# Scientific Data Management (SDM)

Vincent Hardion, on behalf of the KITS Group
and all the stakeholders@MAXIV
User Meeting 2017

# Agenda

S D

M

MAX IV

MAX IV

“

**Every possible way to fail will fail.**

”

# simple as that?

# Only a storage issue?

SDM

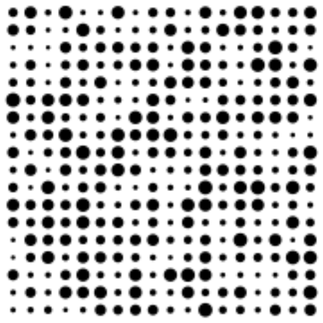
# Data Policy

Scope
Architecture
Data Model
Improvement

# Data Policy

owns

MAX IV stores 3 MONTHS FREE



are

# 2013: Data Generation focused

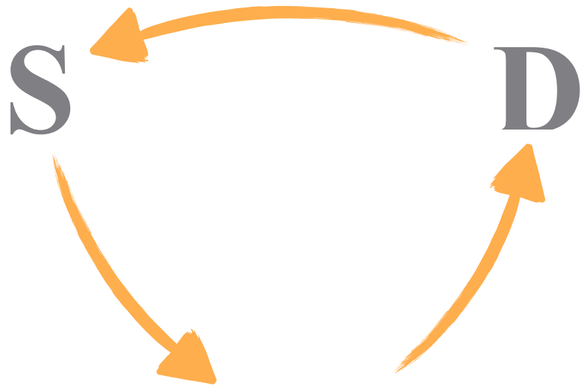
| Use case #1 | Use case #2 | Use case #3 |
| --- | --- | --- |
| Proposal Users can access the beamline | Acquisition Data stored in central storage server | User can retrieve the Proposal Data |

# Concern of the Data

| Submit proposal | Proposal review | Plan Visit + Arrival | Data Collection | Analysis | Scientific Results |
|---|---|---|---|---|---|

**Data Portal**

**User Office**




Data Collection & Analysis

MAX IV

MAX IV

# SDM

## Data Policy

# Scope

## Architecture

## Data Model

## Improvement

# Make it simple and transparent

“The Users have to do the job if the component don’t talk each other”

# Data and Service Interoperability

MAX IV integrates different software at different stages.

The integration is either in-house or by file. Use HDF5 as a standard format.

