Advanced software tools and workflows for synchrotron X-ray scattering analysis at ESRF from a user's perspective

Maciej Jankowski

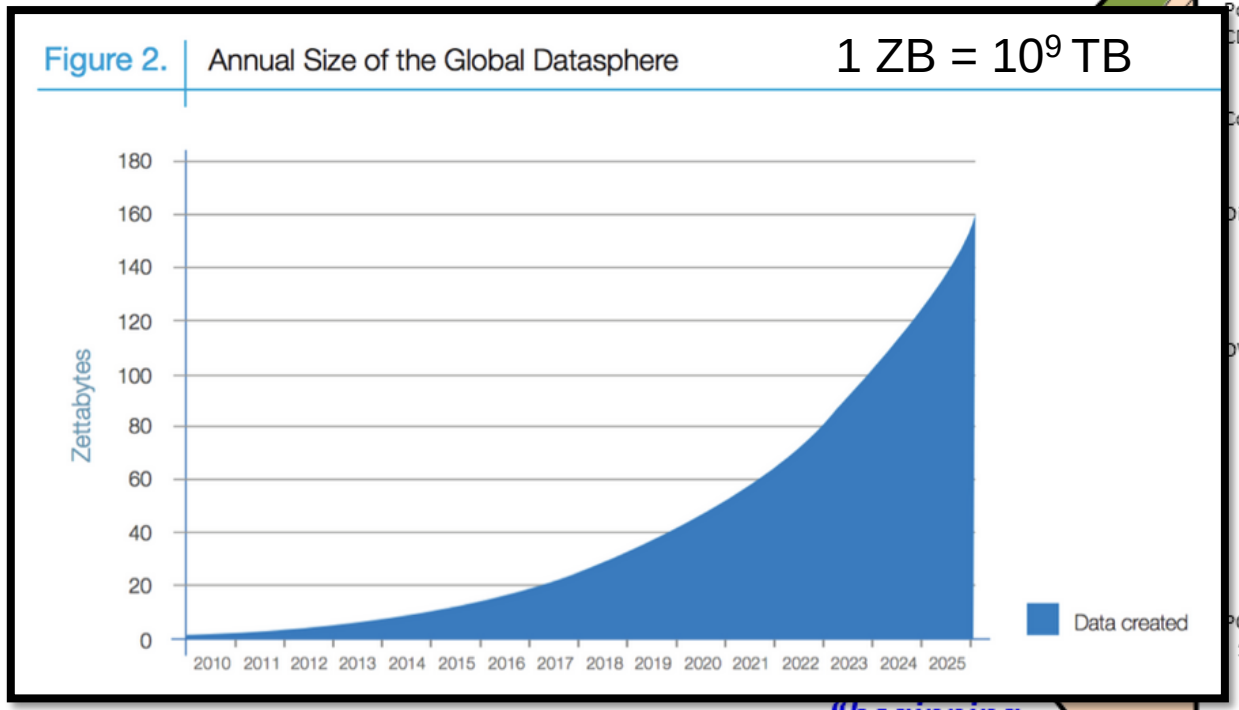
PIONEERING SYNCHROTRON SCIENCE

ESRF
The European Synchrotron

30 years of science

DIGITAL AGE

Global Information Storage Capacity
in optimally compressed bytes



2007 ANALOG
19 exabytes

- Paper, film, audiotape and vinyl: 6 %
- Analog videotapes (VHS, etc): 94 % **ANALOG**

DIGITAL

- Portable media, flash drives: 2 %
- Portable hard disks: 2.4 %
- CDs and minidisks: 6.8 %
- Computer servers and mainframes: 8.9 %
- Digital tape: 11.8 %
- DVD/Blu-ray: 22.8 %

- PC hard disks: 44.5 %
  123 billion gigabytes

"beginning of the digital age"
50%

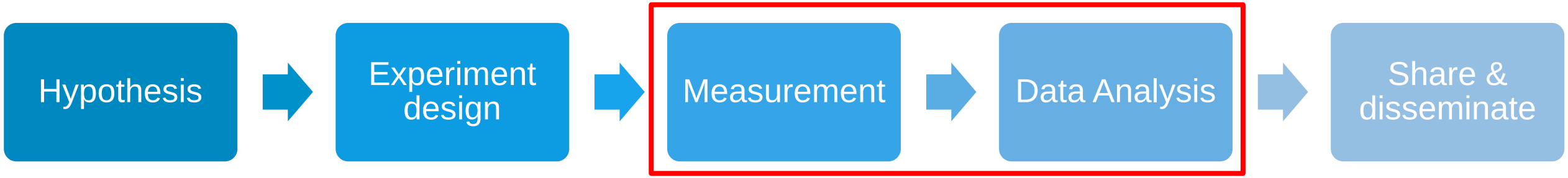
- Others: < 1 % (incl. chip cards, memory cards, floppy disks, mobile phones, PDAs, cameras/camcorders, video games)

