Data Challenges in Photon Science

Manuela Kuhn
GridKa School 2016
Karlsruhe, 29th August 2016

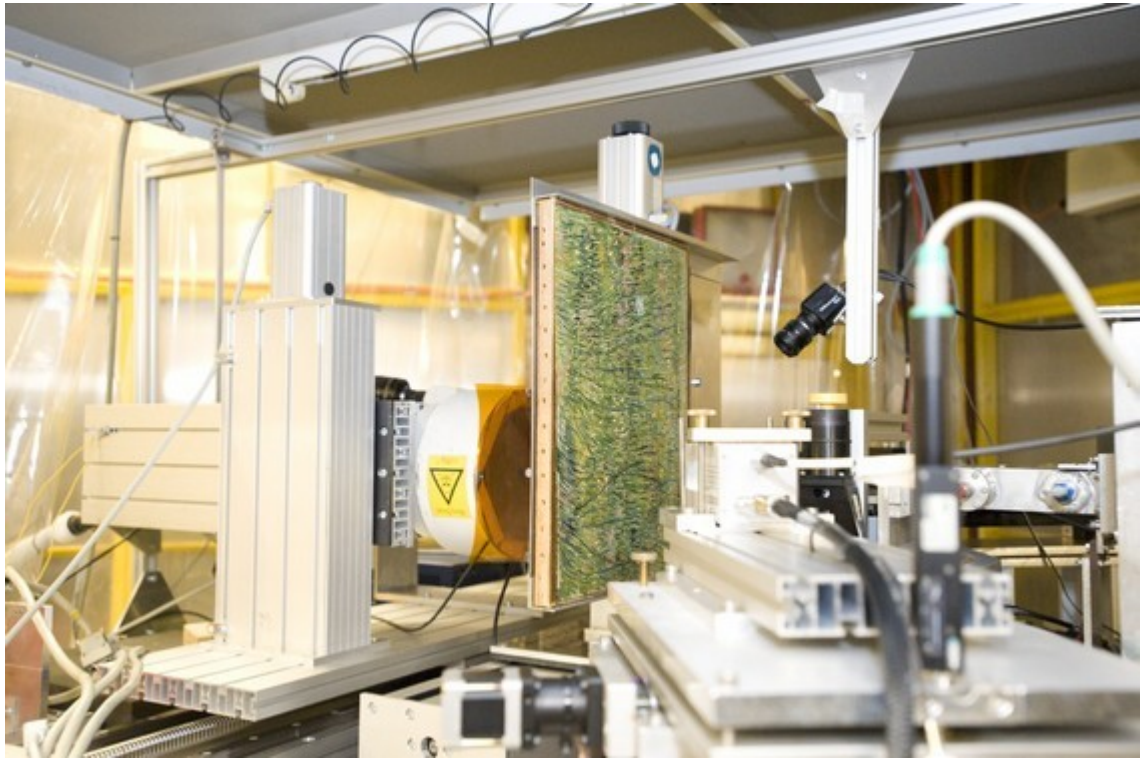
HELMHOLTZ ASSOCIATION

DESY

Photon Science

- Exploration of tiny samples of nanomaterials
- Synchrotrons and free electron lasers generate extremely powerful and focused radiation
- The X-ray beams are so intense that they can reveal even the finest details
- Examples:
  - Find the tiniest cracks and pores in a turbine blade or minute impurities in a semiconductor
  - See the positions of individual atoms in a protein molecule
  - With extremely short X-ray flashes it is even possible to observe ultrafast processes such as those that occur in a chemical reaction