User Office

Data Acquisition

Data Storage

Data portal

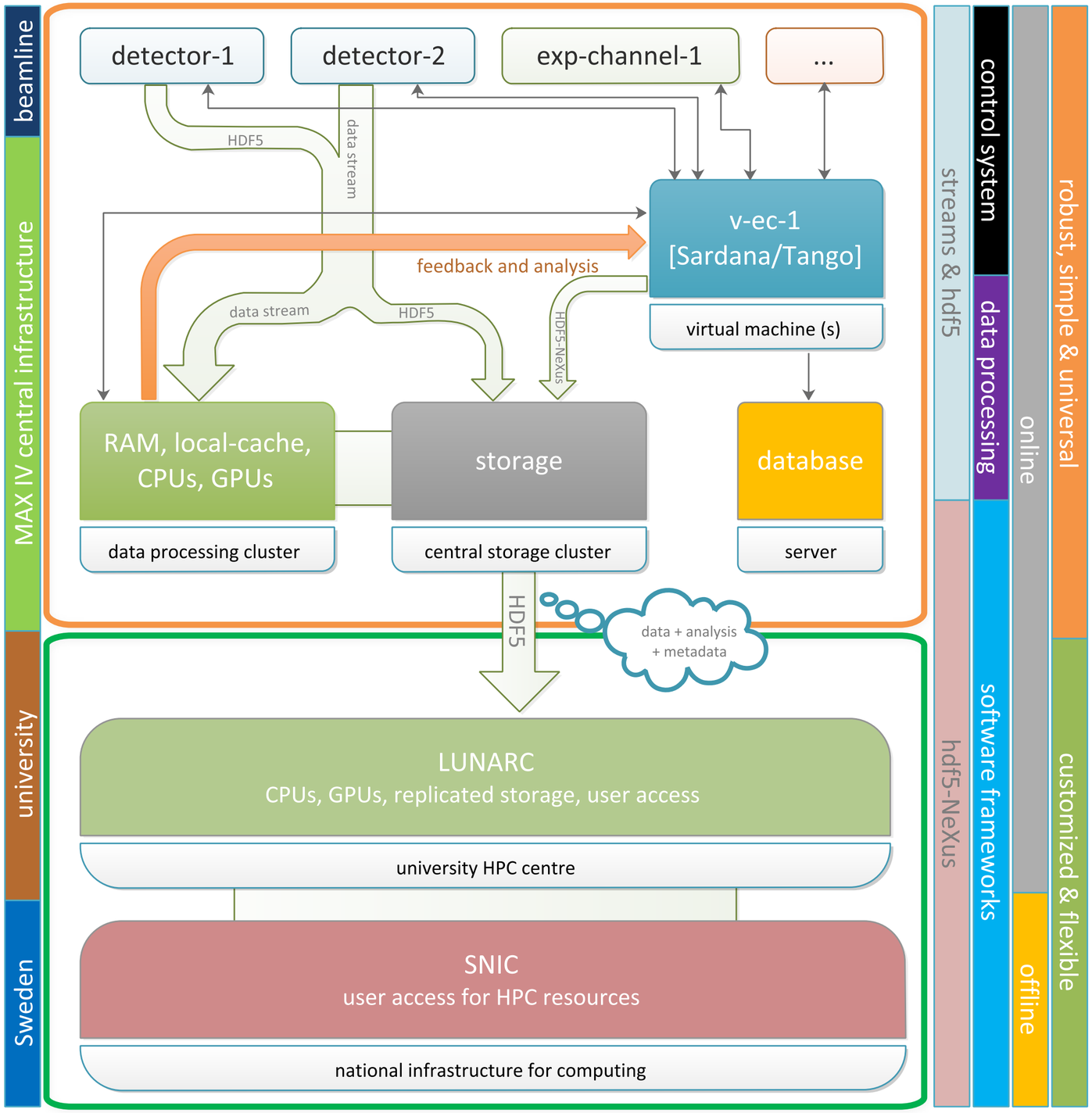
Data Visualisation

Data Analysis

Data Assessment

(Meta) Data catalogue

