High Data Rate Initiative of the Helmholtz Association (HDRI)

Status Report on Workpackage 1 (Data Management)

R. Gehrke, DESY, Hamburg

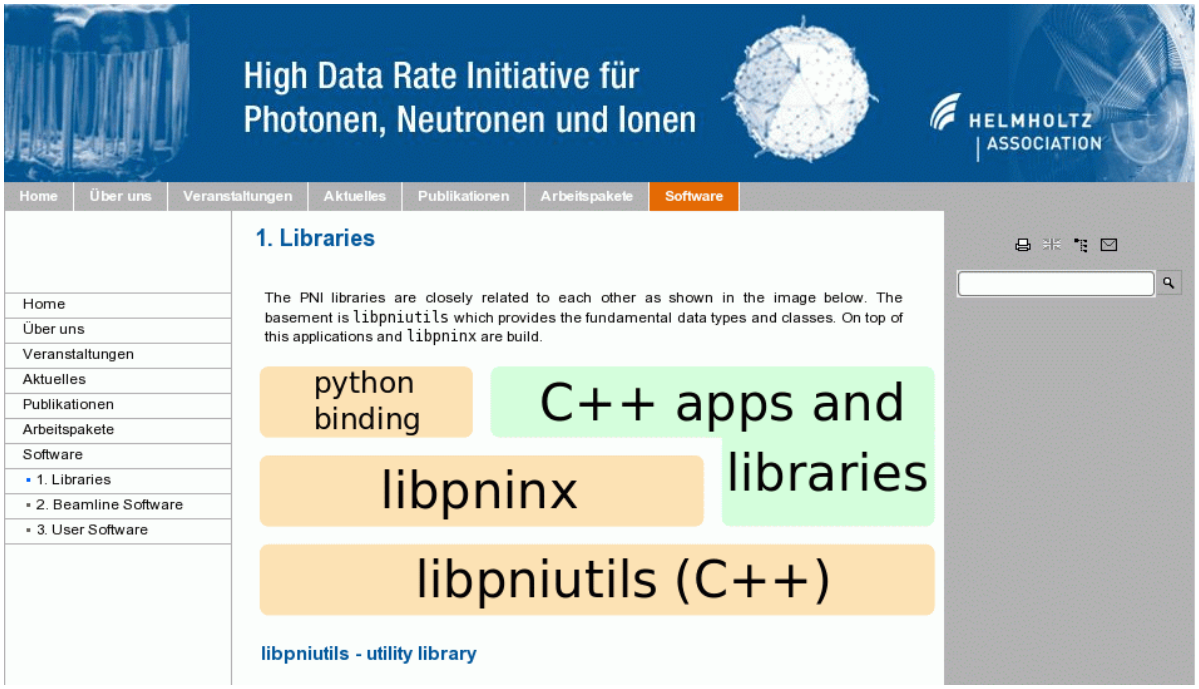
1. Standard Data Format
2. Webportal
3. Data Policy

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”



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

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

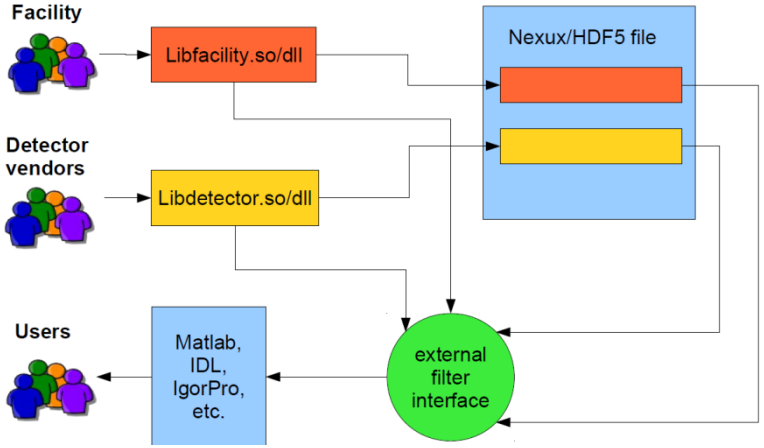
NeXus - Accompanying Measures

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)

