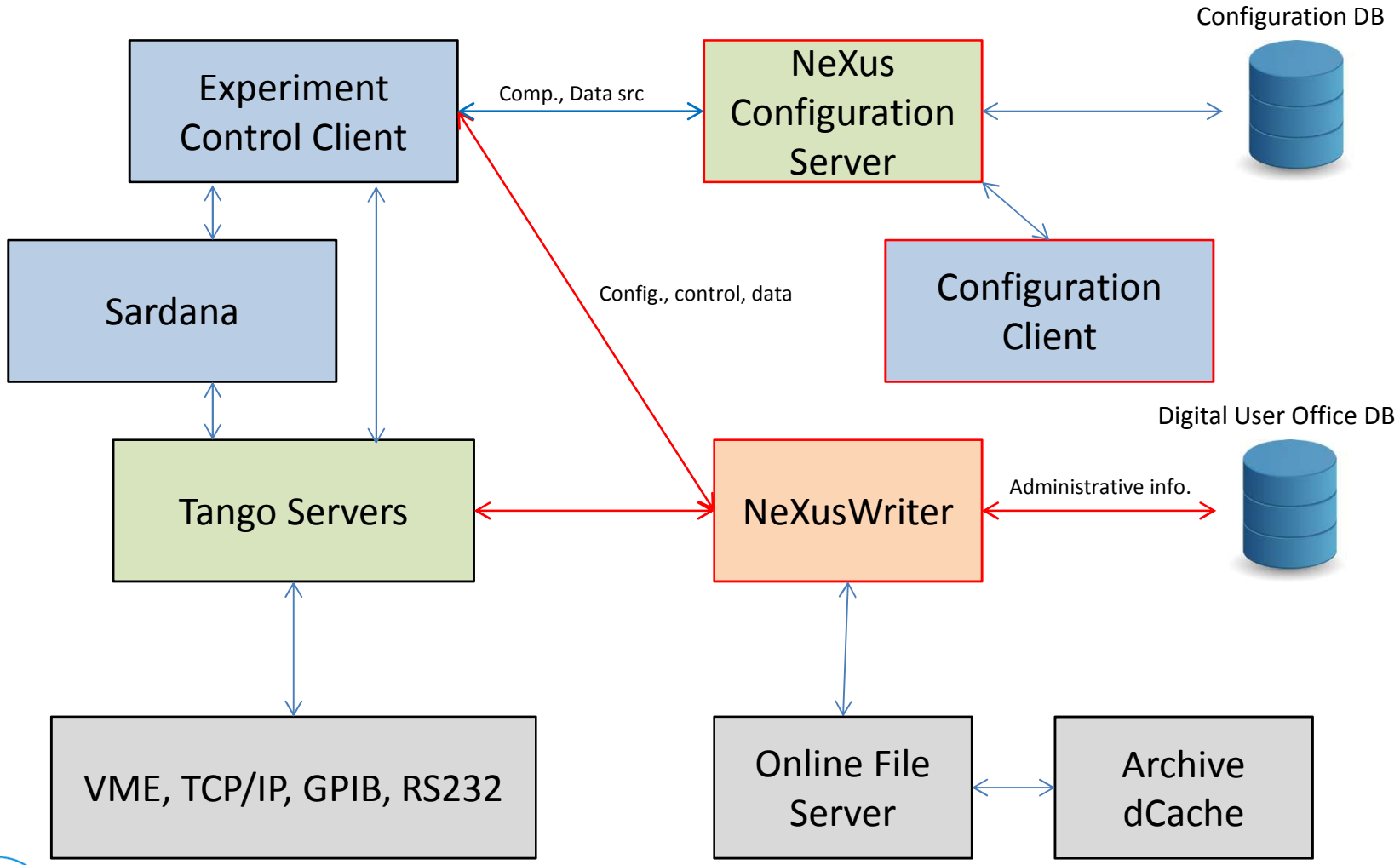
DECTRIS

Open API for new Pilatus and Eiger Detectors (allow to intercept the data stream)