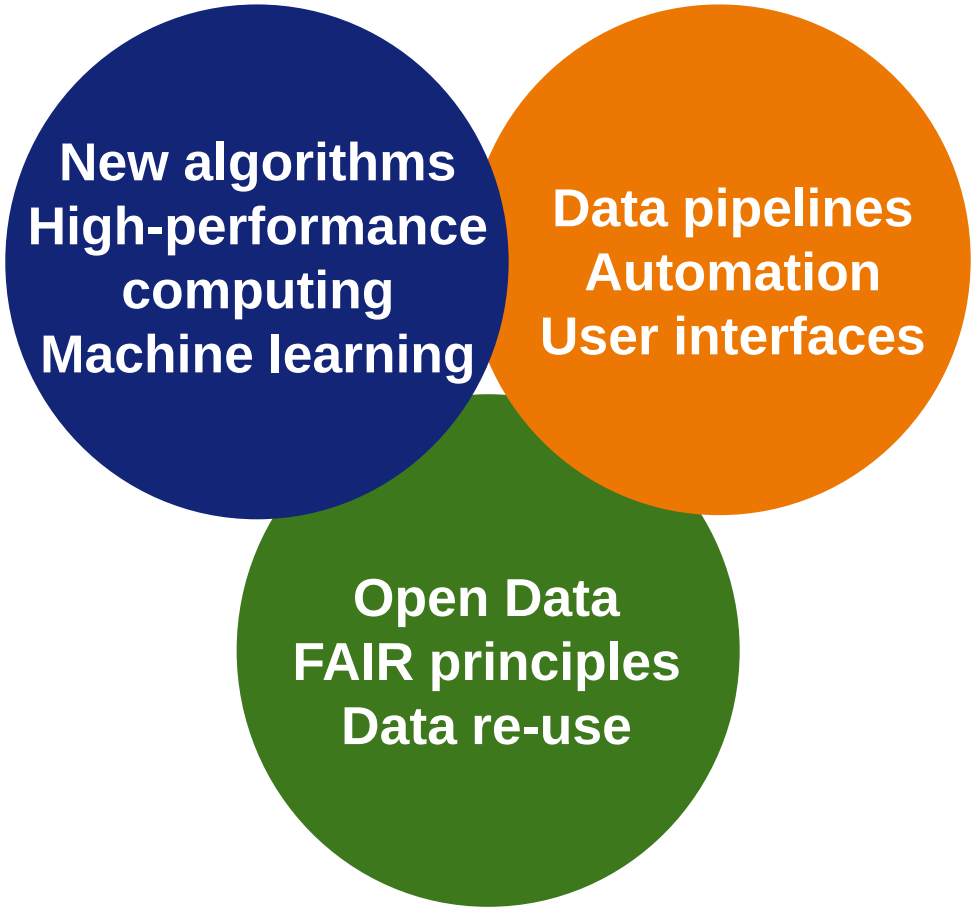
% digital:
1 % **3 %** **25 %** **94 %**

DIGITAL
280 exabytes

Source: Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60 –65. http://www.martinhilbert.net/WorldInfoCapacity.html

RESEARCH LIFECYCLE AND DATA ANALYSIS





3 pillars for (big) data exploitation

Groups involved at ESRF:

- EXP/ Algorithms & scientific Data Analysis (ADA)
- ISDD/Software Data Automation Unit (DAU)
- ISDD/ Beamline Control Unit (BCU)
- TID/IT

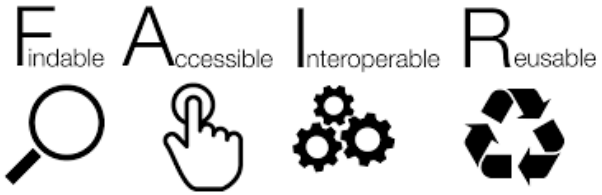
DATA THROUGHPUT AT ESRF AFTER CBS UPGRADE

- **x10-x100 flux for many experiments**
- **Faster detectors (100s of Hz and beyond)**
- **Operando experiments**
- **High-speed/time-resolved measurements**

Detectors for 2024	Max. Acquisition Rate
Pilatus4 4M (ID15A, ID30-A1, ID31, ID30B)	4KHz -> 15GB/s (compressed)
SmartPix 1M (ID31, ID01, ID02, ID09, ID10, ID06-LVP)	6KHz / 6bit -> 6GB/s
Rigaku 1M (ID10)	52KHz / 2bit -> 52 GB/s
Jungfrau (4M - ID29, 1M - ID09)	1kHz -> 8GB/s
Fraunhofer 40000x800 (BM18)	100Hz -> 2.5GB/s
Andor Balor 16M (ID16A)	55Hz -> 1.7 GB/s

BIG DATA

- **Big data**:
  - Some beamlines > 10s TB/day
  - >kHz data collection
  - **Faster data processing**
  - **'Digesting'** a large number of datasets
- **Shift in user community**:
  - Less interested in experimental techniques
  - Focused on results
  - **Need simplified/automated analysis**
- **Smarter use of our facilities**:
  - Automating data collection
  - Less radiation damage
  - Do more with less signal
  - **Open Data**:
    - Truly FAIR data (re-usable)



NICE CLUSTER- ESRF COMPUTING CLUSTER

In 90’ the NICE cluster had ~3GB of storage: 1 PT = 10^6 GB

NICE stands for **Networked Interactive Computing Environment:**

- 161 Nodes (computing units)
- CPUs: Intel Xeon, AMD EPYC 9454- 48 Cores, EPYC 9654- 96 Cores, Power9
- NVIDIA GPUs: V100, A40, A100, H100
- Linux system
- Supports VISA, Jupyter notebooks (Python), CVMFS
- A few tens of PB storage
- Tape storage for archiving

Slurm is an open-source, fault-tolerant, and highly scalable cluster management and job scheduling system for large and small Linux clusters.

