

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”

High Data Rate Initiative für Photonen, Neutronen und Ionen

HELMHOLTZ ASSOCIATION

Home Über uns Veranstaltungen Aktuelles Publikationen Arbeitspakete Software

1. Libraries

The PNI libraries are closely related to each other as shown in the image below. The basement is libpniutils which provides the fundamental data types and classes. On top of this applications and libpninx are build.

python binding C++ apps and libraries

libpninx

libpniutils (C++)

libpniutils - utility library

Home Über uns Veranstaltungen Aktuelles Publikationen Arbeitspakete Software

- 1. Libraries
- 2. Beamline Software
- 3. User Software

Search

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

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



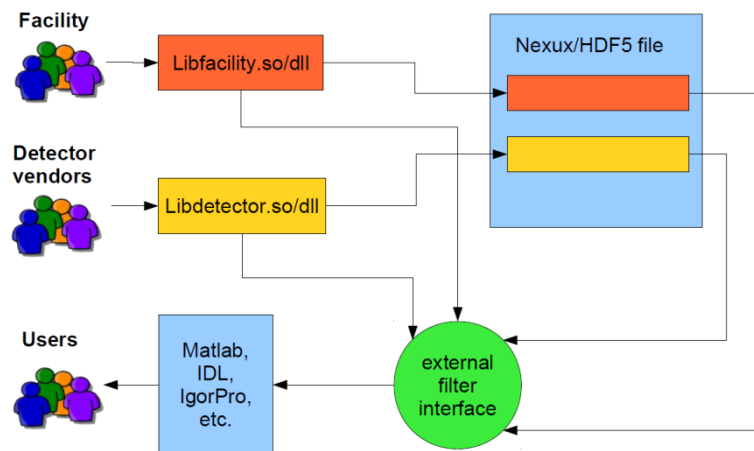
NeXus - Accompanying Measures

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

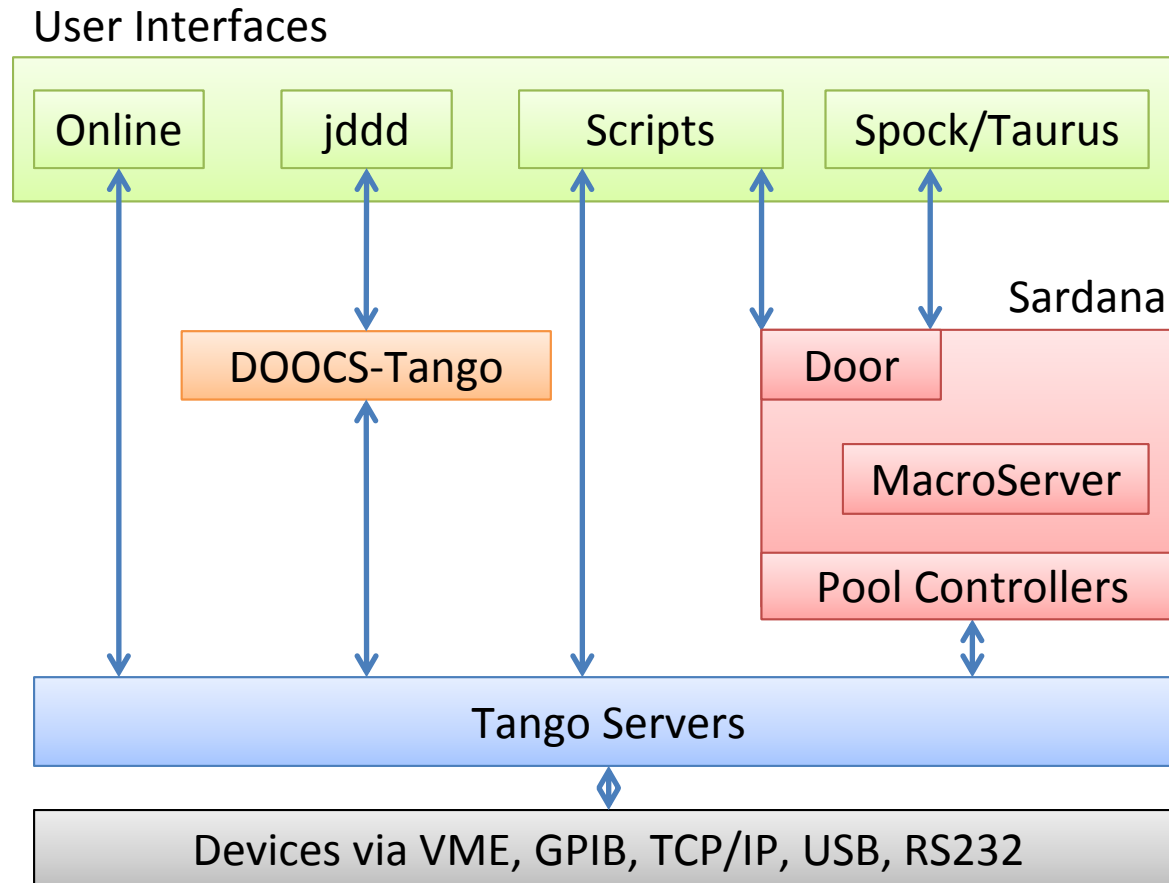
NeXus will adopt IUCR standards for describing geometries