NeXus Implementation – Step 2 – NeXus Data Server

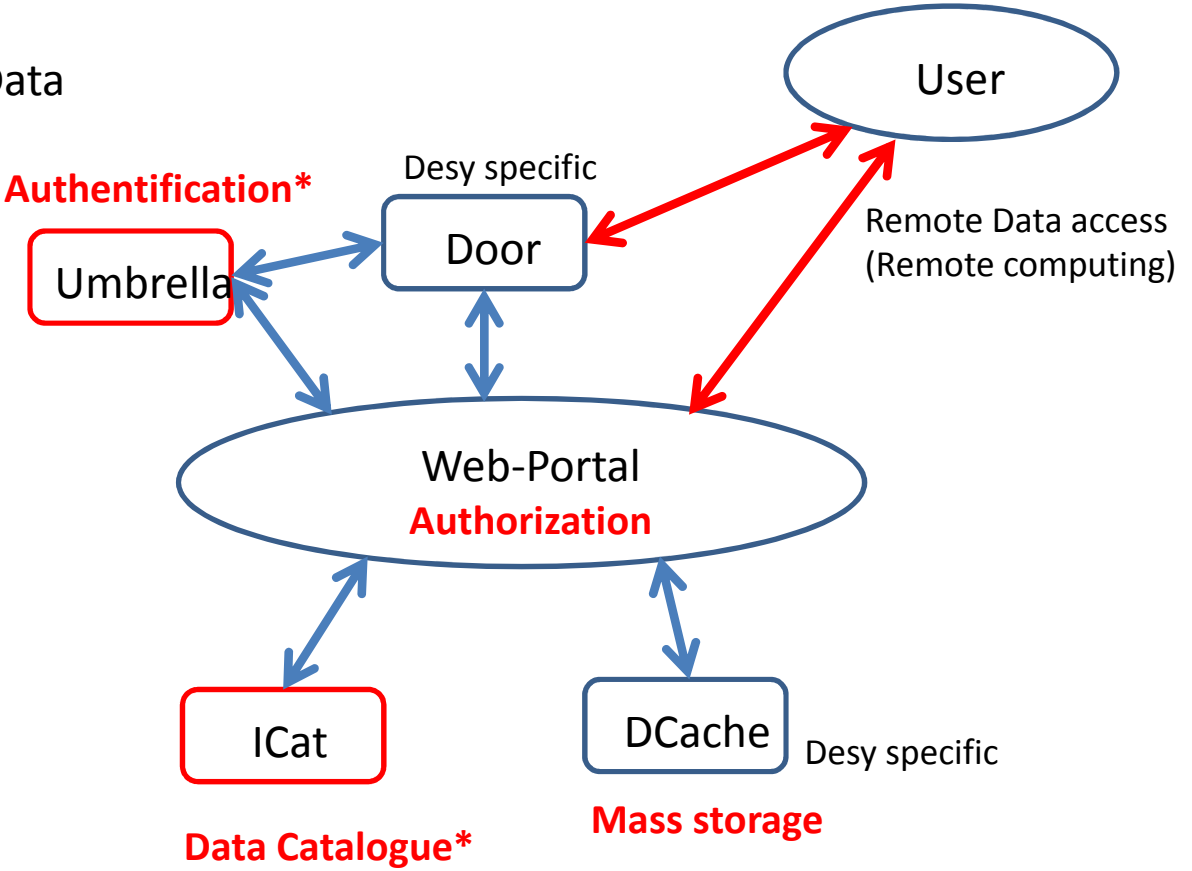
Status: All Milestones reached

J. Kotanski



Remote Access: Web Portal

*In co-operation with PanData



Status: In Progress

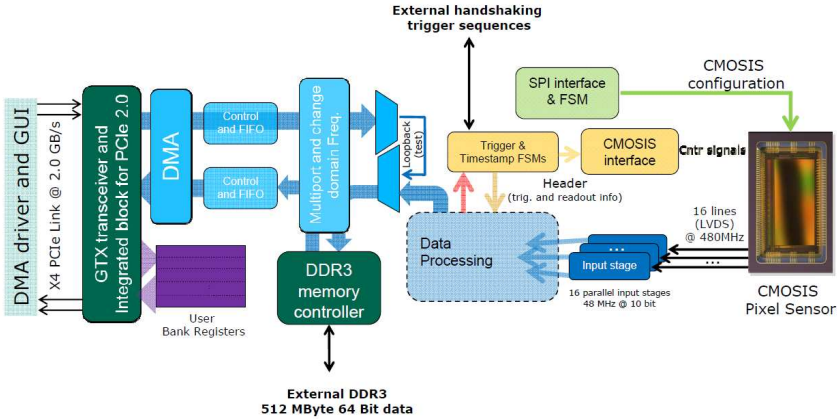
Web-Portal	Web Access	γ-Portal under construction	DESY
Umbrella	Authentification	Testphase finished	PanData
ICAT	Metadata Catalogue	implemented (compilation of meta data in progress)	PanData
dCache	Data Archive	NFS 4.1 consolidated	DESY

WP2: Real-time Data Processing (FPGA)