- Allocates access to computer nodes for users
- Framework for starting, executing, and monitoring jobs on nodes
- Assigning resources and managing a queue of pending work

Cluster and beamline computers can be accessed by remote desktop and web browser

remote.esrf.fr

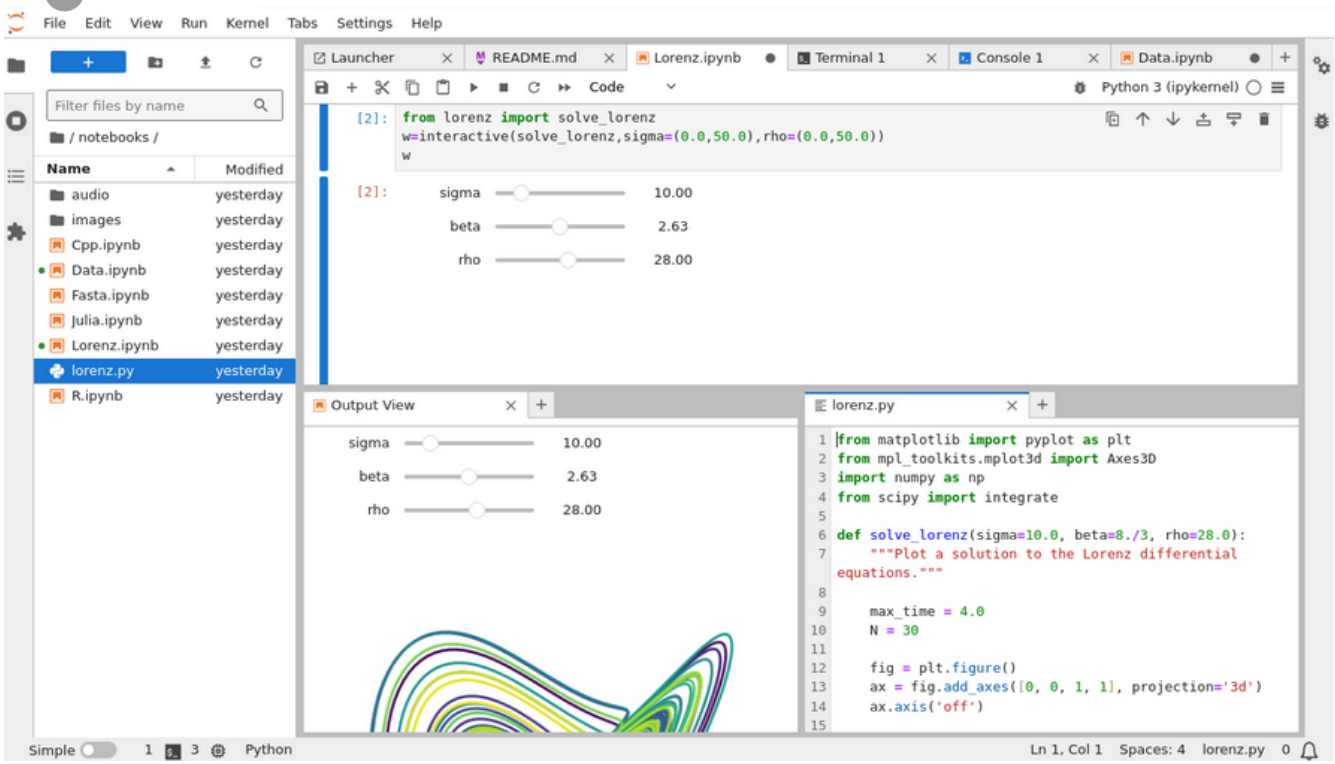
Ubuntu jupyter VISA CernVM File system slurm workload manager

COMPUTING ON CLUSTER

visa.esrf.fr

VISA (Virtual Infrastructure for Scientific Analysis) makes it simple to create compute instances (a kind of private virtual computer) on the data analysis infrastructure to analyze your experimental data using just your web browser

jupyter-slurm.esrf.fr



CernVM File system

APPTAINER

The **CernVM** File System is a software distribution service that allows users to load modules (software packages, containers (apptainer), Python environments,…) on beamline computers or ESRF clusters: ready-to-use software, no need for installation.

JupyterLab is the latest web-based interactive development environment (Python) for notebooks, code, and data.

EXAMPLE visa.esrf.fr

VISA

Home

Support

Help

Sign out

Maciej's compute instances

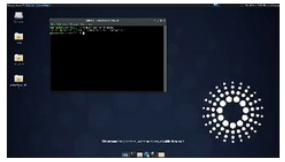
CREATE A NEW INSTANCE

Filter instances by experiment...