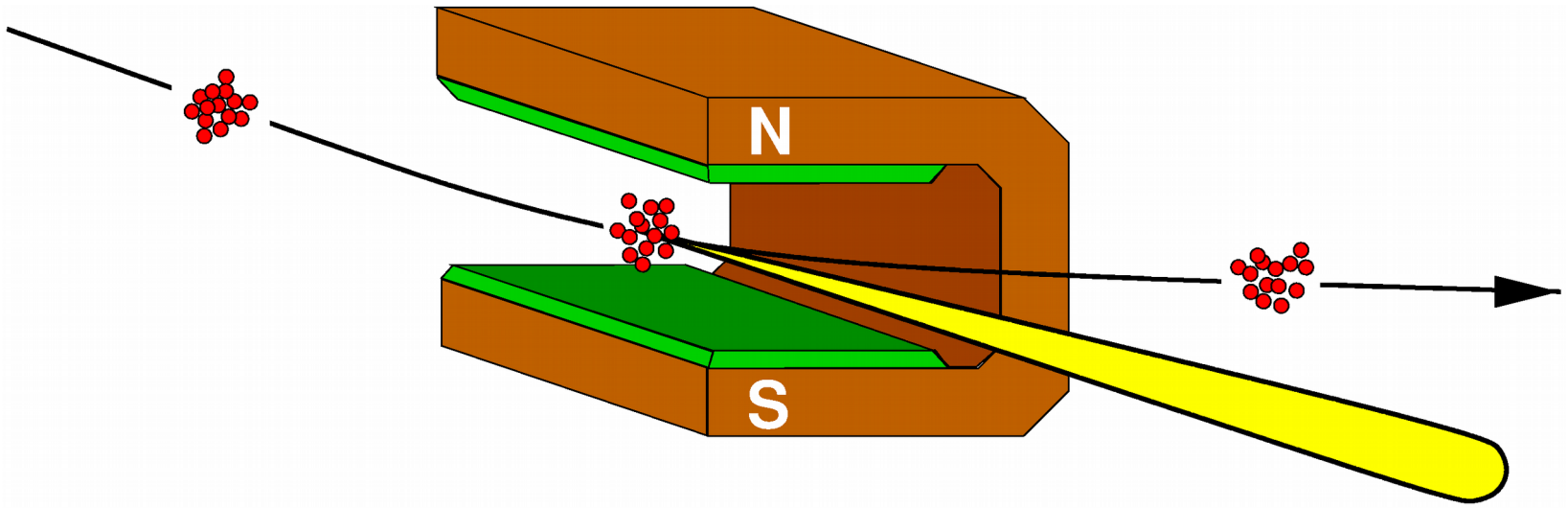
Photon Science



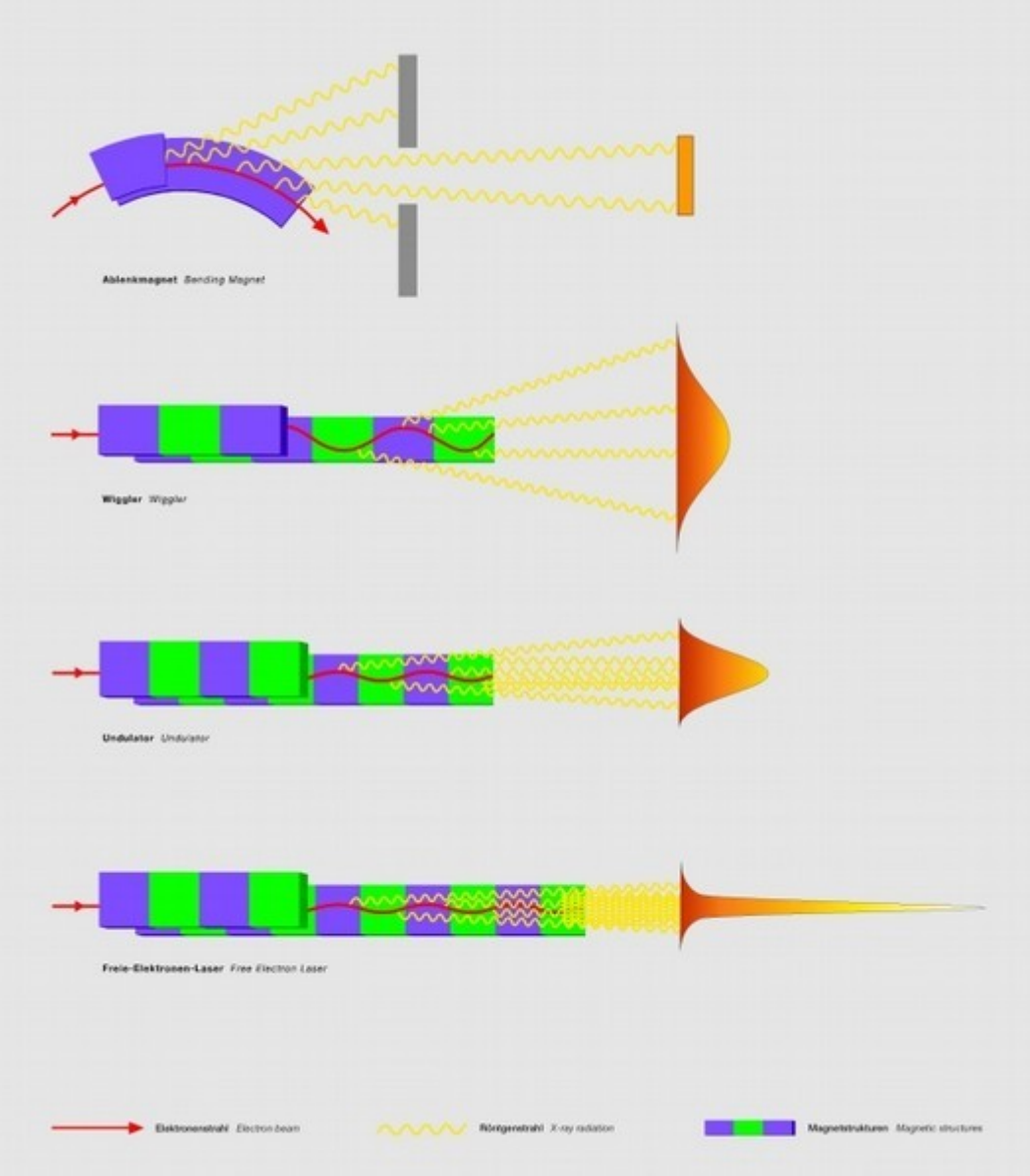
> Visualizing a lost painting by Vincent van Gogh using X-ray fluorescence mapping