# Data flow




The MAX IV imaging concept
Matěj et al. ASCI 2017

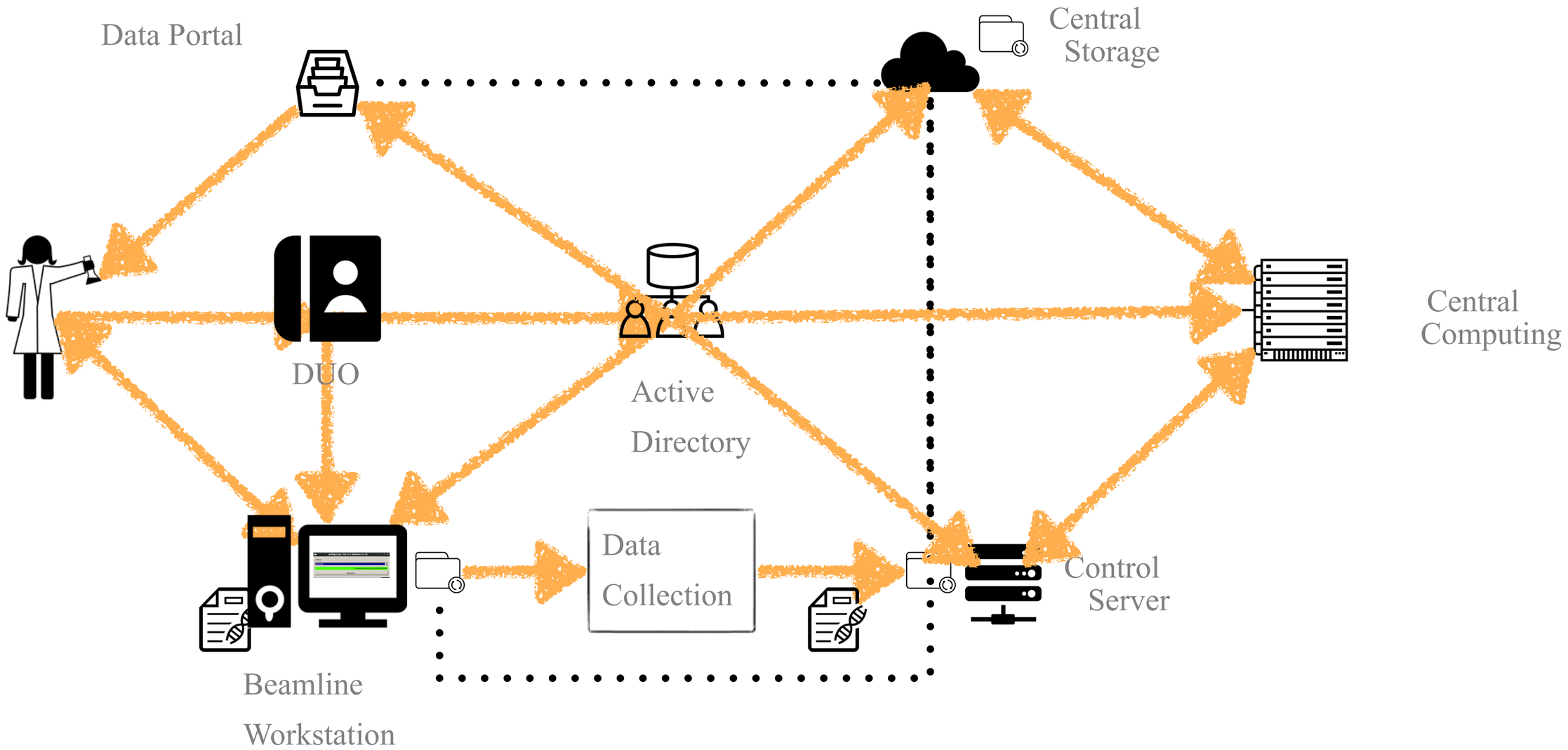
# SDM

## Data Policy
## Scope

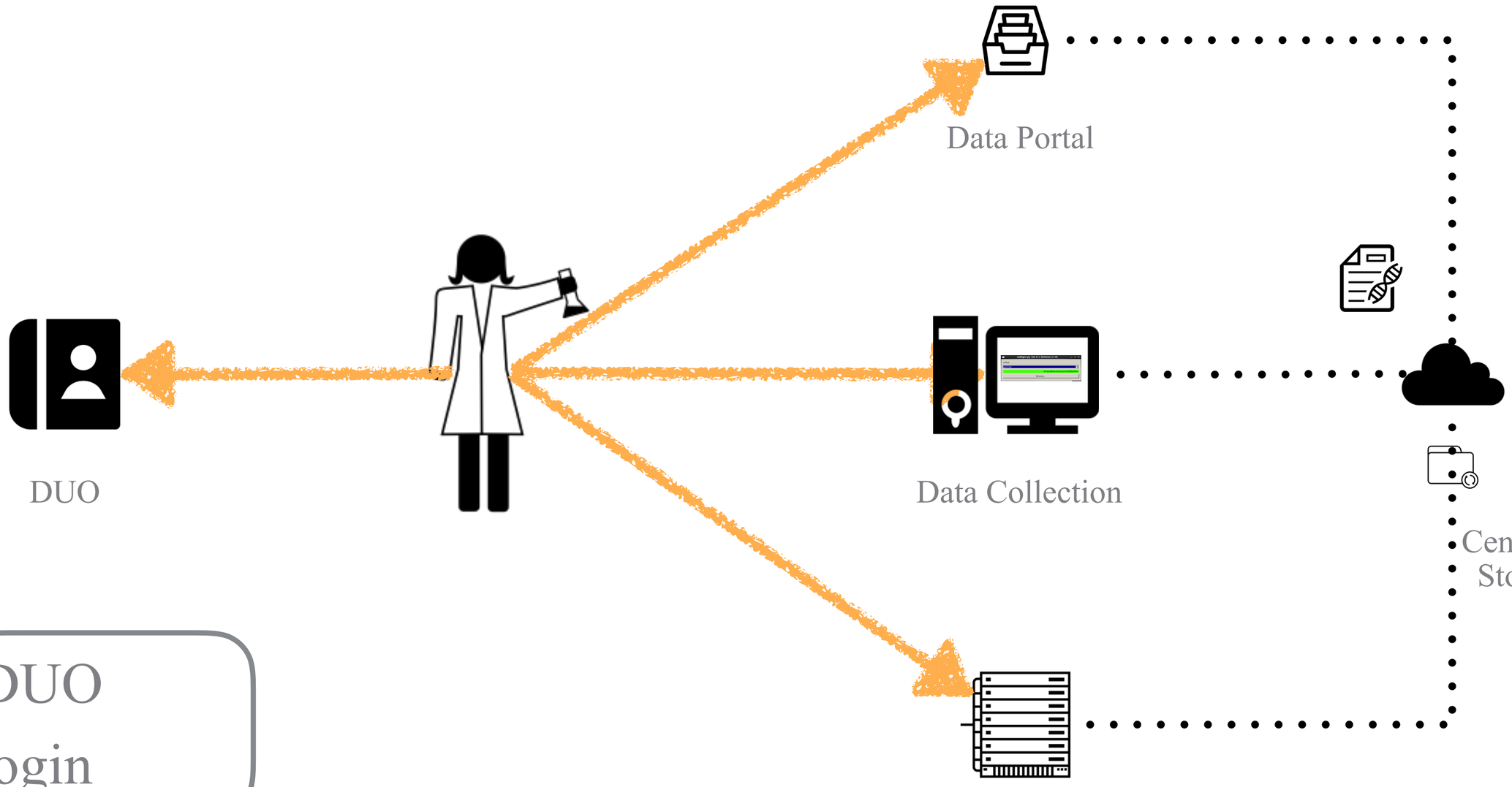