My instances 1 | Instances shared with me 0

nebula_seeder

active



Desktop (latest)

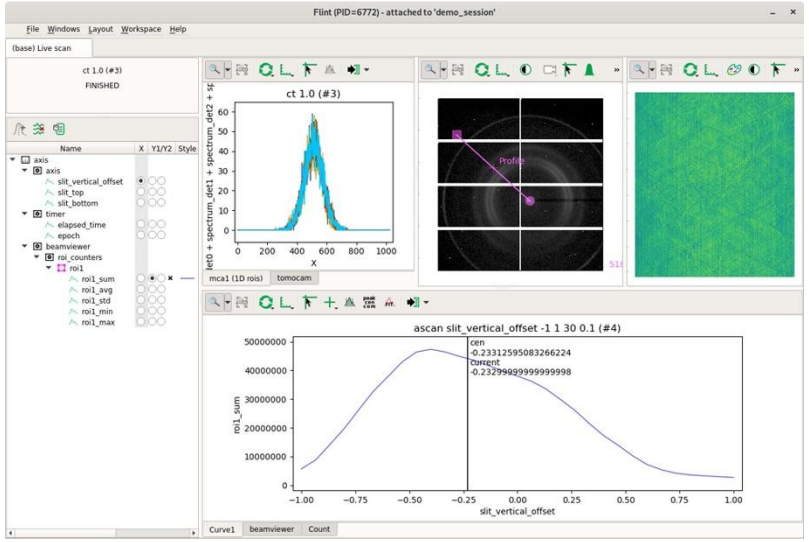
4 GB · 4 VCPUs

Instance **2958** created on **10 Sept 2024** and due to expire on **9 Nov 2024**

Settings | Connect

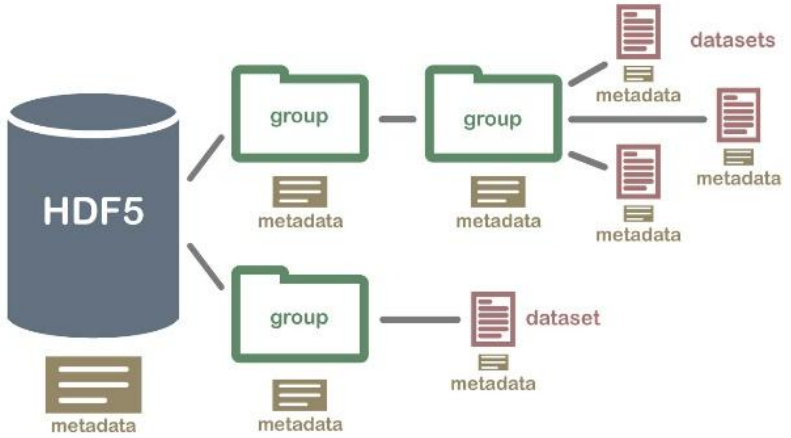
DATA FORMAT AND COTROL SOFTWARE

BLISS is the new ESRF control system for running experiments, with full deployment for the EBS upgrade program. It has replaced SPEC, but the syntax is very similar, and the programming language is Python.



Flint is a graphical application companion for BLISS for online data visualization.

Synchrotron Radiation News, *36*(6), 12–19.



The Hierarchical Data Format version 5 (**HDF5**) is an open-source file format that supports large, complex, heterogeneous data. **Main format of ESRF**

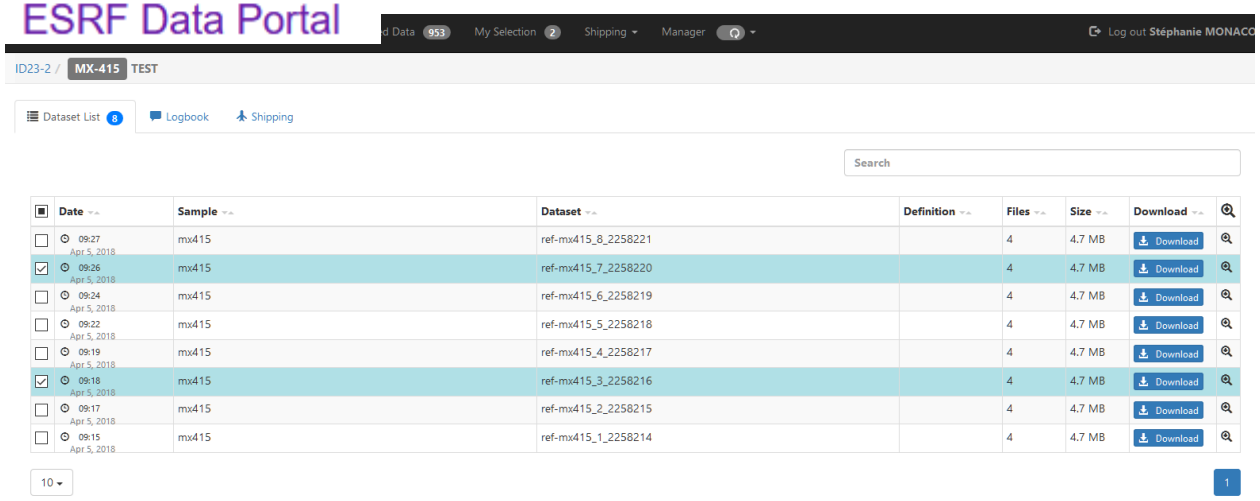
Collette, A. (2013). *Python and HDF5*. O'Reilly Media, Inc.