NIAC

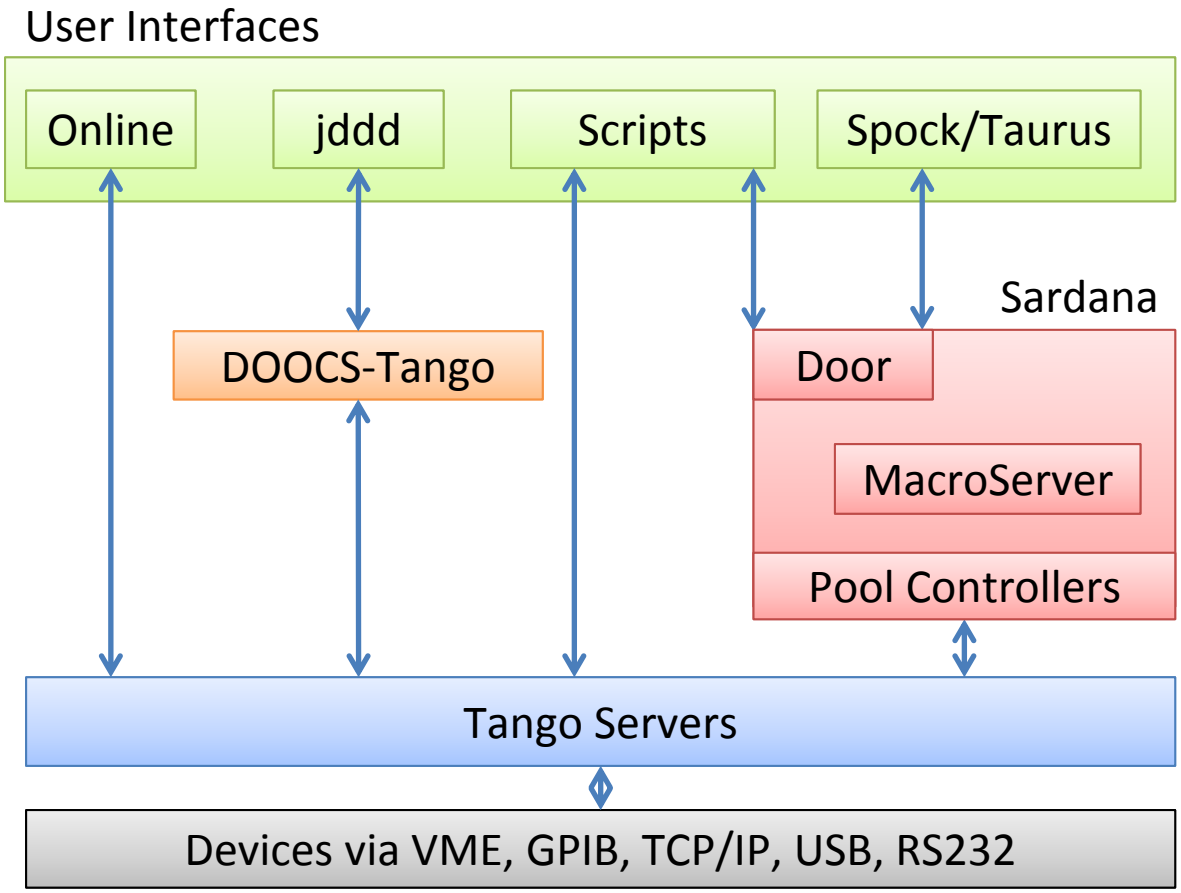
IUCR Proposal to make NeXus an IUCR file standard