CBF library will read NeXus files



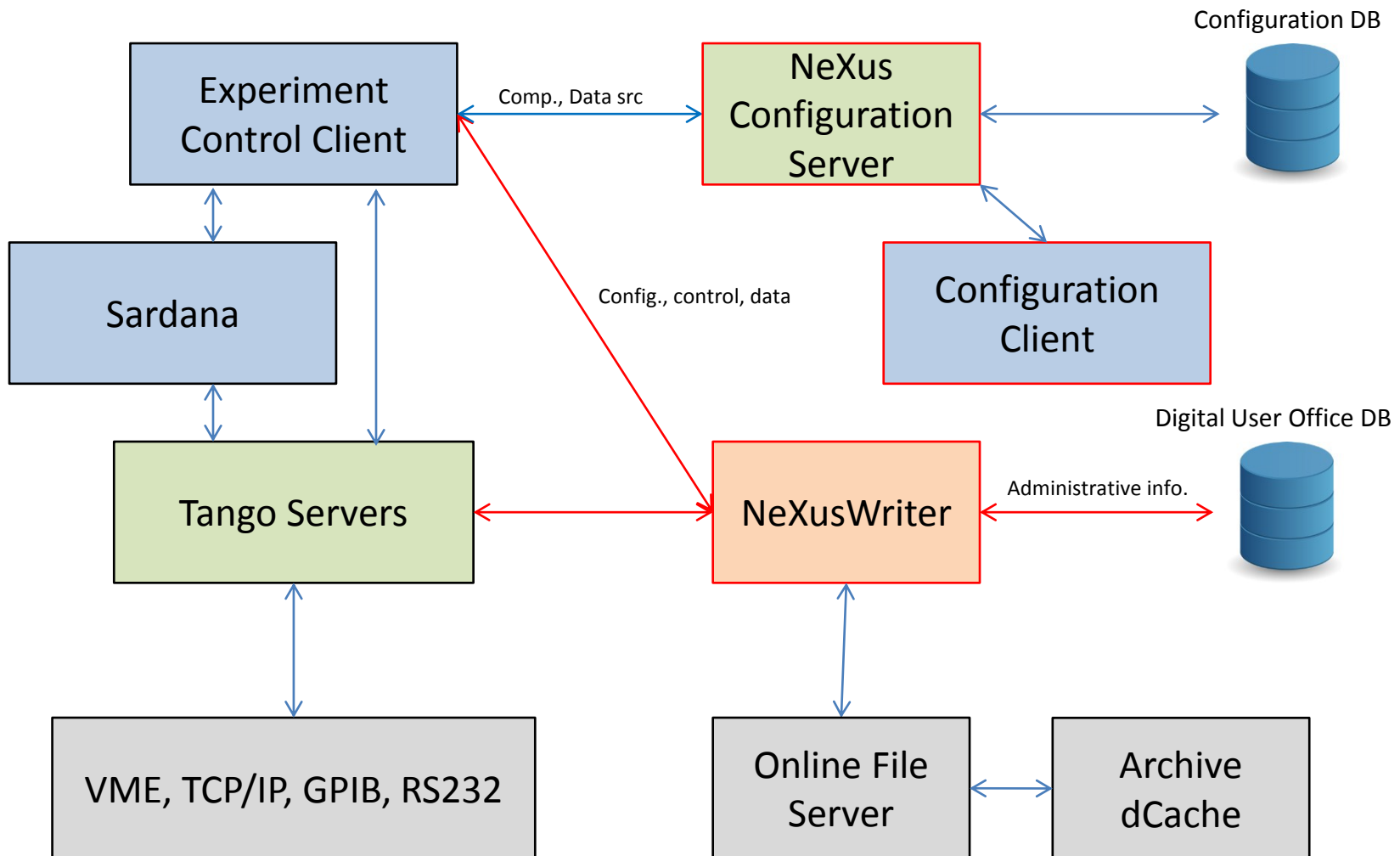
The Experiment Control System at PETRA III