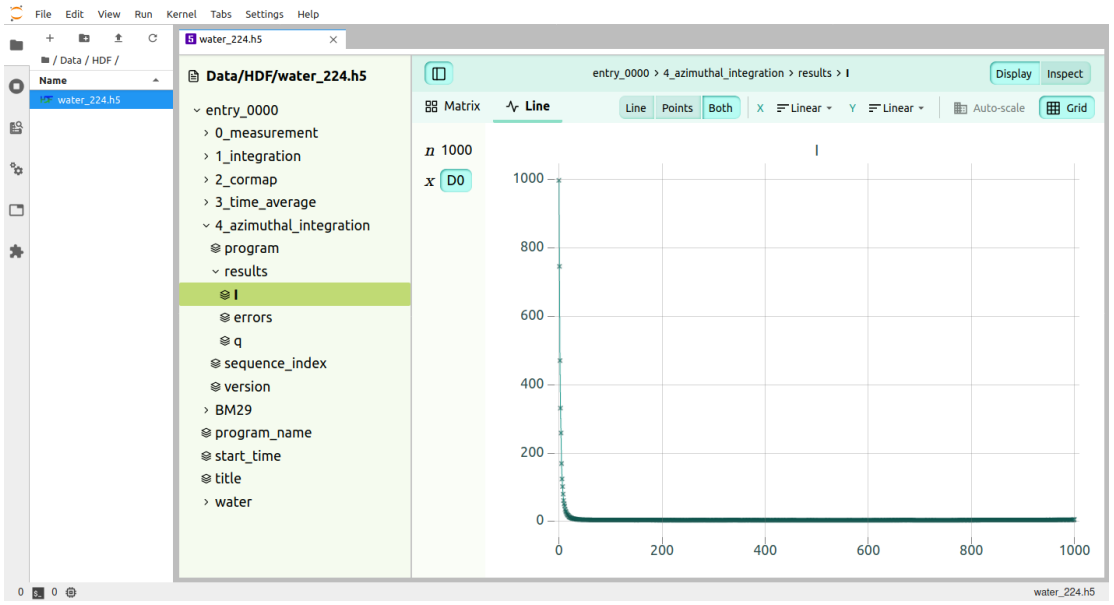
NeXus is an effort by an international group of scientists motivated to define a common data exchange format for neutron, X-ray, and muon experiments. NeXus is built on top of the scientific data format HDF5 and adds domain-specific rules for organizing data within HDF5 files in addition to a dictionary of well-defined domain-specific field names.

J. Appl. Cryst. (2015). **48**, 301-305

DATA BROWSING AND VISUALIZATION

ESRF Data Portal
https://data.esrf.fr





h5Web

SILX aims to provide a collection of Python packages to support the development of data assessment, reduction and analysis at synchrotron radiation facilities. It intends to provide reading/writing tools for different file formats, data reduction routines, and a set of Qt widgets to browse and visualize data.

silx.io
- HDF5, SPEC files, other image formats

silx.math:
- histogramming, fitting, median filtering

silx.gui:
- 1D and 2D visualization tools, data viewer and browser

silx.opencl:
- Image alignment, image processing, tomography tools

EWOKS- WORKFLOW SYSTEM MANAGEMENT

EWOKS

The main goal of Ewoks is to automate data processing and experiments at large-scale facilities and make data processing more reproducible.

DOI 10.5281/zenodo.12704398

EWOKS is fully integrated with BLISS for real-time data analysis during acquisition

ESRF Data Portal
https://data.esrf.fr

ESRF Data Portal

The workflows can be executed from the data portal which allows post experiment analysis and reprocessing