NeXus will adopt IUCR standards for describing geometries

CBF library will read NeXus files

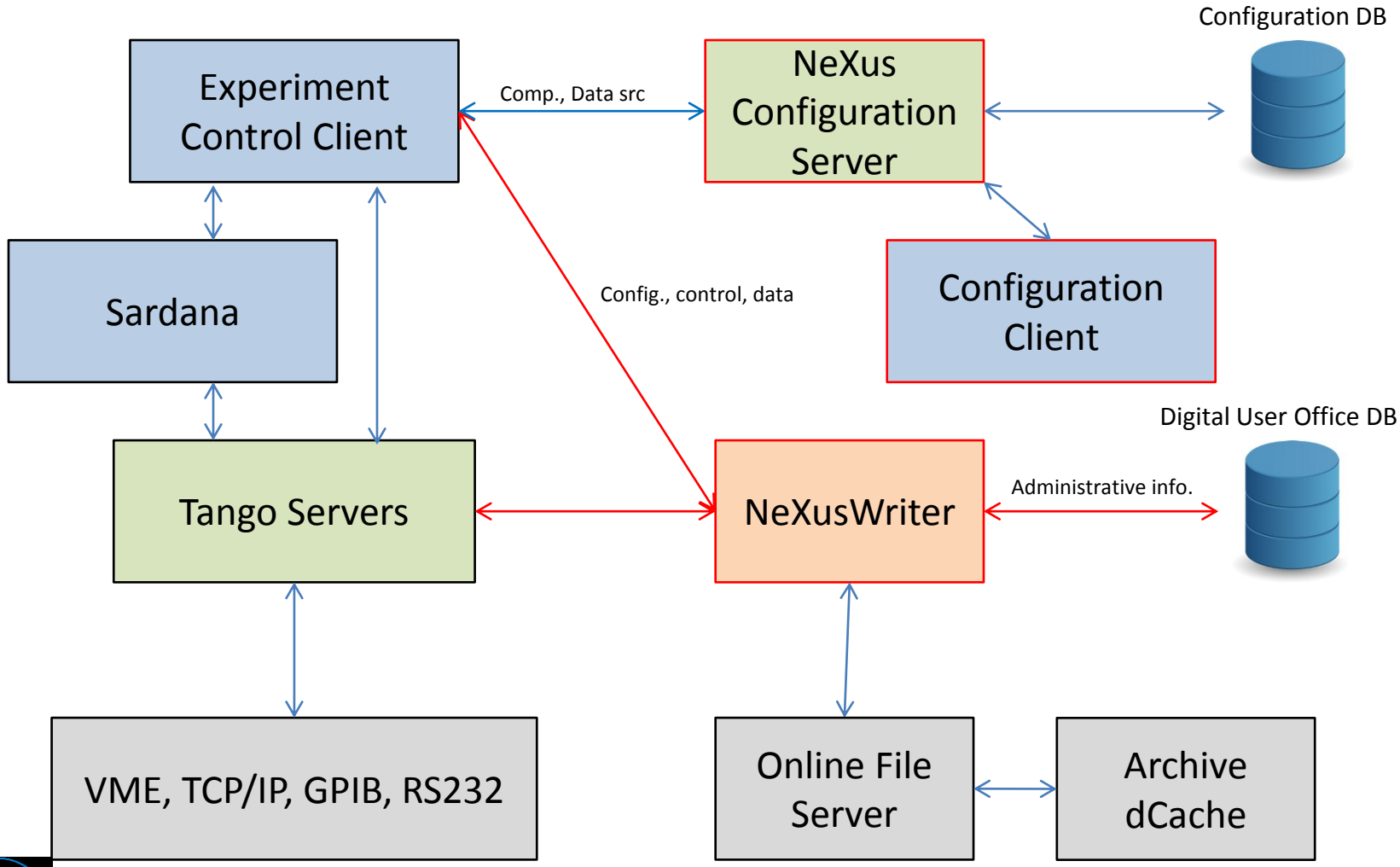
The Experiment Control System at PETRA III

T. Kracht, T. Núñez, H. Blume, M. Fleck, J. Grabitz, Y. Yu



NeXus Implementation – Step 2 – NeXus Data Server

J. Kotanski



NeXus Configuration

J. Kotanski, E. Wintersberger

The NeXusWriter uses a **configuration** to create a NeXus file structure.