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



NeXus Implementation – Step 2 – NeXus Data Server

J. Kotanski



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

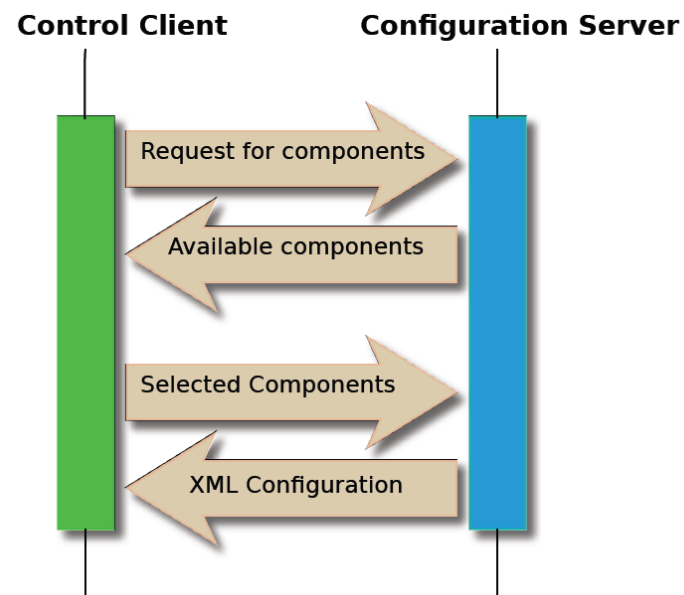
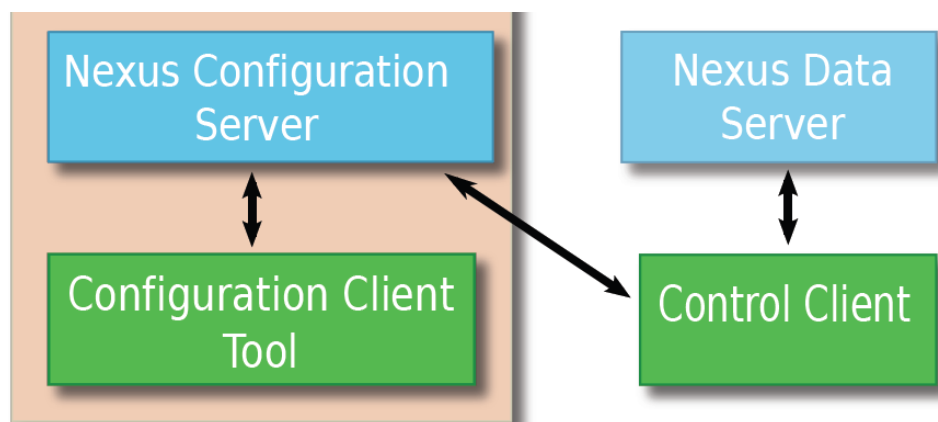
Configuration: a collection of **components**

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

The screenshot displays the Component Designer Configuration Client Tool with three open windows. On the left, a 'Collections' panel shows a list of components and data sources. The 'Components' tab is active, listing: MNI, beamline, diffractometer, mythen (selected), scan, scan2, scan3, test, and trigger. The 'DataSources' tab lists: dataSource1, dataSource2, dataSource3 (selected), and dataSource4.