It can be installed locally on PC or cluster

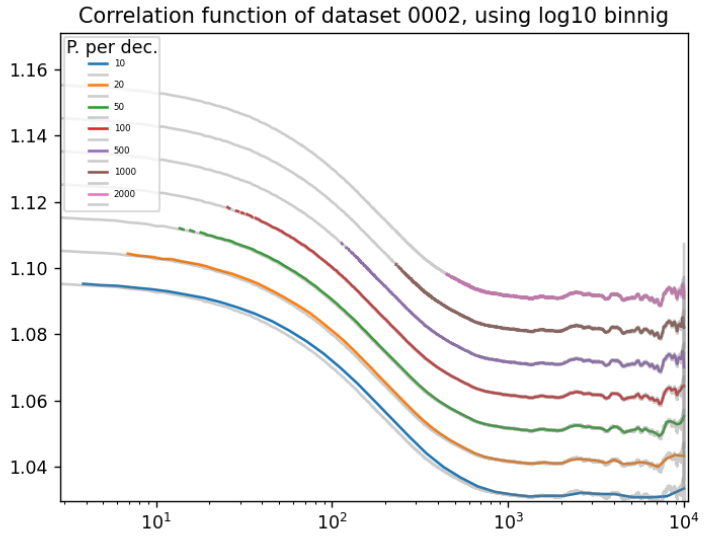
SOFTWARE FOR SCATTERING

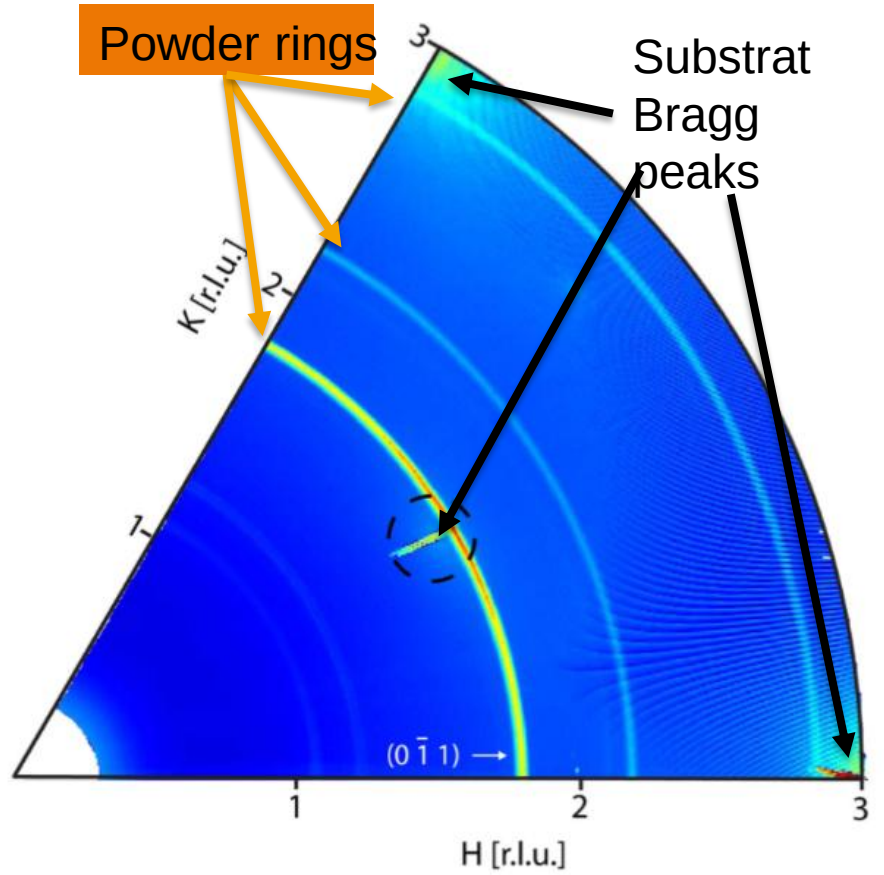
Free SAS

Small angle scattering tools ... but unlike most others, free and written in Python (3.6+) The FreeSAS tool suite is licensed under the MIT license.

New XPCS correlator- https://gitlab.esrf.fr/xpcs/correlator

- Sparse data format
- Compatible with Lima2
- Fast calculation algorithms using GPU
- In-fly data binning for massive datasets





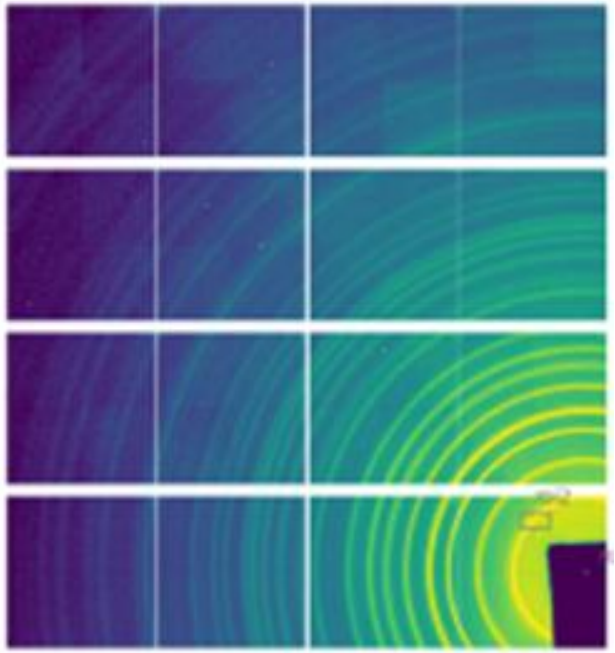
Bismuth (111) polycrystalline thin film on Al_2O_3 (0001), surface unit cell, integrated from ~100k images (80 Gb of data), using fast scan mode

BINoculars

J. Appl. Cryst. 48, (2015).

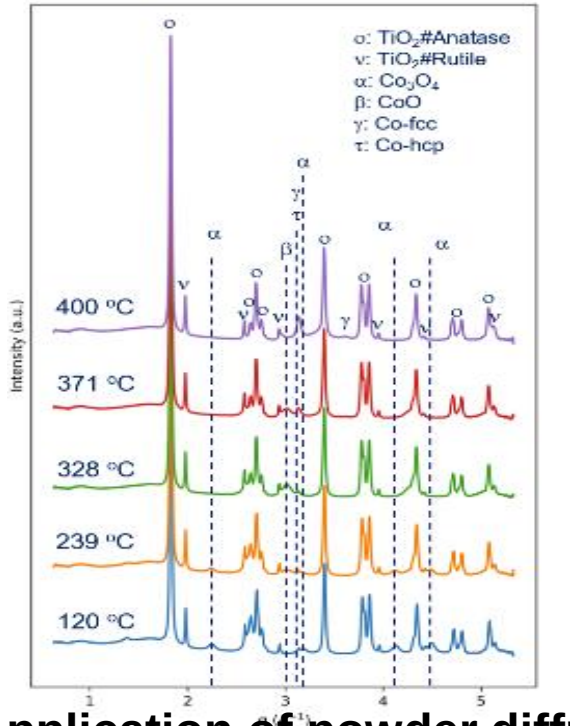
PYFAI- AZIMUTHAL INTEGRATOR OF 2D DETECTOR SIGNAL

2D detector image of Cobalt NPs



Fast transformation

Integrated powder pattern



Application of small-angle scattering

- Micro/nano-scale structure
- Particle shape
- Protein domains
- Protein folding
- Colloids
- Fiber orientation