FPGA based Hardware for data acquisition and analysis

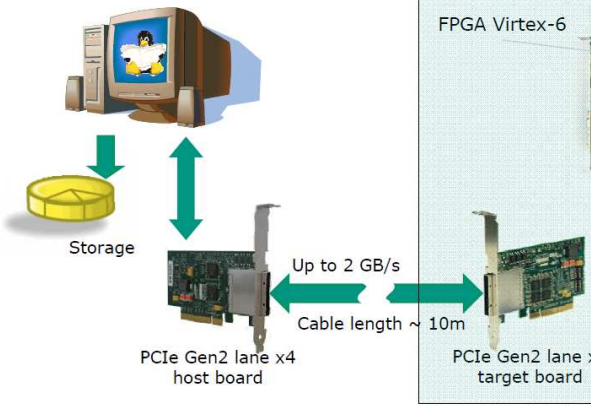
Case 1: High speed camera demonstrator for SR tomography applications (KIT)

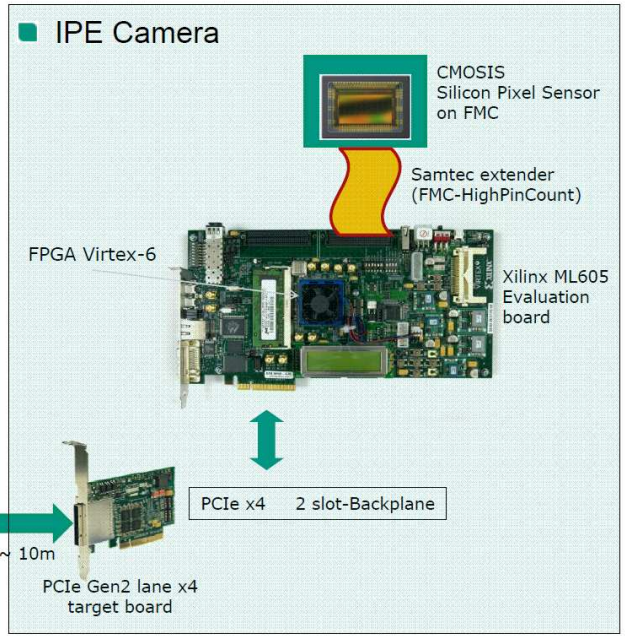
(2.2 Mpixel, 12 Bit, 340 frames/s, preprocessing, DMA transfer)



FPGA Function Blocks

Reconstruction server & online monitor





Case 2: Si(Li)-Compton Polarimeter Readout (GSI/SPARC)

(Replace NIM/VME by custom designed FPGA hardware)

Status: All Milestones reached

WP2: Real-time Data Processing (GPU)