Configuration: a collection of **components**

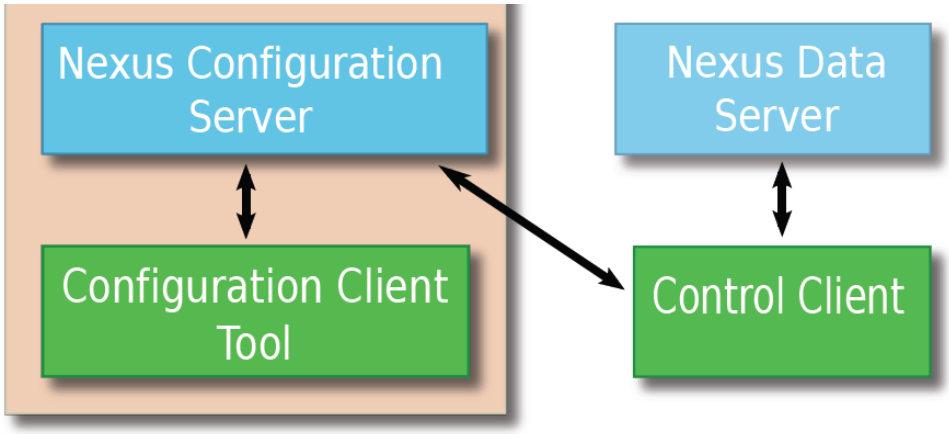
Components…

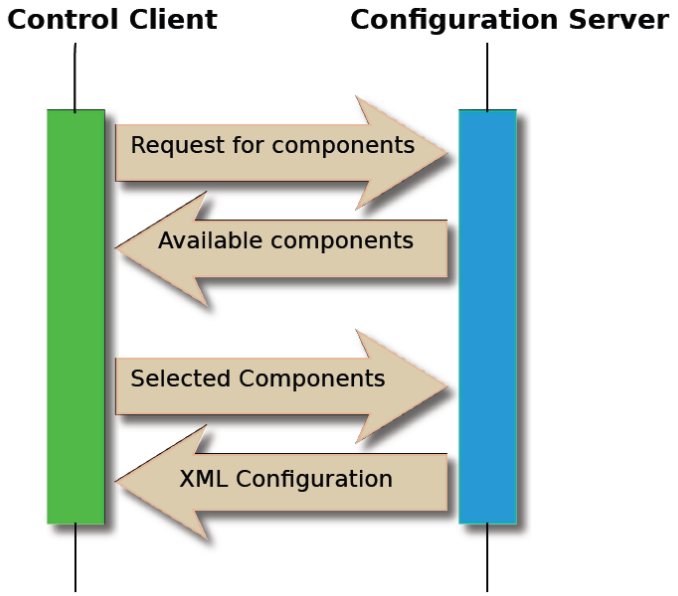
- …Represent
  - Primitive devices: counter, MCAs, 2D detectors
  - Composite devices: monochromator, diffractometer
  - Structures: beamline description
- …Can be inserted and removed without side effects
- …Contain several NeXus classes (elements)
  - The elements are specified using absolute paths, which allows them to be merged

- The elements have data sources or links
- Application definitions can be expressed in terms of components

- Data sources
  - Tango servers, json strings from the control client, DB, factory for external inputs