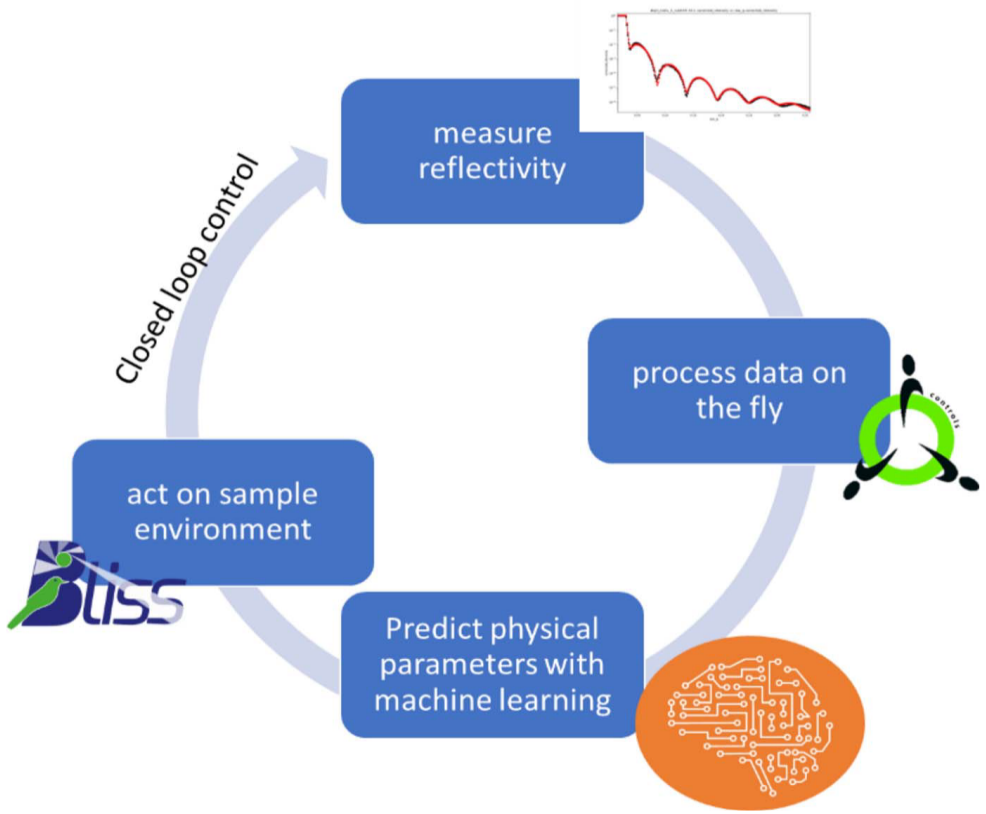
Application of powder diffraction:

- Phase identification (mapping)
- Crystallinity
- Lattice parameters
- Thermal expansion
- Phase transition
- Crystal structure
- Strain and crystallite size

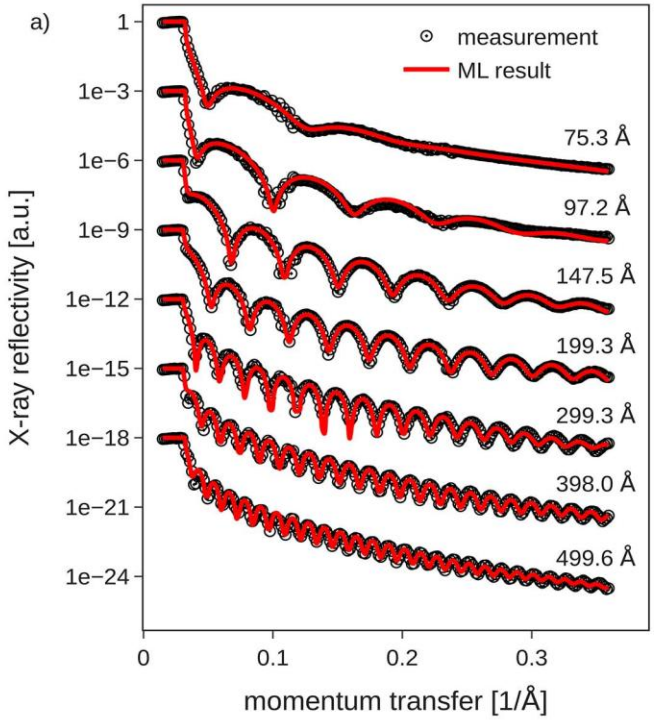
J. Kieffer 2013 J. Phys.: Conf. Ser. 425 202012

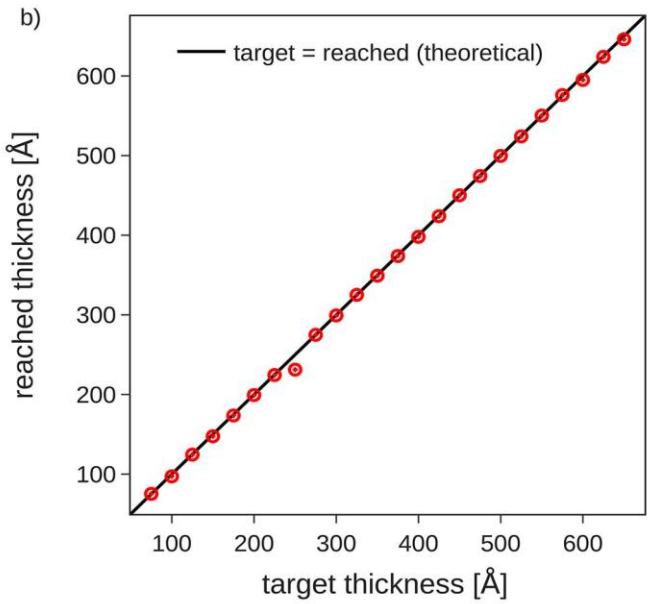
The European Synchrotron | ESRF

XRR CLOSED LOOP CONTROL EXPERIMENT USING ML



Using ML to tune/trigger actions during an experiment: live processing of data on a dedicated server for ML inference.