Synchrotron radiation

- Synchrotron radiation is the radiation coming from a beam of electrons turning in a magnetic field
- The angular acceleration induces the radiation of photons which emerge tangentially to the curvature of the beam



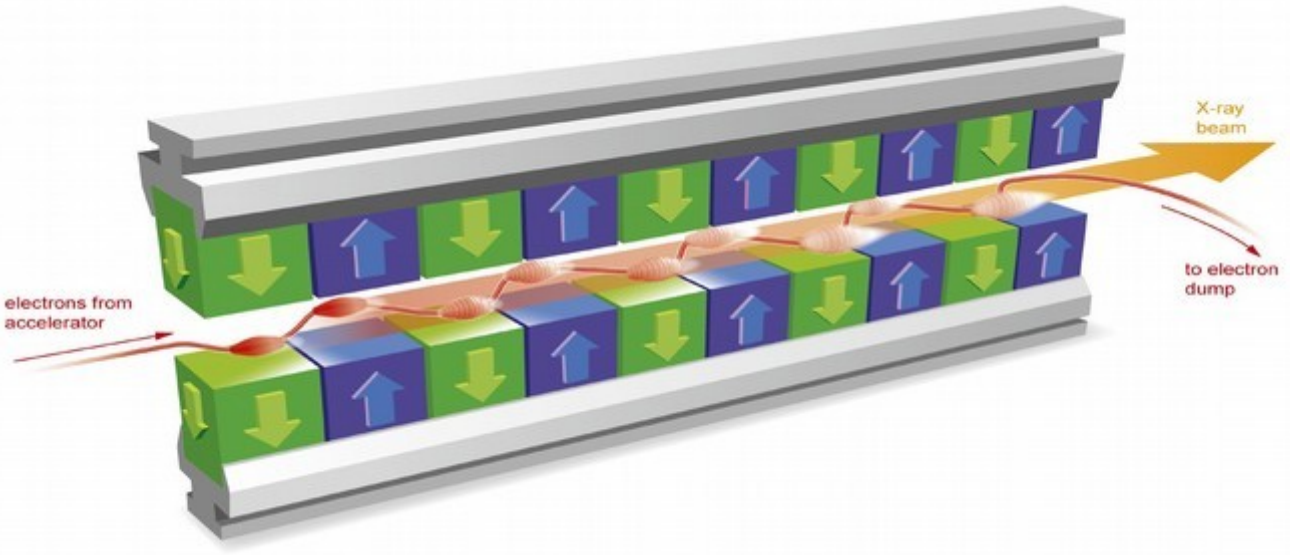
Synchrotrons and Free Electron Lasers



> Both synchrotron radiation and FEL have a particle accelerator to generate the beam

Free Electron Lasers

- Utilizes the synchrotron effect in order to create a linear photon beam
- The beam is similar to the one from lasers in its directionality and coherence
- The photon beam from a FEL can be controlled in its range in frequency and power (this is not true for synchrotron radiation)



DESY

- Founded December 1959
- Accelerator Center
  - Research, construction and operation
- Research topics
  - Particle Physics (HEP)
    - e.g. Higgs@LHC, Gluon@PETRA
  - Photon Science
    - X-ray crystallography, broad spectrum of application
  - Astro Particle Physics
- 2 Sites
  - Hamburg
  - Zeuthen (Brandenburg), near Berlin
- ~2300 employees, 3000 guest scientists annually