# Architecture

## Data Model
## Improvement

# Architecture



# User Interfaces and one login



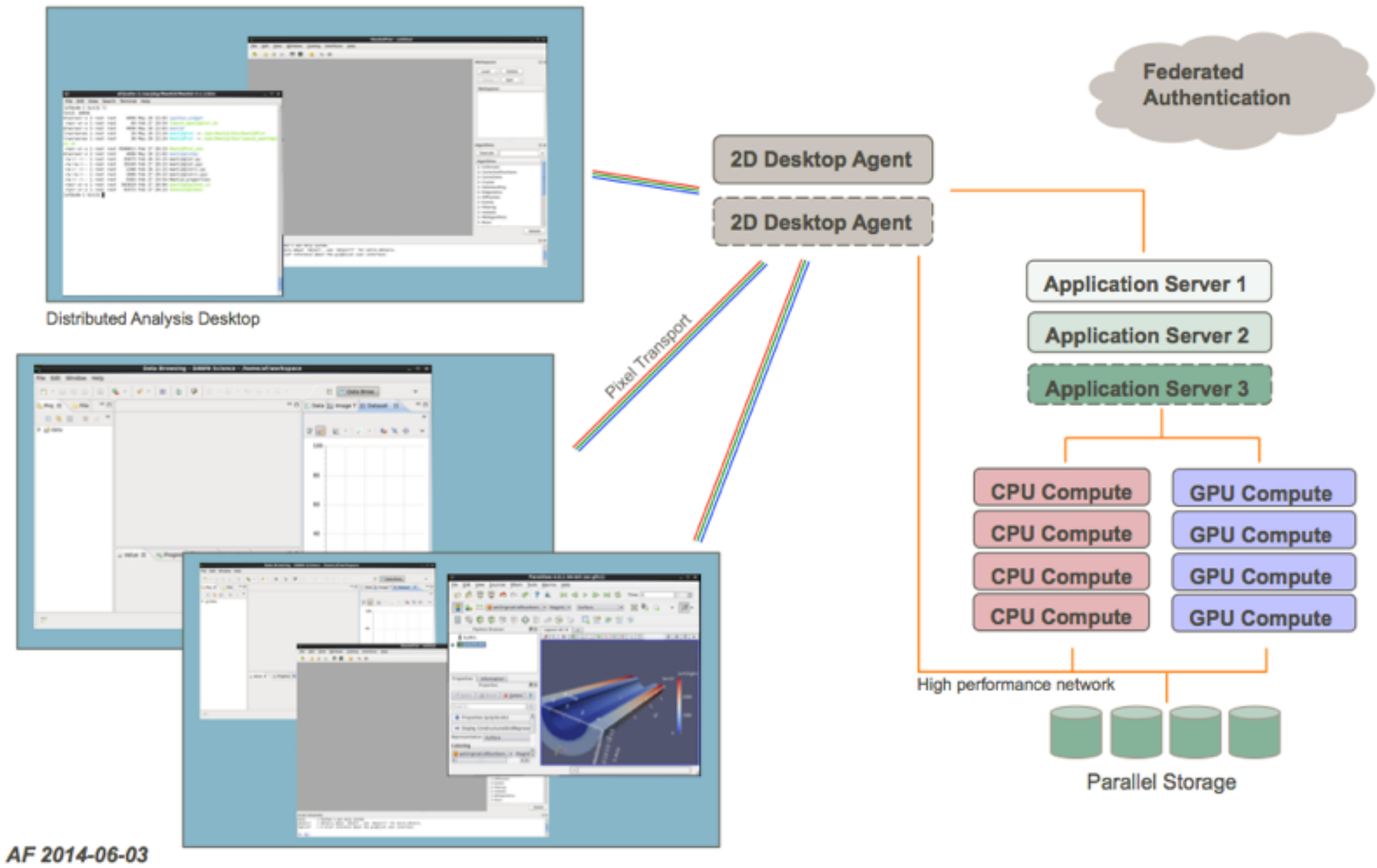

DUO login

Proposal login

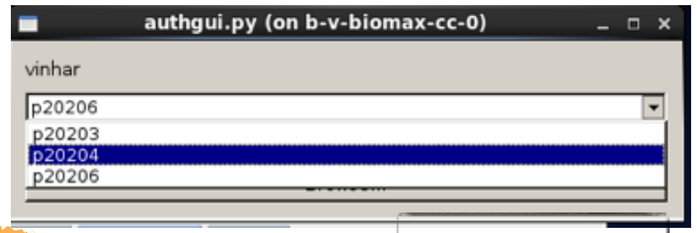
# Beamline experimental station

Authentication

- Group login during beamline shift
- Personal login for:
  - Data Analysis
  - Data Portal
  - DUO
  - …