The 'Component: beamline' window shows a hierarchical tree structure for configuration:

- definition
 - group: entry
 - group: instrument
 - group: source
 - field: distance (NX_FLOAT)
 - field: type (NX_CHAR)
 - field: probe (NX_CHAR)
 - field: emittance_x (NX_FLOAT)
 - field: emittance_y (NX_FLOAT)
 - field: name (NX_CHAR)
 - field: sigma_x (NX_FLOAT)
 - field: sigma_y (NX_FLOAT)
 - field: energy (NX_FLOAT)
 - field: current (NX_FLOAT)
 - group: undulator (NXinsertion_device)
 - field: type (NX_CHAR)
 - field: gap (NX_FLOAT)
 - field: poles (NX_INT)
 - field: length (NX_FLOAT)
 - field: energy (NX_FLOAT)
 - group: monochromator (NXmonochromator)
 - field: wavelength (NX_FLOAT)
 - field: energy (NX_FLOAT)
 - group: crystal (NXcrystal)
 - field: experiment_identifier (NX_CHAR)
 - field: experiment_description (NX_CHAR)
 - field: start_time (NX_DATE_TIME)
 - field: end_time (NX_DATE_TIME)
 - group: user (NXuser)
 - group: sample (NXsample)
 - field: name (NX_CHAR)
 - field: description (NX_CHAR)

The 'Component: diffractometer' window shows a similar structure:

- definition
 - group: entry
 - group: sample
 - field: x_translation (NX_FLOAT)
 - field: distance (NX_FLOAT)
 - field: y_translation (NX_FLOAT)
 - field: azimuthal_angle (NX_FLOAT)
 - field: polar_angle (NX_FLOAT)
 - group: instrument (NXinstrument)
 - group: detector (NXdetector)
 - field: distance (NX_FLOAT)
 - field: polar_angle (NX_FLOAT)
 - field: azimuthal_angle (NX_FLOAT)

The 'Component: mythen' window shows:

- definition
 - group: entry
 - group: instrument (NXinstrument)
 - group: detector (NXdetector)
 - field: data (NX_INT32)
 - field: x_pixel_offset (NX_FLOAT)
 - field: y_pixel_offset (NX_FLOAT)
 - field: description (NX_CHAR)
 - field: type (NX_CHAR)
 - field: layout (NX_CHAR)
 - field: count_time (NX_FLOAT)
 - field: x_pixel_size (NX_FLOAT)
 - field: y_pixel_size (NX_FLOAT)
 - field: beam_center_x (NX_FLOAT)
 - field: beam_center_y (NX_FLOAT)

Each window has 'Save' and 'Close' buttons at the bottom.