Hamburg

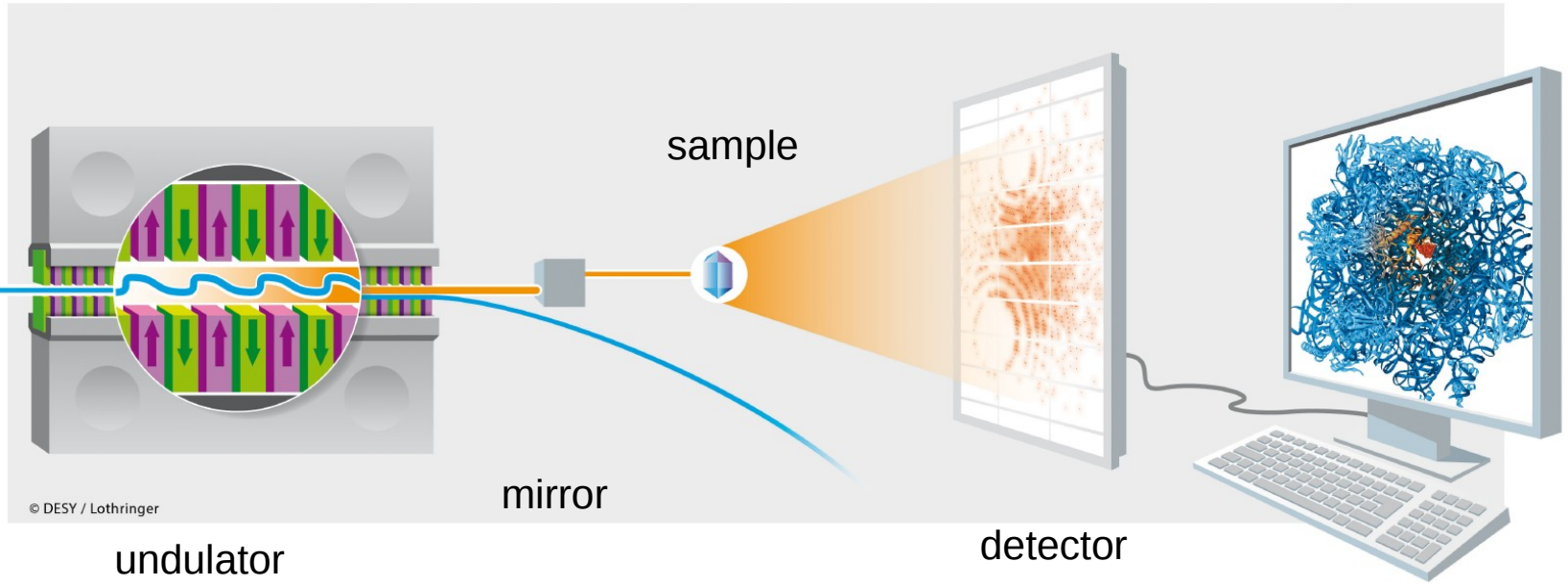
Zeuthen

Light Sources on the DESY campus

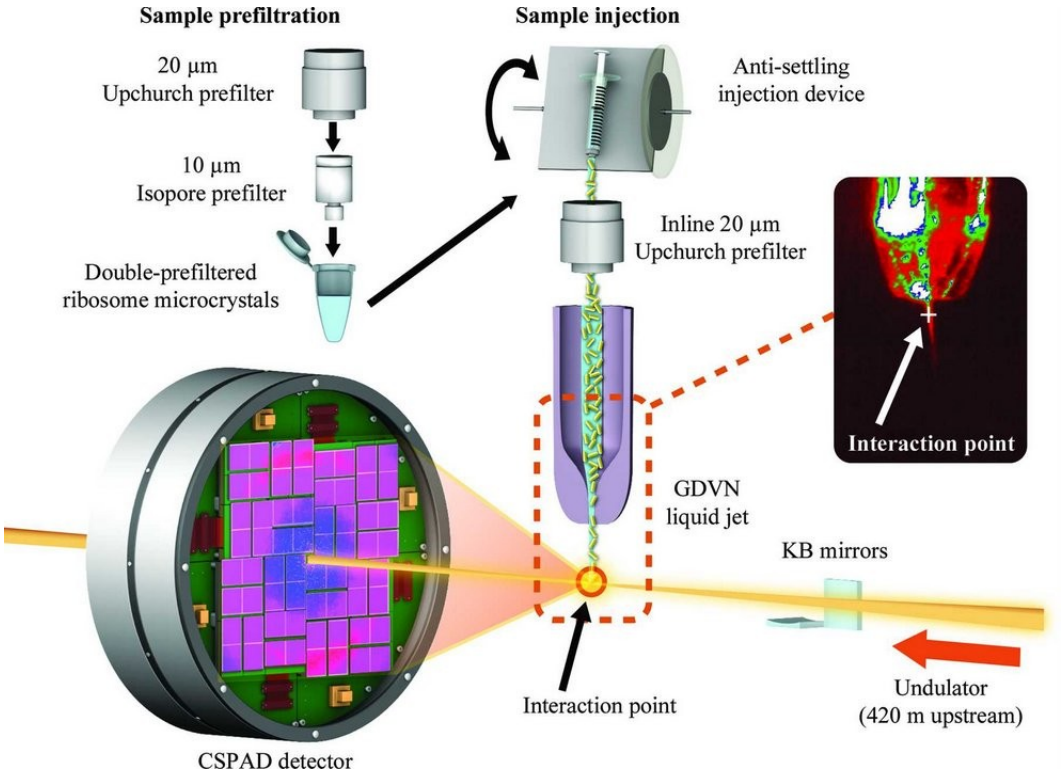
- PETRA III:
  With a circumference of 2.3 km the biggest and most brilliant synchrotron light source in the world
- FLASH:
  The first free-electron laser worldwide to produce femtosecond pulses of soft X-rays