# Data Collection - Proposal choice

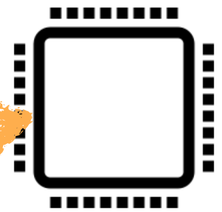


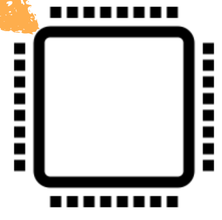

authgui

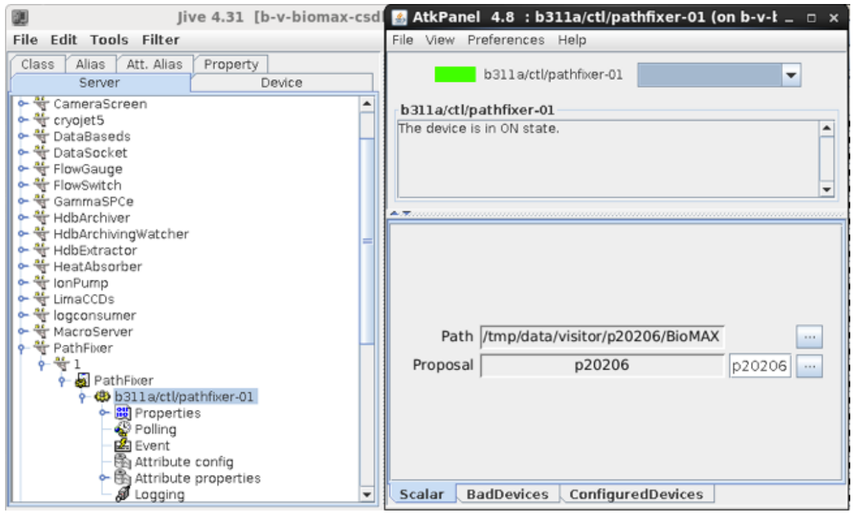
Control Server

PathFixer

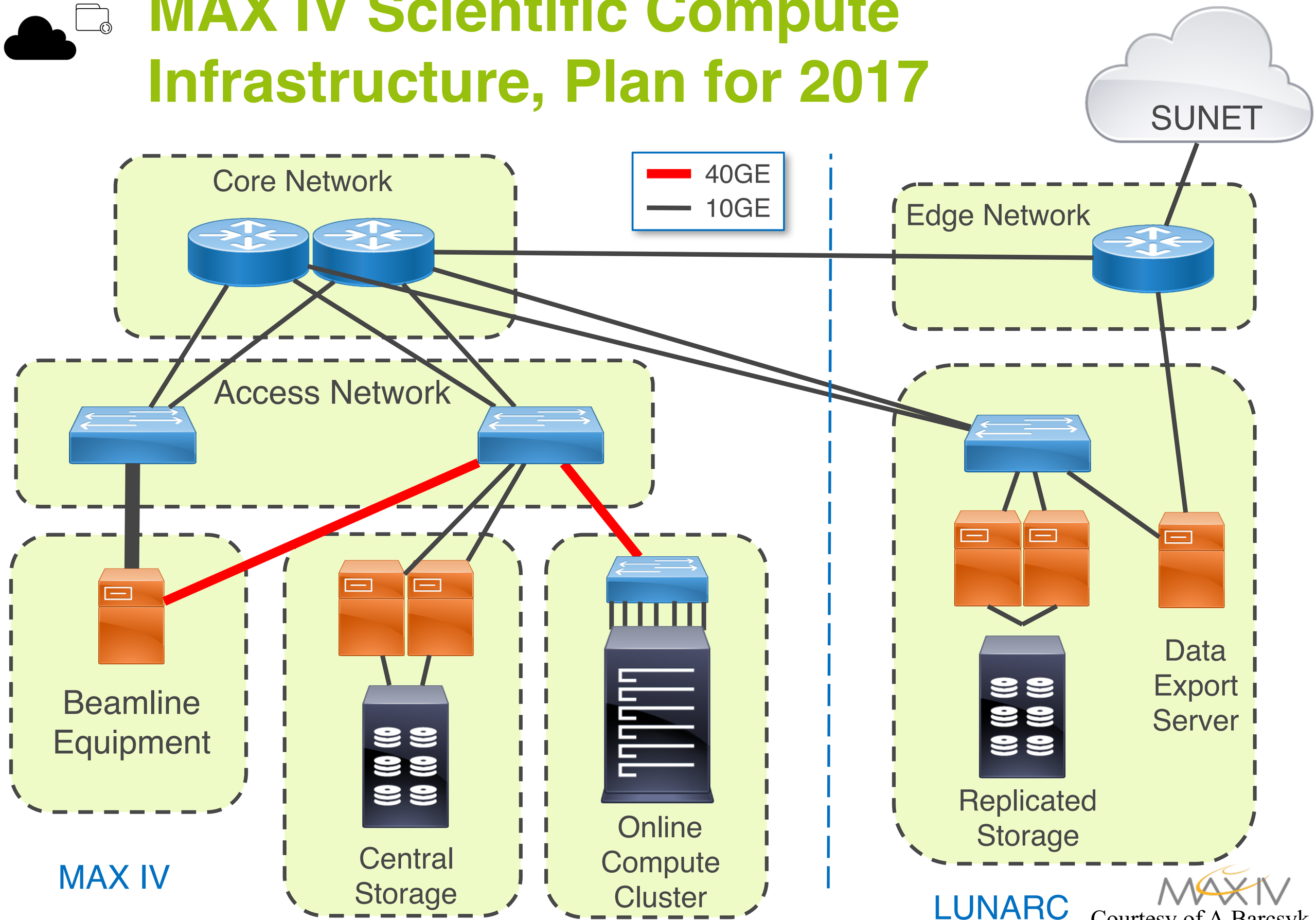
Sardana







# MAX IV Scientific Compute Infrastructure, Plan for 2017




Courtesy of A.Barcsyk

# Analysis software

## BLs sw

- **BioMAX**
  - on-site XDS and other relevant MX sw - deployed, final tuning
  - MX sw PreSTO pkg – just coming to LUNARC and MAX IV
- **NanoMAX**
  - Data explorer for Coherent scattering and XRF – based on SILX gui (ESRF
  - Ptypy – ptychography for python (DLS)
- **BLOCH**
  - ARPES visualization with Qt
- **BALDER**
  - GPU accelerated data reduction/ processing pipeline based on SILX (ESRF)
- **BioMedMAX** project
  - Log-polar Radon tomo reconstructio

