Outline

- Introduction: objectives and tasks of WP-5
- WP-5 activities at DESY
- Requirements for the virtual labs

WP-5 Objectives

- Demonstrate PANDATA services and JRA results using 3 different techniques
  - Structural 'joint refinement' against X-ray and neutron powder diffraction data, VL1
  - Simultaneous analysis of SAXS and SANS data for large scale structures, VL2
  - Access to tomography data exemplified through paleontological samples, VL3

VL1, VL2 Tasks

- An authenticated user searches data in a catalogue
- Access is authorized and data downloaded from facility archive
- Search relevant software in a database
- Download software and run it locally or at the facility
- Analyze data
- Upload results and any relevant reduced data to facility archive

VL3 Tasks

- Setup a public access DB for storing tomographic raw and processed data of paleontological data e.g. 2D tomographs and 3D processed images
- Provide authorized access from multiple institutes to store processed data in the data base
- Enable public access to data in database
- Implement long term archiving for database

How to proceed

- Collect the requirements for the WPs
- Implement SDF, install Umbrella and ICAT at DESY
- Download/upload software from/to PANSOFT
- Make Tests
  - need help for the neutron analysis
- Give feedback to the service providers
- Report on the implementation
- VL3 questions
  - Who has data of paleontological samples?
  - Who is going to work with the data?
  - Is VL3 just a specific view in the data catalogue?

WP-5 activities at DESY

NeXus @ DESY

- new c++ library, exclusively based on HDF5
  - Purely object-oriented interface, for details see Eugen Wintersbergers talk
  - Coordinated with NIAC
  - Status: API published
  - Code on source forge
    - https://sourceforge.net/projects/libpninx
    - https://sourceforge.net/projects/libpniutils
- Python bindings
  - Status: ongoing

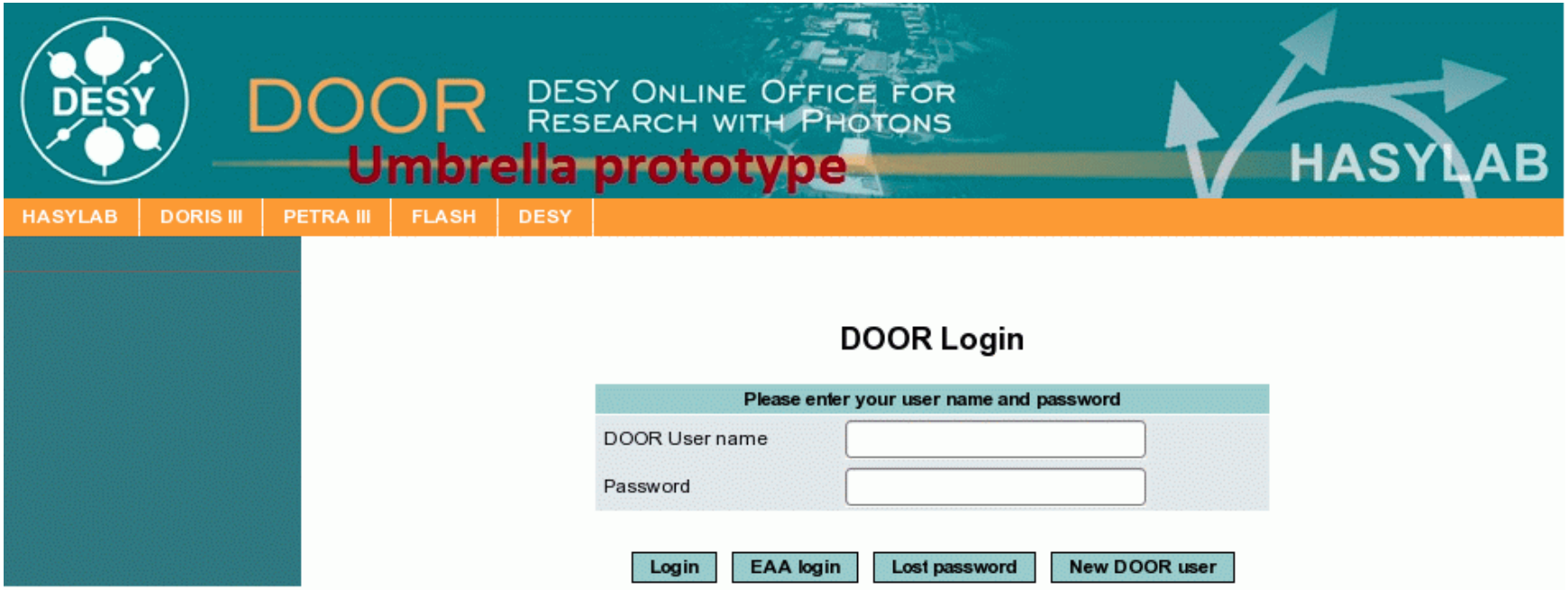
NeXus @ DESY

- Finalizing NeXus application definitions for scattering experiments, see talk of Rainer Gehrke
- Next steps:
  - Implementation at PETRA III beamlines, starting with SAXS, date: Apr, 1st, new colleague
  - NeXus writer : Tango server
    - Configuration files to generate meta-data
    - Collect data from other Tango servers
  - Adaptation of analysis software
  - Setup work flow

@ DESY

- DOOR/Umbrella prototype, https://shibs.desy.de
  - (Door login)
  - Register at Shibboleth (umbrella.psi.ch)
  - Match DOOR & Shibboleth accounts
  - EAA Login (Umbrella)

DOOR/Umbrella Prototype



@ DESY

- DOOR/Umbrella prototype, https://shibs.desy.de
  - (Door login)
  - Register at shibboleth (umbrella.psi.ch)
  - Match Door & Shibboleth account
  - EAA Login (Umbrella)
- Friendly user phase
- Status: Test system is operational
- Future plans: use Umbrella for DOOR and data portal

Requirements

WP-5 collects the requirements for the other work packages → D 5.1

- Standard data format
- Authentication
- Data catalogue
- Software catalogue
- Provenance
- Preservation
- Scalability

User vs. service provider perspective.