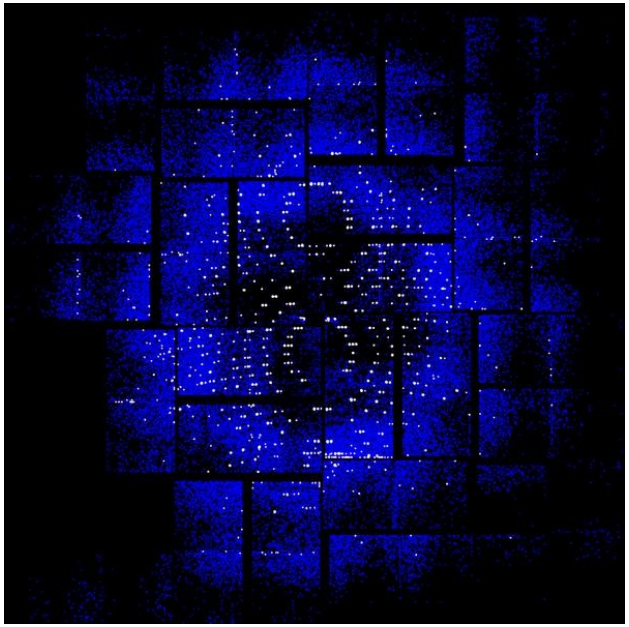
- European XFEL:
  Once finished, this facility will deliver hard X-ray pulses far shorter than those from any other X-ray source, and their peak brilliance will be up to eight orders of magnitude higher

Experiment example - schematic



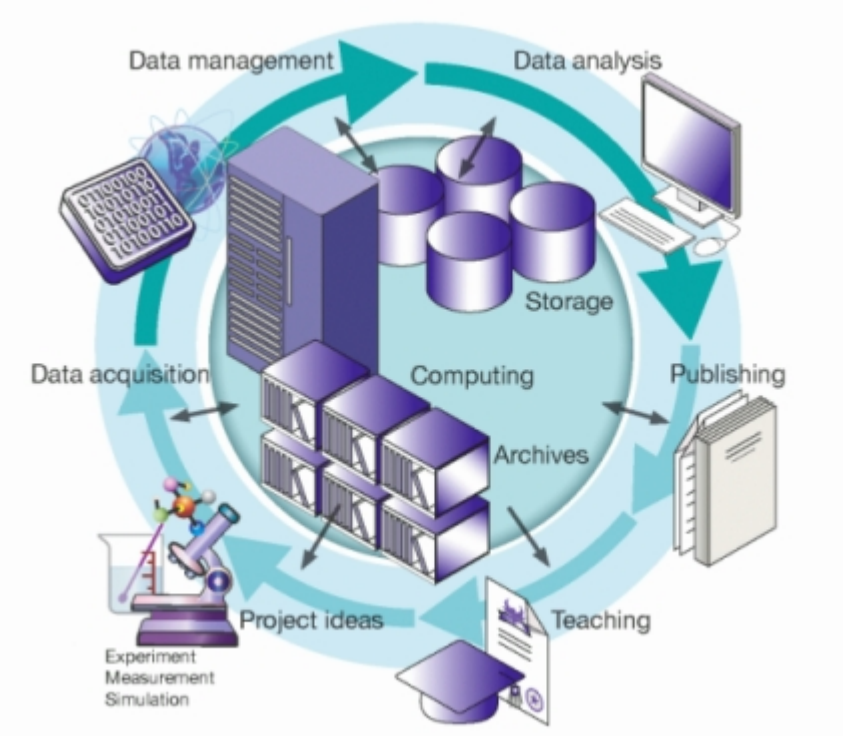
Experiment example





Data Life Cycle

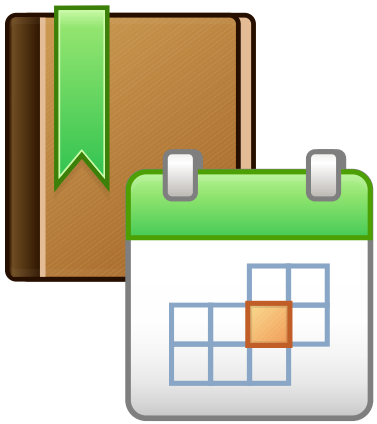
- Apply for an experiment
- Experiment preparation
- Start of the experiment
- Data acquisition
- Activities during the experiment
- Stop of the experiment
- Data access after the experiment
- Data archival