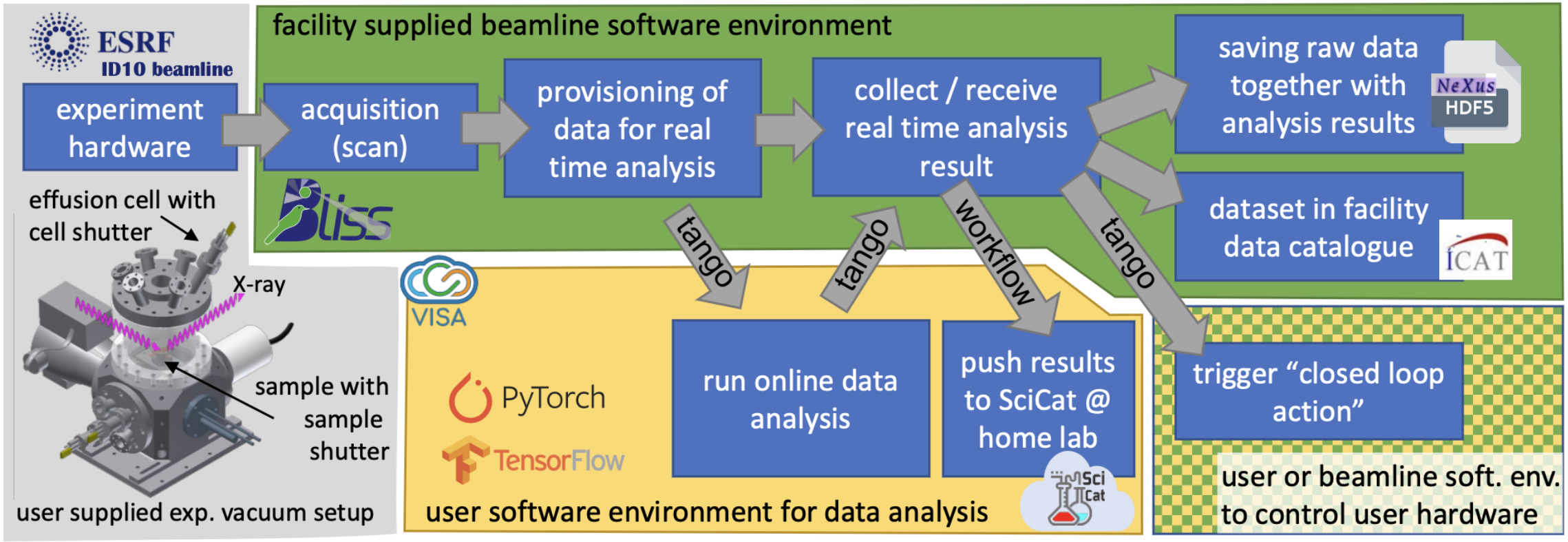




ML can be used to automatically fit reflectivity data.
A good example of ML application is the non-linearity of the reflectivity vs the model, which can make the fit complicated.

Pithan, L., Starostin, V., Mareček, D., Petersdorf, L., Völter, C., Munteanu, V., Jankowski, M., Konovalov, O., Gerlach, A., Hinderhofer, A. and Murphy, B., 2023. **Closing the loop: autonomous experiments enabled by machine-learning-based online data analysis in synchrotron beamline environments**. *Journal of synchrotron radiation*, *30*(6).

EXPERIMENT WORKFLOW AT ID10 ESRF- TOWARD AUTONOMOUS BEAMLINES



Reflectivity @ ID10 general workflow

THANK YOU FOR YOUR ATTENTION

GIANT INNOVATION CAMPUS

ESRF The European Synchrotron

cea · cnrs · EMBL · ESRF The European Synchrotron · GRENOBLE ECOLE DE MANAGEMENT · Grenoble INP | UGA · NEUTRONS FOR SOCIETY · UGA Université Grenoble Alpes · The European Synchrotron | ESRF

FREE-SAS

- **supycomb:** a re-implementation of supcomb as described in J. Appl. Cryst. (2001). 34, 33-41
- **autorg**: automatic guess of the Guinier region in a SAXS curve, based on BioXTAS-RAW
- **auto_guinier:** automatic guess of the Guinier region in a SAXS curve, based on J. Appl. Cryst. (2007). 40, s223-s228
- **auto_gpa**: automatic Guinier peak analysis, based on J Appl Cryst. (2016). 49, 1412–1419.
- **bift**: an Bayesian inverse Fourier transform, based on J. Appl. Cryst. (2006). 39, 797-804 & BioXtas-RAW
- **cormap:** a tool to assess the similarity of saxs curves, based on Nature Methods volume 12, pages 419–422 (2015)