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

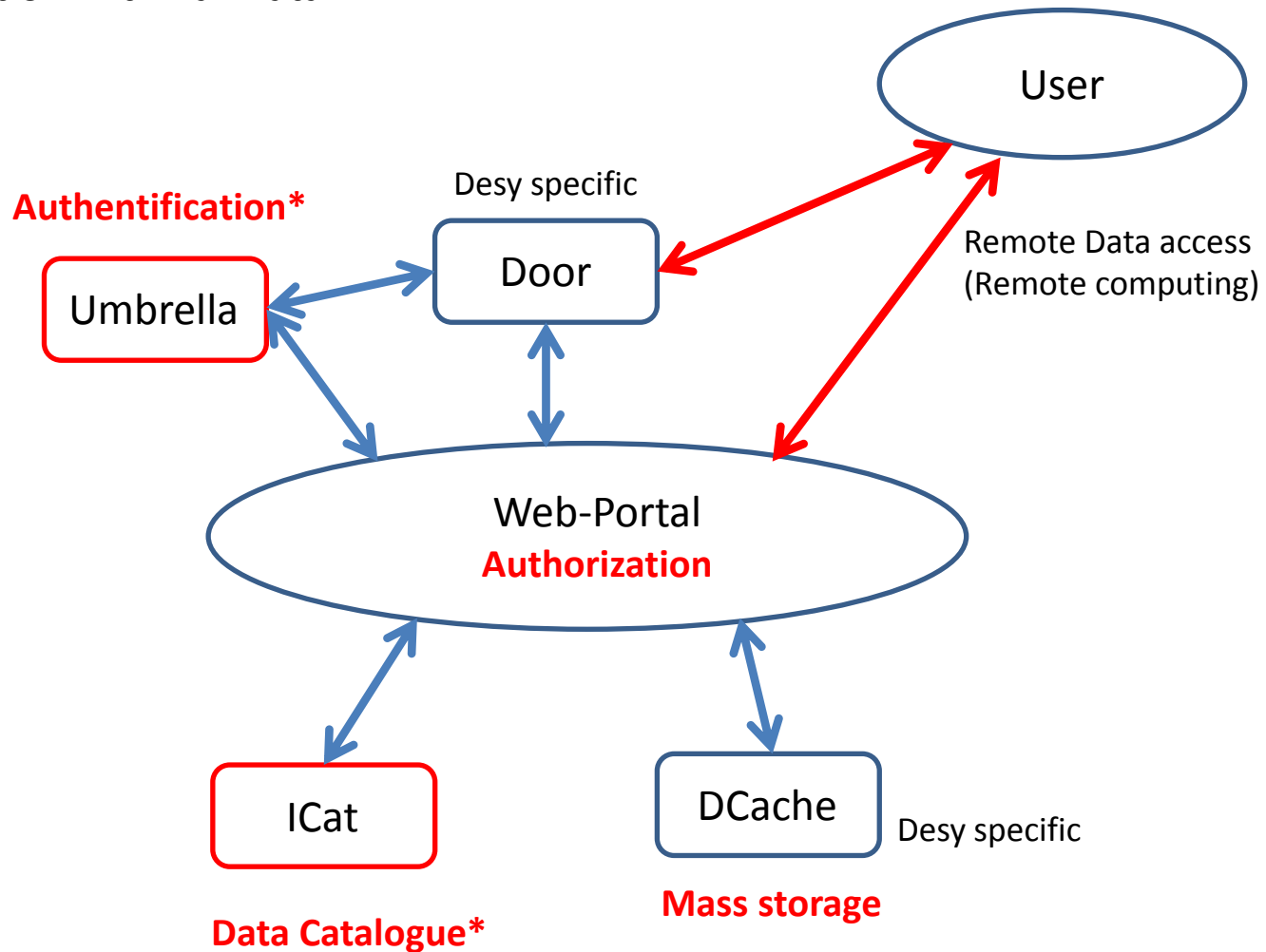
Integration into beamline control (showcase P03, microSAXS/WAXS)

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	Authentication	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

The screenshot shows the gamma-portal website in a Mozilla Firefox browser window. The page title is 'Details - Mozilla Firefox'. The address bar shows 'https:gamma-portal.desy.de'. The page header includes the DESY logo, the text 'gamma portal', and 'HASYLAB'. A 'Logout' link is visible in the top right corner. The main content area is titled 'Browse archive / Details'. On the left, there is a sidebar with links: 'Home', 'Browse archive', '-Details', 'Staging status', and 'Migration status'. The main content area displays a search bar with a 'Go' button and an 'Actions' button. Below the search bar, there is a table of archived data. The table has columns: 'Archived Directory', 'Format', 'Size in Mb', 'Archived Date', 'Version', 'Last Changed', and 'On Disk'. The table contains three rows of data, all with a size of 0.01 Mb and a version of 0. Below the table, there are 'Stage' and 'Download' buttons. A note at the bottom states: '* Stage means to copy multiple datasets to DESY disk space'.

Archived Directory	Format	Size in Mb	Archived Date	Version	Last Changed	On Disk
experiment3	tar	0.01	06-FEB-2013 13:40	0	06-FEB-2013 13:40	no
experiment3/kupfer	tar	0.01	06-FEB-2013 13:40	0	06-FEB-2013 13:40	no
experiment3/protein1	tar	0.01	06-FEB-2013 13:40	0	06-FEB-2013 13:40	no

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



Questionnaire 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 %)

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