Data Life Cycle

- > Apply for an Experiment
- > Experiment Preparation
- > Start of the Experiment
- > Data Acquisition
- > Activities during the Experiment
- > Stop of the Experiment
- > Data Access after the Experiment
- > Data Archival

- Submit research proposals or experiment applications
- Complete all administrative steps required prior and after the experiment



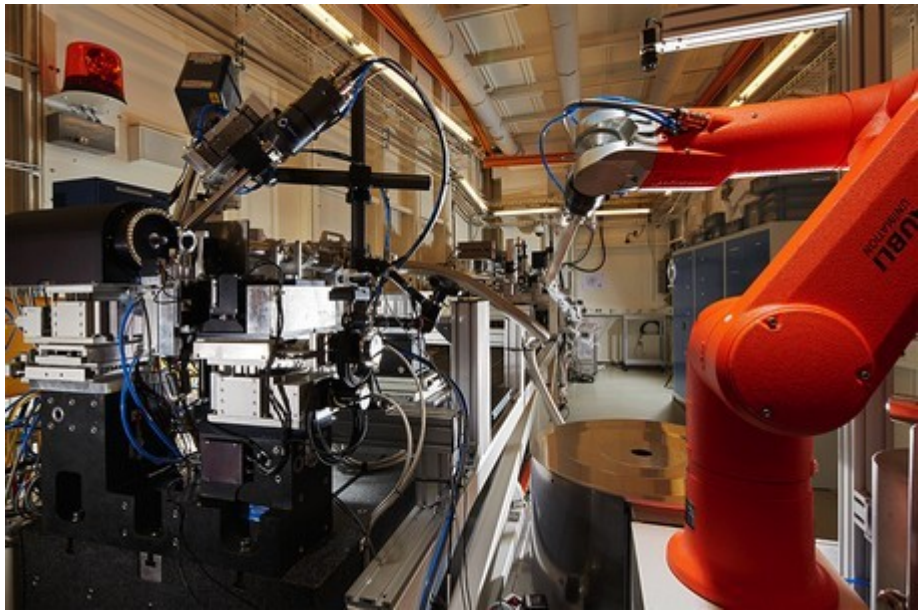
Challenge:

- Photon science community not really aware of computing & storage problem, 100s of small groups, few computing experts, no ecosystem

Data Life Cycle

- Apply for an Experiment
- Experiment Preparation
- Start of the Experiment
- Data Acquisition
- Activities during the Experiment
- Stop of the Experiment
- Data Access after the Experiment
- Data Archival

- Retrofitting of the experiment station
  - Integrate brought in equipment into the facility environment
- Start / stop commissioning



Data Life Cycle

- > Apply for an Experiment
- > Experiment Preparation
- > Start of the Experiment
- > Data Acquisition
- > Activities during the Experiment
- > Stop of the Experiment
- > Data Access after the Experiment
- > Data Archival

System setup:

- Give access to the storage space
- Configure access control for the users
- Configure exports and endpoints

Challenges:

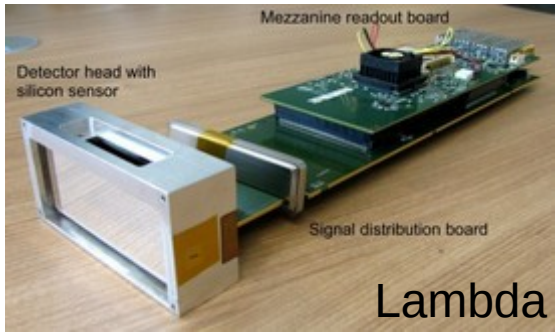
- Lazy account/credential handling
- Mix of OS's at experiment stations – Windows & Linux of all ages and conditions (HW too)
- Black box detector PCs
- Open access vs. safety (brought in equipment even has root)

Detectors

Detector	OS/Access	File size/rate	Bandwidth
Pilatus 300k	Linux (Black box)	1,2 MB Files @ 200 Hz	240 MB/s
Pilatus 6M	Linux (Black box)	25 MB files @ 25 Hz 7 MB files @ 100 Hz	625 MB/s 700 MB/s
PCO Edge	Windows	8 MB files @ 100Hz	800 MB/s
PerkinElmer	Windows	16 MB + 700 Byte files @ 15 Hz	240 MB/s
Lambda	Linux	60 Gb/s @ 2000 Hz	7.5 GB/s
Eiger	Http (Black Box)	30 Gb/s @ 2000 Hz	3.8 GB/s

Pilatus 6M

PCO Edge



Lambda

Data Life Cycle

- > Apply for an Experiment
- > Experiment Preparation
- > Start of the Experiment
- > Data Acquisition
- > Activities during the Experiment
- > Stop of the Experiment
- > Data Access after the Experiment
- > Data Archival