## KITS

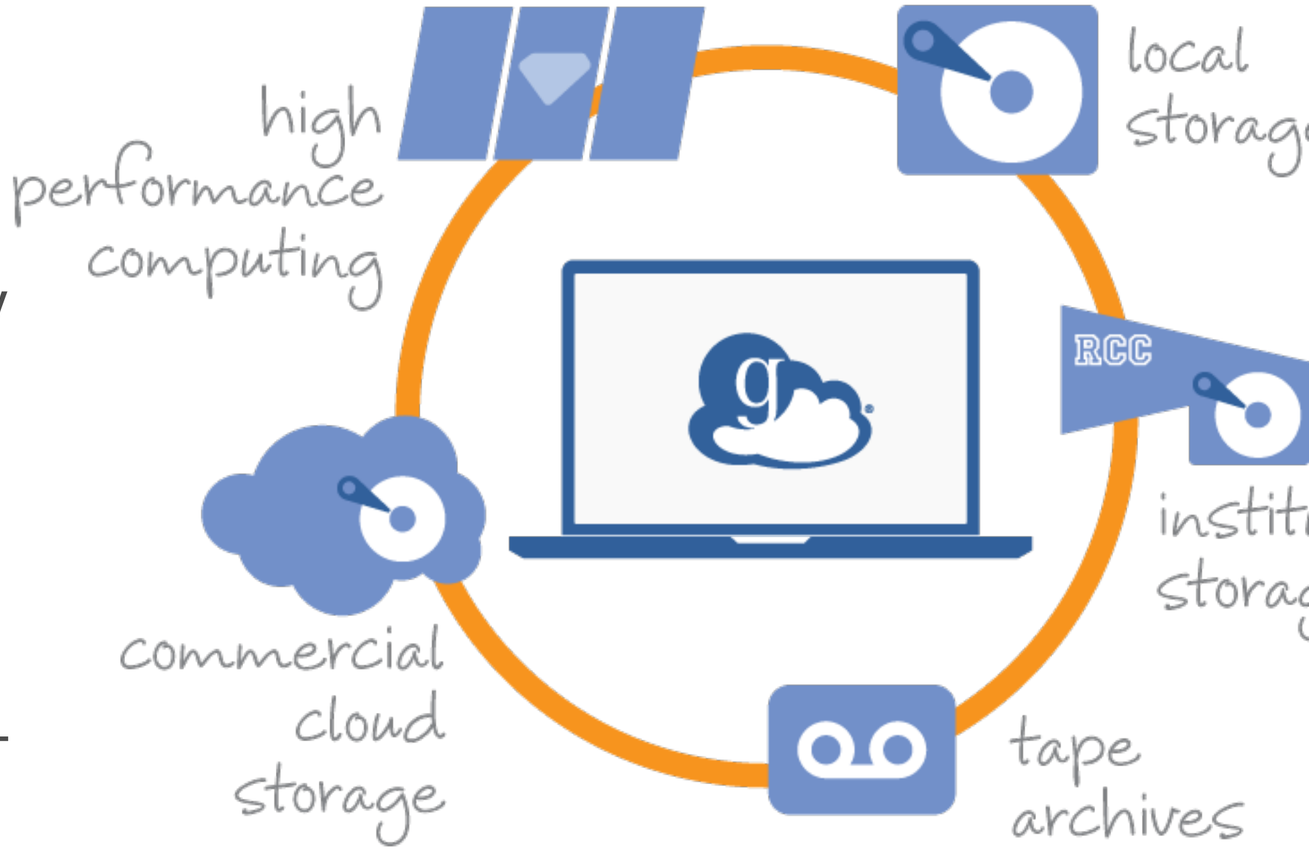
- **Live view data visualization** in bls (PyMca data source for MAX IV/ SILX)
- **HDF5 web gui**
  - quick HDF5 data inspection everywhere (“initially in MAX IV”)

## MAX IV

- **reliable data high throughput data processing** (parallel HDF5 writers)
- **infrastructure for MX experiments**
- **tools for processing large data** (0.1 - 1 TB datasets)
  - testing data processing framework (Savu – DLS) with tomography reconstructions

# Data Portal

- A user's data access rights are determined by proposal she belongs to
  - Independent whether connected inside or outside MAX IV
- Within the facility, the storage directory can be mounted via NFS or SMB
- Outside facility, a Data Transfer System will provide fast data movement capability
  - Globus/gridftp services
  - MAX IV is directly connected to SUNET
  - 10Gbps initially, 100Gbps later




Courtesy of A.Barcsyk

# SDM

**Data Policy**
**Scope**
**Architecture**

# Data Model

## Improvement

# User Data Collection

All software aware of the visitor and proposal save automatically in the right place.

In any case the storage is accessible everywhere on the beamline

Data generated by default as HDF5 or Nexus

Courtesy of A.Barcsyk

# Ownership and Access

**`/data/(user-type)/(beamline)/(proposal)/(visit)/raw`**

The (user-type) includes **visitors** (bread and butter academic users), **staff** (MAX IV staff) and **proprietary** (mostly industrial users with a different need for data security).
The (visit) should begin with the date when the first shift starts.
In addition to /raw there will be a /process folder at the same level with full access.

| path: | **/data/visitor** | **/beamline** | **/proposal** | **/visit/raw** |
|---|---|---|---|---|
| owner | root | beamline-services | beamline-services | beamline-services |
| group | root | beamline-staff | proposal-group | proposal-group |
| posix | 2755 | 2775 or 2755* | 2750 | 2770 |
| | static | static | dynamically created | dynamically created |
| | created on the central storage | created on the central storage for each beamline | dynamically created by a service account | dynamically created by a service account |

# Group and User model

Users:

- The same user can be part of several proposals groups
- The beamline-staff group is a member of the proposal group
- The beamline staff people is indirectly part of the proposal group via their beamline-staff group
- The beamline technical account are indirectly part of the proposal group via their beamline-staff group

| group | memberof | Comment | GID |
|---|---|---|---|
| Visitor | | primary group for any visitor | 1332 |
| Staff (MAX-Lab) | | primary group for any maxiv staff | 1330 |
| <proposal>-group | Visitor | group all visitors associated to the proposal | 200000-399999 |
| <beamline> | Staff, <proposal>-group | group all beamline staff, automatically associated to the <proposal> group for maintenance purpose | 1329-1400 |

| User | Primary Group | Proposal Group | Proposal Group member | Source | Directory | Comment | UID |
|---|---|---|---|---|---|---|---|
| <proposal> | Visitor | <proposal>-group | | DUO | AD only | generic proposal account,<br>mainly used to share the same session on the beamline computers | 400000-599999 |
| <user> | Visitor | <proposal>-group | | DUO | AD and DUO | visitor account,<br>can be used for all case and to read the data outside of the beamline | 400000-599999 |
| <maxiv account> | Staff | no | <beamline> | AD | AD | staff account,<br>can be used for own research, commissioning, maintenance of the Visitor | xxx |
| <beamline>-user | Staff | no | ???<beamline> ??? can have too much right | AD | AD | generic beamline account,<br>mainly used to share the same session on the beamline computers | 1990-2099 |
| <beamline>-service | Staff | no | <beamline> ??? can have too much right but limited if no login | AD | AD | generic beamline account,<br>mainly used by the software which create the data files; daemon service not intended to | 1990-2099 |

# SDM

## Data Policy
## Scope
## Architecture
## Data Model

# Improvement

# Data Collection - More

Use case #4

---

Privacy of the Proposal

Use case #5

---

Handle different Data layout

Use case #6

---

Merge Data from different detector

Use case #7

---

Manage Meta Data

Use case #8

---

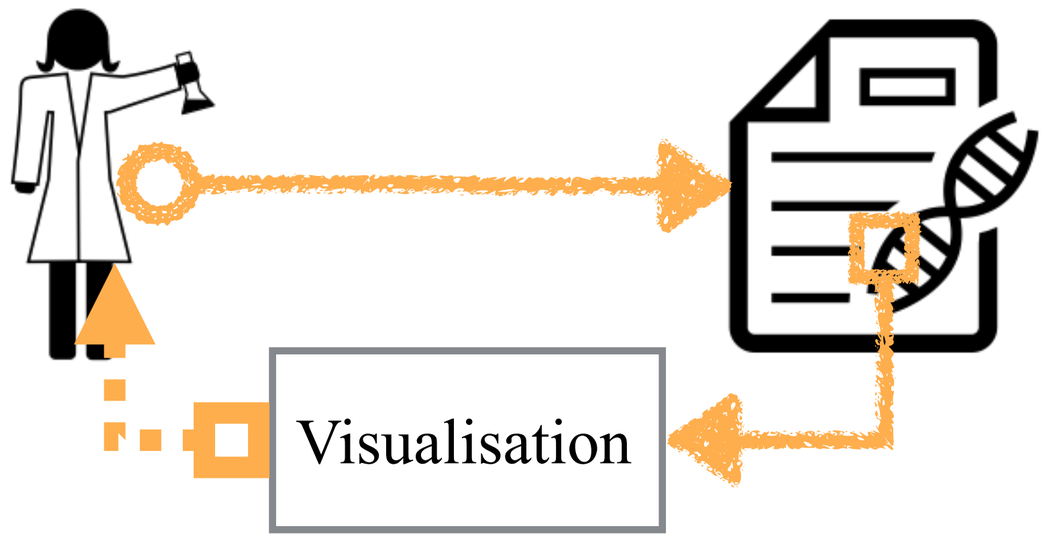
Simultaneous Read/Write

Use case #…

---

Performance

# On-the-fly analysis



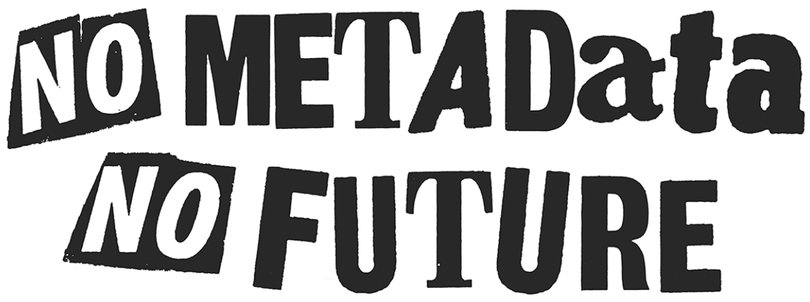

#streaming

# Future

Transfer data of User where it’s the best

iRODS

Meta Data Catalogue



Thanks for your attention

# Questions?

Credits:
Special thanks to Krister Larsson
All KITS members and honorary members.
Accelerator and Beamlines Staff
All our collaborators