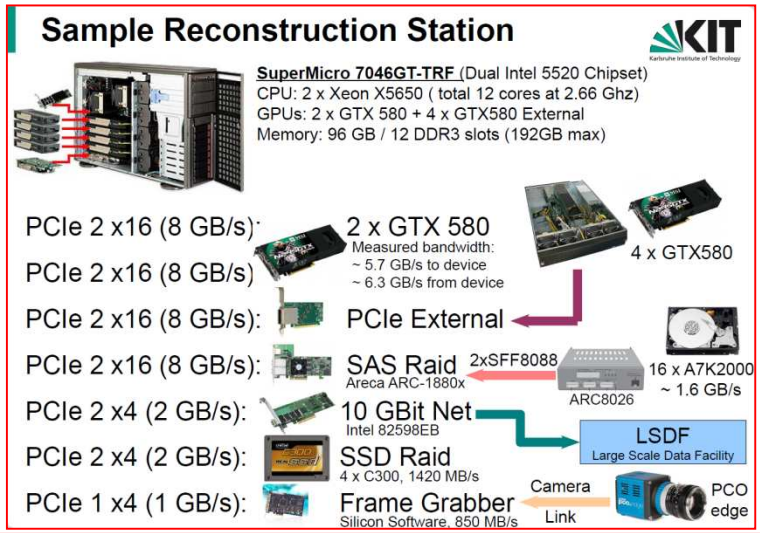
GPU based data analysis (system layout, benchmarking, demonstrators)

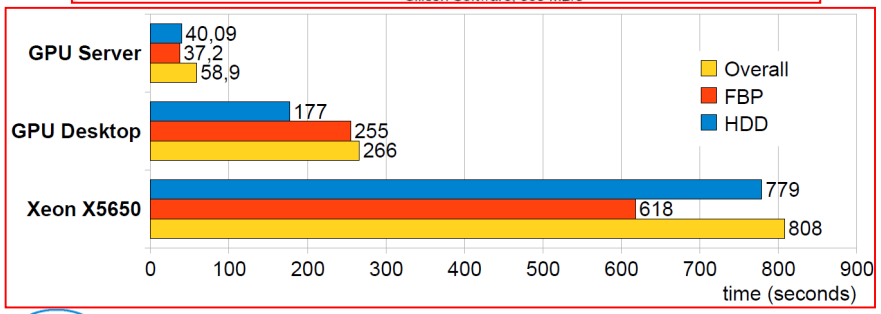
Case 1: Fast reconstruction of tomography data (filtered back projection) (KIT)

(demonstrator with 12 CPU cores and 6 Nvidia GTX580 Fermi-GPUs

→ 24 Gbyte input data to 11 Gbyte output in some ten seconds)

Case 2: to be defined (MX, SAXS)





Sample Data-Set

Porose polyethylene grains in a conical plastic holder

KIT

Sample Data Set:

Number of Projections:	2000
Resolution of Projections:	1776x1707
Total Size:	24GB

Resulting 3D Image:

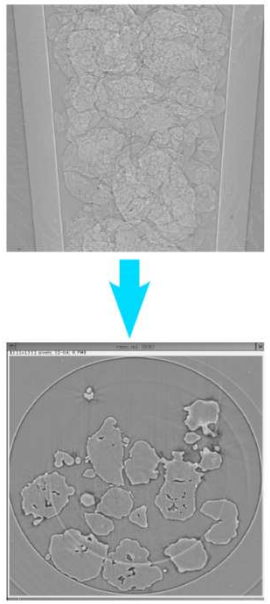
Resolution:	1691x1331x1311
Total Size:	11GB

Required Operations:

Data I/O:	35GB
Filtering:	~ 0.6 TFlop
Back Projection:	~ 53.0 TFlop

Maximum Performance Estimation:

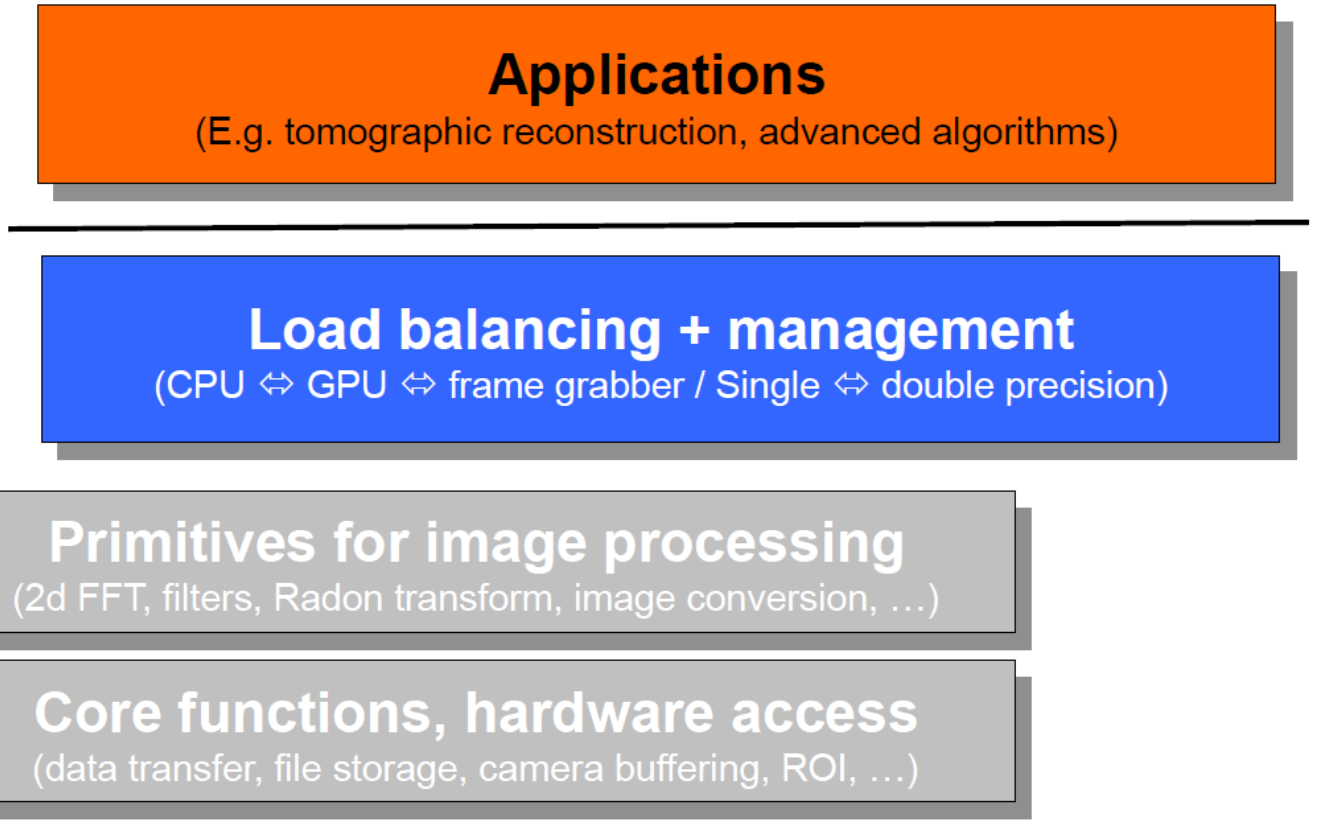
Opteron 6176 (~ 110Gflops)	~ 10 minutes
HDD I/O (~ 50 MB/s)	~ 15 minutes



Status: All Milestones reached

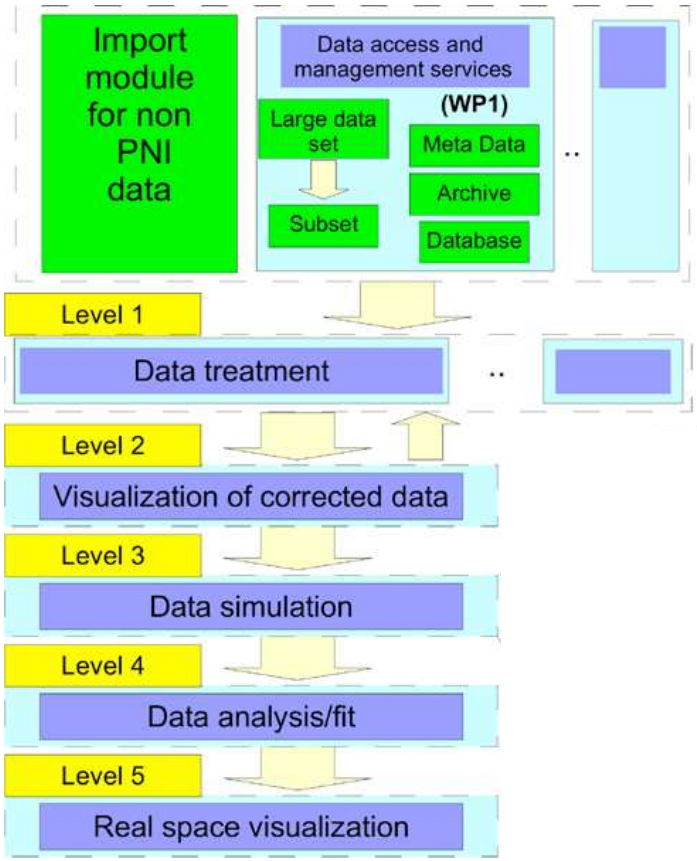
Parallel Computing Framework (UFO)