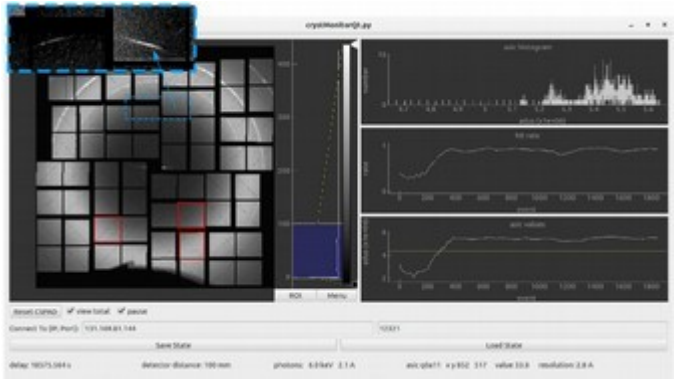
Challenges:

- Data rates of the detectors
  - > High demanding detectors
  - > Multiple detectors in parallel
- Data flow from experiment stations to storage infrastructure (network limitation)
- Data distribution (to storage system, online analysis,…)
- Data reduction (e.g. XFEL expects 50 GB/s of data)
- Wide variety of data formats

Data Life Cycle

- > Apply for an Experiment
- > Experiment Preparation
- > Start of the Experiment
- > Data Acquisition
- > Activities during the Experiment
- > Stop of the Experiment
- > Data Access after the Experiment
- > Data Archival

- System monitoring
  - > Experiment equipment
  - > Storage system
- Data monitoring and analysis
  - > Live view
  - > Hit rate detection
  - > Dark+gain correction
  - > 3D reconstruction



- Parallel data access from outside the facility

DESY

Data Life Cycle

- > Apply for an Experiment
- > Experiment Preparation
- > Start of the Experiment
- > Data Acquisition
- > Activities during the Experiment
- > Stop of the Experiment
- > Data Access after the Experiment
- > Data Archival

Challenges

- Data distribution
- Access data without interfering with data taking
- Online analysis has to run on powerful infrastructure with fast access to the data
- Guarantee data safety even with external access during the experiment

Data Life Cycle

- > Apply for an Experiment
- > Experiment Preparation
- > Start of the Experiment
- > Data Acquisition
- > Activities during the Experiment
- > Stop of the Experiment
- > Data Access after the Experiment
- > Data Archival

- Removal of exports and endpoint

  → Data not accessible for next user group

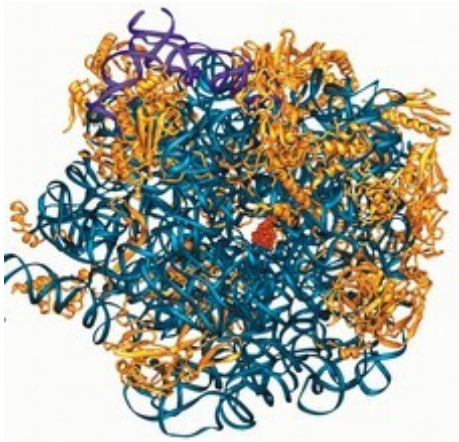
- Extract brought in equipment

  > Remove access to facility infrastructure

Data Access after the Experiment

- > Apply for an Experiment
- > Experiment Preparation
- > Start of the Experiment
- > Data Acquisition
- > Activities during the Experiment
- > Stop of the Experiment
- > Data Access after the Experiment
- > Data Archival

- Offline analysis on- and off-site



Challenges

- Authenticated access to data
- Manage authentication
- Analysis infrastructure on-site
- Brought in analysis software
- Fast data export