Control Client – Configuration Server: Basic Workflow

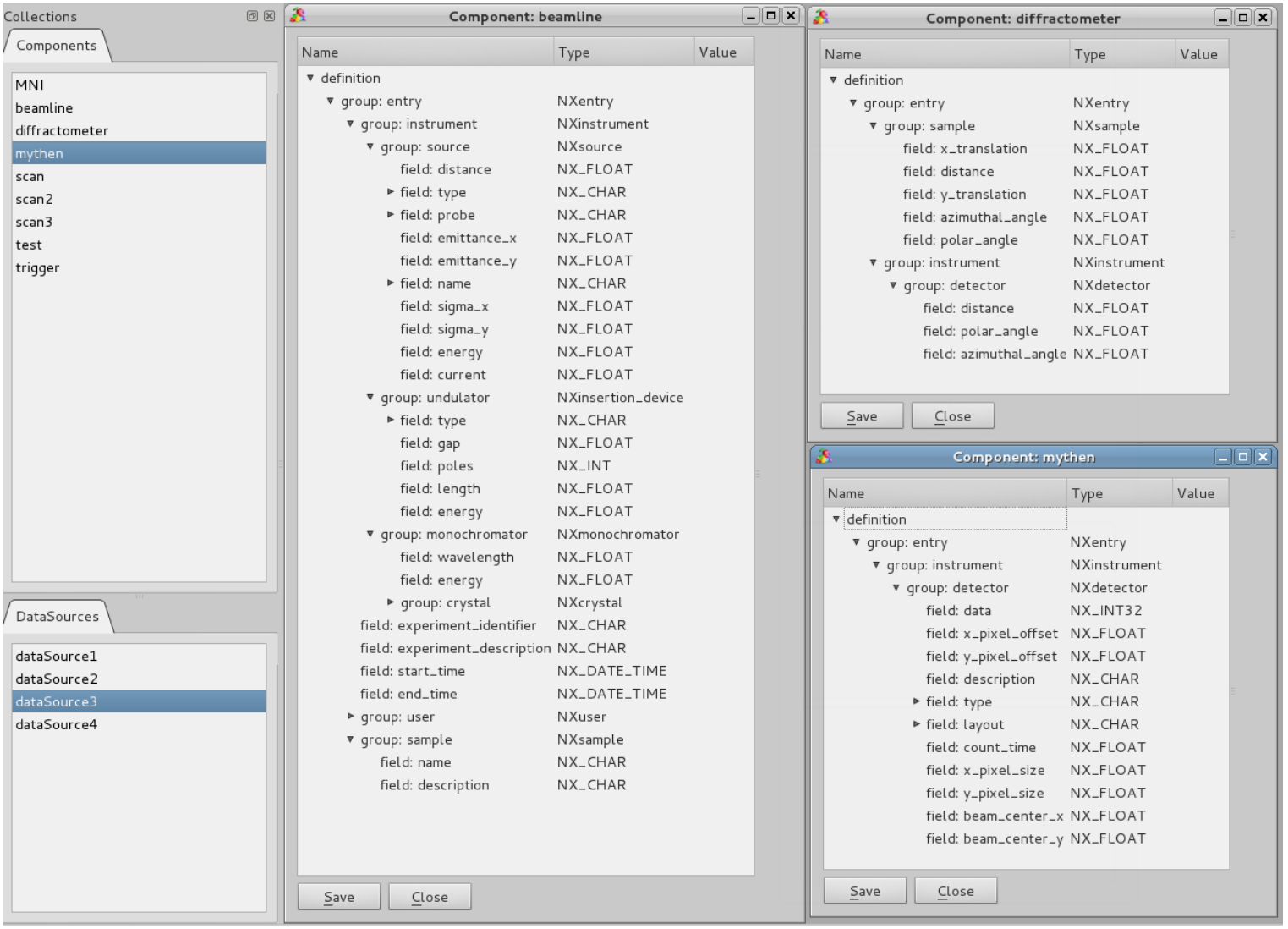




Component Designer

Configuration Client Tool

J. Kotanski



NeXus Integration – Status Summary

✔ C++ and Python APIs are stable and ready for production code

✔ Basic Matlab, IgorPro, IDL support for HDF5

✔ Nexus data server

✔ Nexus data server configuration tool

✔ Nexus configuration server

- Nexus / CBF / CIF integration
- External filter support
- Full documentation of beamline software (servers)