Status: In Progress



WP3: Data Analysis, Modeling, and Simulation

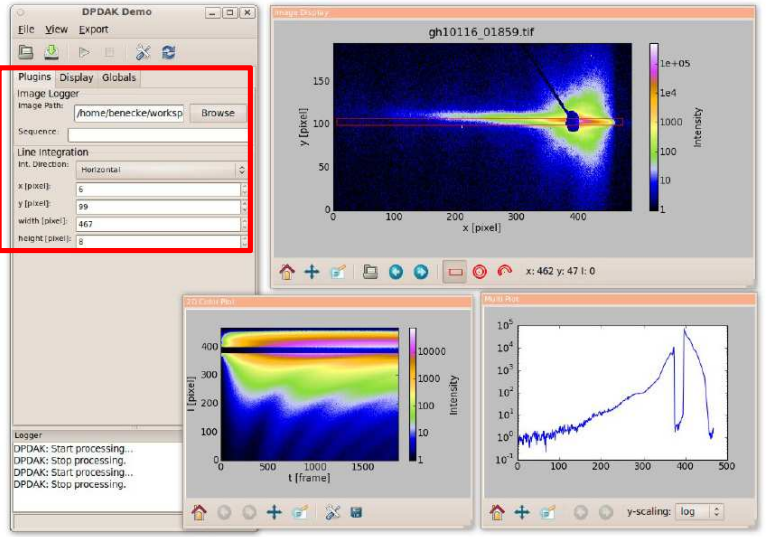
Status: In Progress



Current activity at DESY (co-operation with MPI Golm)
Prototype: Directly Programmable Data Analysis Kit (DPDAK)

Max-Planck-Institut für Kolloid- und Grenzflächenforschung

Plugin controls



Python based Framework and Plugins,
GUI for data flow setup
At present: Implementation of part of the GISAXS data evaluation workflows

Remaining HDRI activities in POF-2

Advancing the UFO parallel Computing Framework

- Features: automatic Scheduling and Optimization, GUI
- Content: New algebraic reconstruction algorithm for tomography data

Using UFO for application with streamed data

- Integration with DAQ hardware platform

Harmonization of Web Portals…

NeXus (Standard Data Format) Implementation at DESY, KIT, HZG, HZB

Data Server & Configuration Server, DESY Petra III beamlines P02, P03

ANKA/KIT image beamline - Tomography

NeXus Tools

MPI support for NeXus API

Fuse Filesystem for NeXus Files

Adding new HDF5 features to NeXus API (external filters…)