A questionnaire will be prepared that should cover all relevant issues.

Standard Data format

- PANDATA Europe: HDF5 agreed, NeXus is ‘starting point’ for file structures and application definitions
- General implementation of the SDF
- Application definitions
- Full description of the experiment and measurement within the files
- Data compression
- Converters to other file formats

Interfacing applications to files

- Problem:
  - NeXus files from different facilities may not be fully standardized
  - Old, non-NeXus files exist and need to be processed
- Possible solution: customize applications

cdma
Common Data Model API

- A dictionary translates application notions to file entities
- It uses plugins to adapt to different formats
- Was presented on a PANDATA workshop by Alain Buteau and Majid Ounsy
- DESY decided to join this initiative

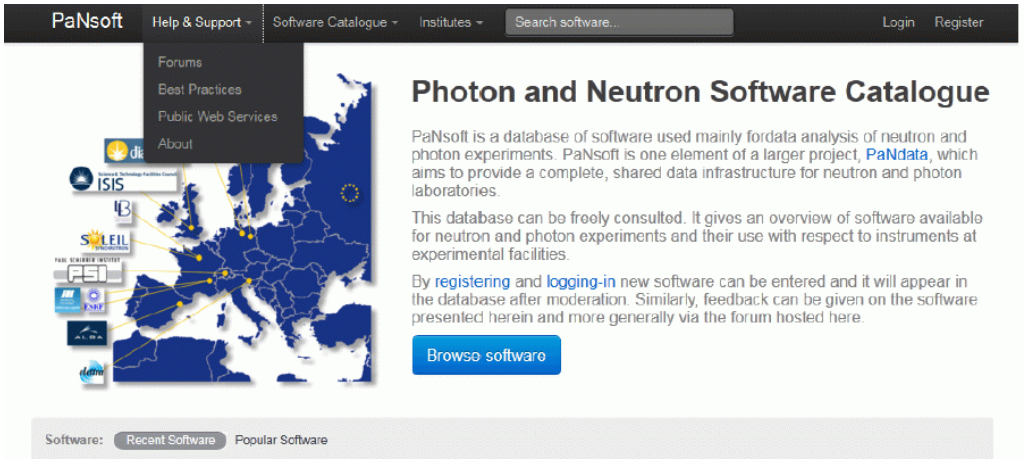
Common Authentication System

- PANDATA-Europe: Umbrella
- Single-sign-on to PANDATA partners
  - Digital user office
  - Data portal
- Common proposal submission system (?)

Data catalogue

- ODI proposal: ICAT is a strong candidate
- Data portal
- Federated catalogue
- Authentication
- Data download
- Keyword based searches
- Specify metadata
- Data viewer (?)
- Authorization management
- Performance
  - Response time for searches
  - Size of the catalogue
  - Download speed
- Cite data and link data and publications

Software catalogue



- PANDATA-Europe: PaNsoft
- List of available software
  - Versions identified by DOIs
- Metadata
  - Author, maintainer
  - Download server
  - Supported data formats
  - Supported platforms
  - Detailed description, including experimental techniques using a ‘common vocabulary’
  - Usage guide
- Forum
  - Bug fixing
  - New features
  - Ranking

Work flow management

- Data download may not be possible due to large data volumes
  - Current data rates: 1-10 TB per day per experiment
  - Network speed: 2 x 10 Gbps (DESY) → 30 TB/day
  - Will the network bandwidth cope with the new detector outputs?
- Reconstruction/analysis/simulation preferably guided by the facility experts
- Once established, work flows can also be used for near real time analysis for a fast feedback of the ongoing measurement
- Candidates: DAWB, …

Provenance

- Allow analysis validation
- Record the analysis process
- Common information model
  - Domain ontologies: beamline, instrument, equipment, techniques
  - Corresponds to SDF goals

Preservation

- Long term data storage: 10 years (?)
- Facility handles media changes
- Data access independent of platform
- Metadata management during the whole data live time

Scalability

Parallel file system & parallelized SDF

- Fast data storage has to cope with modern detectors
- Analysis within reasonable time:
  - Disk space is limited, e.g. DESY Online file servers: 140 TB
  - Goal: steady state data flow
  - Sufficient compute resources (not by PANDATA)
  - Fast data access, possibly adapted to specific techniques
- Near online analysis needs fast data access
- Users and facility scientists need a work flow management system

The next steps

- Discussion, feedback
- Prepare a questionnaire and send it to WP-5 contact persons
- Collect the results
- Prepare D 5.1

WP5 contacts

- ALBA, Joerg Klora
- CEA, Alain Menelle, Stephane Longeville
- DESY, Thorsten Kracht
- DIAMOND, Bill Pullford
- ELETTRA, Milan Prica, George Kourousias
- ESRF, Andy Goetz
- HZB Dietmar Herrendoerfer
- ILL, Miguel Gonzales
- PSI, Heinz-Josef Weyer
- SOLEIL, Alain Buteau
- STFC, Debbie Greenfield

Thanks for your attention

and let's open the discussion.

- Comments
- Missing items