✗ Multithreading support

✗ MPI support

✗ Application definition integration

NeXus Implementation – Step 3 – Beamline Integration

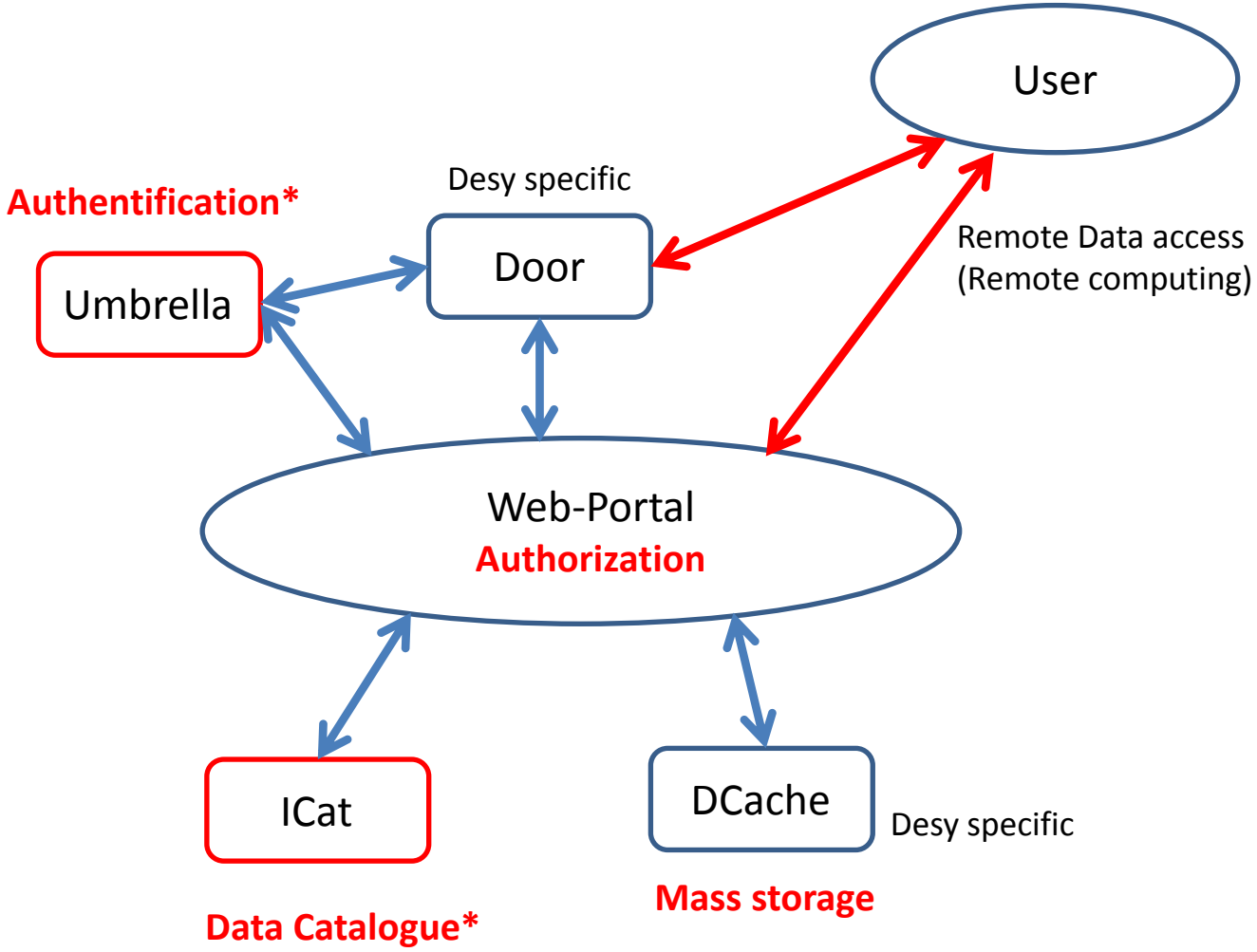
Integration into beamline control (showcase P03, microSAXS/WAXS)

Application definition (Generic Scattering Experiment)

- Content Definition is available (Guideline for implementation)
- General formal NeXus compliant description is very difficult
- „pragmatic approach“: First – Practical implementation according to guidelines
  Second – Formal description