Benchmark tools for NeXus API

Dissemination of NeXus at the Helmholtz Centers

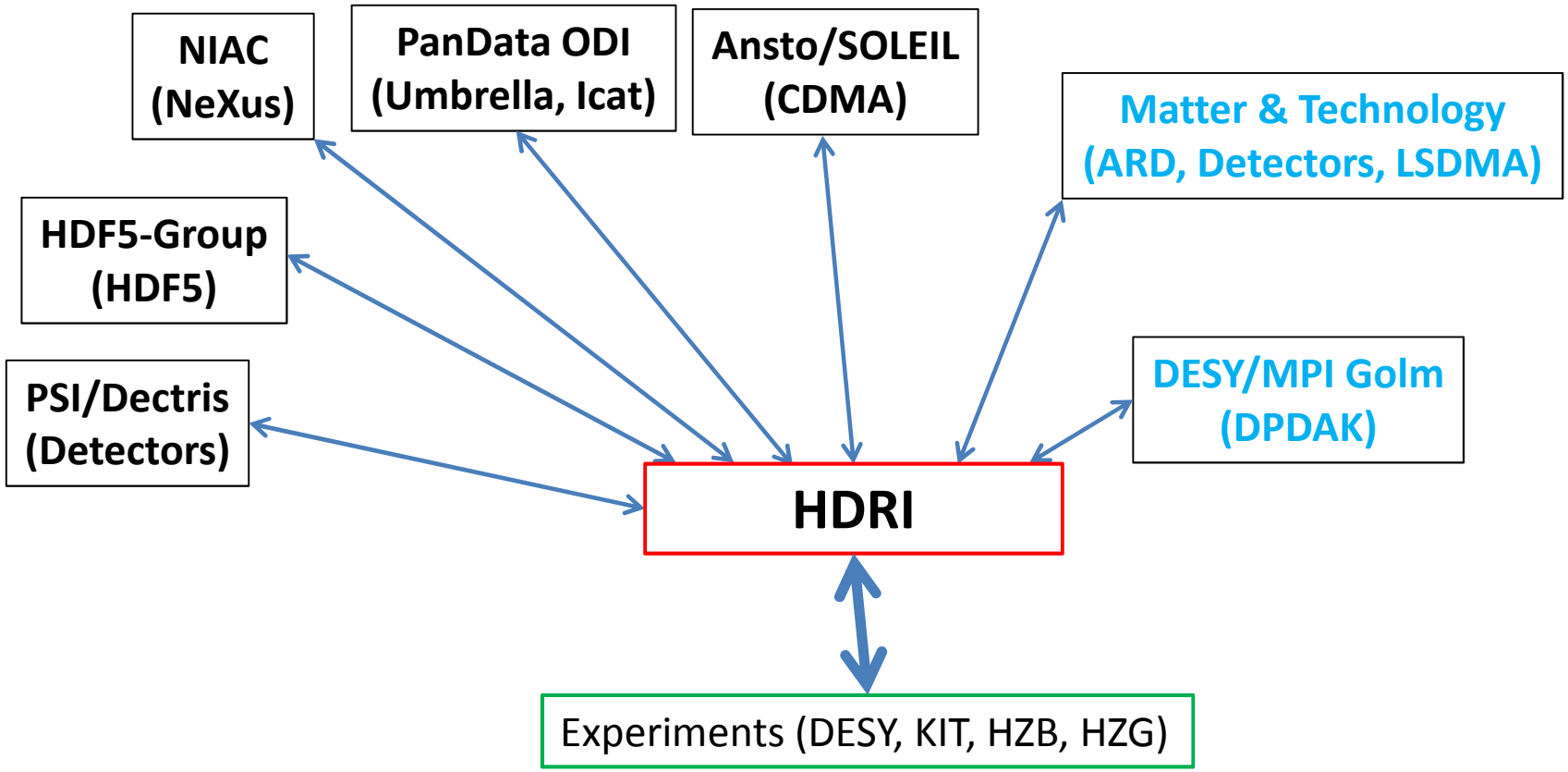
HDRI Activities in POF-3 (preliminary topics)

Quality Assurance for Scientific Software
Concurrent Computing at beamlines with high data rate detectors
(Multithreading, MPI, 0MQ) (realtime reduction and visualization)
NeXus Application Definitions
Training of Users

Workbenches and Workflow Management (DPDAK, DAWN)
Web-Portal assisted remote computing

Scalable computing solutions
(Clusters of GPU nodes)
Common infrastructure for beamlines (DAQ) and Computing (Analysis)

Role of HDRI



Implementation of data related issues at the Helmholtz Centers‘ Experiments on the basis of own and third party developments