Data Life Cycle

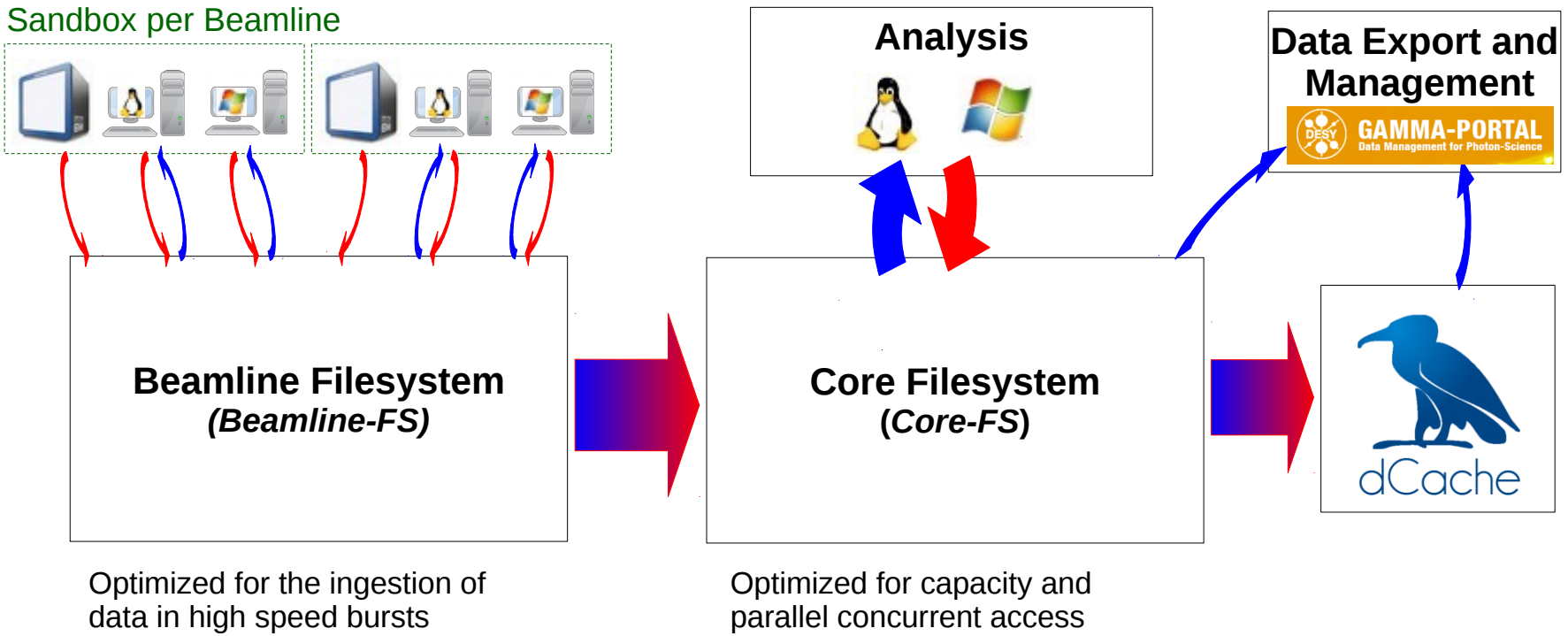
- > Apply for an Experiment
- > Experiment Preparation
- > Start of the Experiment
- > Data Acquisition
- > Activities during the Experiment
- > Stop of the Experiment
- > Data Access after the Experiment
- > Data Archival

- Data copied into long term storage
- Data access after archival

Challenges

- How long preserved?
- Technology
- Access control
- Data export

Example of the realization - ASAP3



> ASAP3 project was developed for Petra3 during the LSDMA portfolio extension of the Helmholtz Association

> Covers all data challenges mentioned on the previous slides

ASAP3 - The User's View of the System

Apply for Beamtime

- Apply for experiment
- Metadata source for experiment

Start Experiment

- Automated system setup
- Create filesets
- Create default directories
- NFSv4 ACL setup
- Initialize Gamma-Portal

Stop Experiment

- Remove fileset on Beamline-FS
- Data not visible for next group
- Fileset on Core- FS remains

Data Archival

- 7 days after stop experiment
- Data copied to dCache
- Copied to tape library for long term storage
- Accessible for user

Exp. Preparation

- Retrofitting of the experiment place
- Start/stop commissioning

Data Acquisition

- Write data to Beamline-FS
- Automatic migration to Core-FS (4 minutes)
- Machine based authentication
- Data distribution with HiDRA

Online Activities

- System monitoring
- Data monitoring
  - OnDa
  - Live-Viewer
- Data access
- Maxwell resources (HPC)
- SMB/NFS access
  - Requires DESY account

Offline Data Access

- Maxwell resources (HPC)
  - CPU and GPU
  - Native GPFS
  - SLURM (partially)
  - DESY account required
- Gamma Portal
  - DOOR account required
- NFS/SMB
  - DESY account required

DESY

Current Status and Outlook

- ASAP3 running successfully for one and a half years
- Flash is currently joining
- Other DESY labs (detector development, microscopy,…) join
- XFEL: similar architecture + components (ASAP3 as blueprint)

→ will become the only system for the DESY light sources and labs

- http://photon-science.desy.de/
- https://www.xfel.eu/
- http://www.helmholtz-lsdma.de/
- https://confluence.desy.de/display/ASAP3/ASAP3++Data+Storage+for+PETRA+III
- Strutz et al (2015) ASAP3 - New Data Taking and Analysis Infrastructure for PETRA III. J. Phys.: Conf. Ser. (JPCS), Volume 664, doi:10.1088/1742-6596/664/00/001001
- Mariani et al (2016) OnDA: online data analysis and feedback for serial X-ray imagingThis article will form part of a virtual special issue of the journal on free-electron laser software. J. Appl. Cryst. 49, 1073-1080. doi:10.1107/S1600576716007469