Remote Access: Web Portal

*In co-operation with PanData



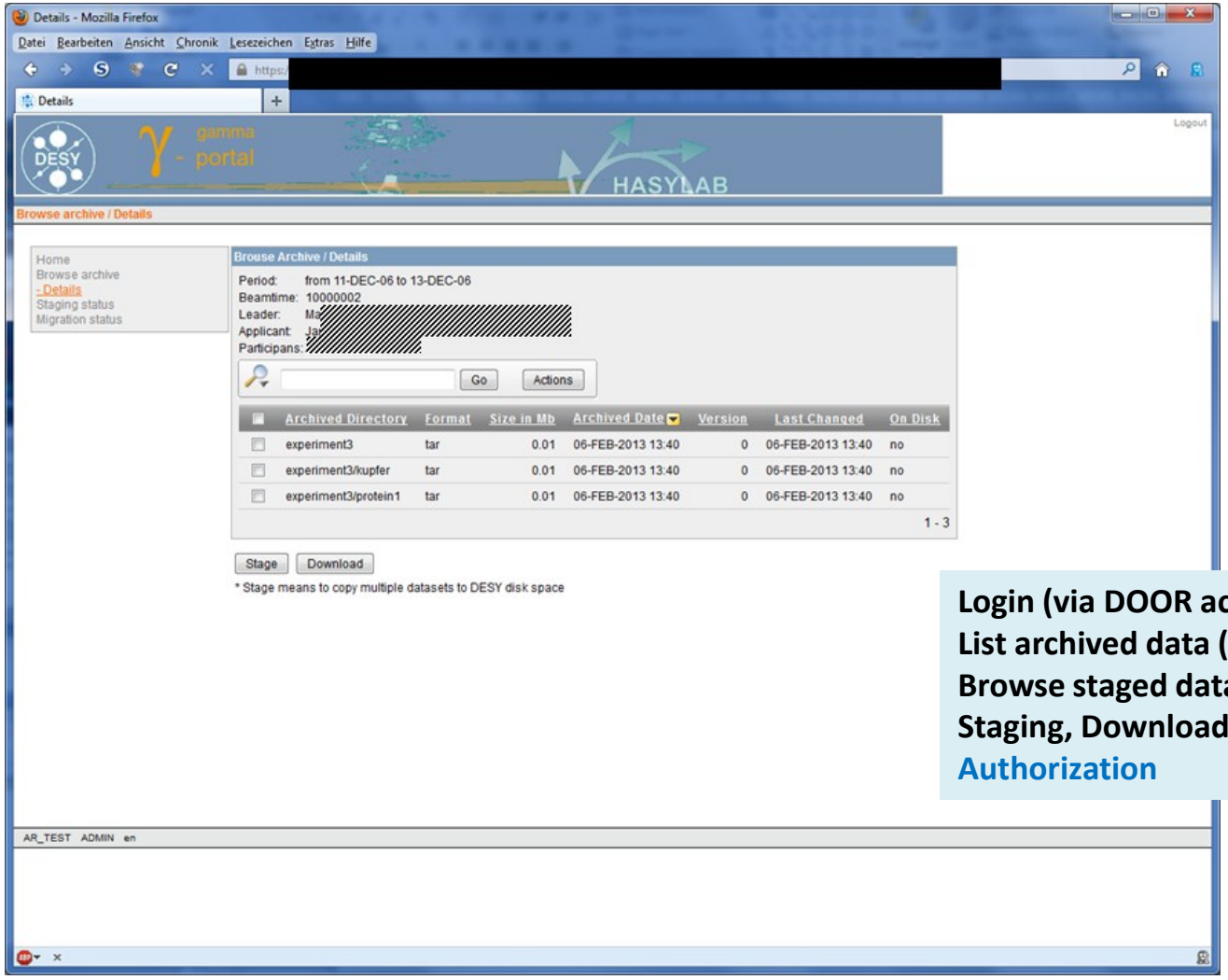
Web Portal: Components

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

*Compilation of meta data : international concerted action (PanData)

γ-Portal

https:gamma-portal.desy.de



Login (via DOOR account)
List archived data (keyword search)
Browse staged data
Staging, Downloading (present: single tarballs)
Authorization

Questionaire on data storage preferences June 2012

Interrogated Projekt Leaders in DOOR: **1181**
Response: **407 (34,5 %)**

Groups from:	Number	Storage at DESY	Storage at Institution
Private companies	3	1	2
Publicly funded institutions	159	55	104
Universities	245	81	164
Total	407	137 (33,7 %)	270 (66,3 %)

Data Policy

Principal Investigator (PI) has full responsibility for the data

This includes adherence to all rules imposed by legislative and funding bodies

Facility offers Data Storage Services for a ten years life cycle

- provides web access for authorized persons
- fast access during first three months for all data
- archiving on user demand (fee may be charged) otherwise data may be deleted
- archived data will be staged on demand

Disclaimer