HDF5 technology and NeXus data format usage at Diamond Light Source

Peter Chang

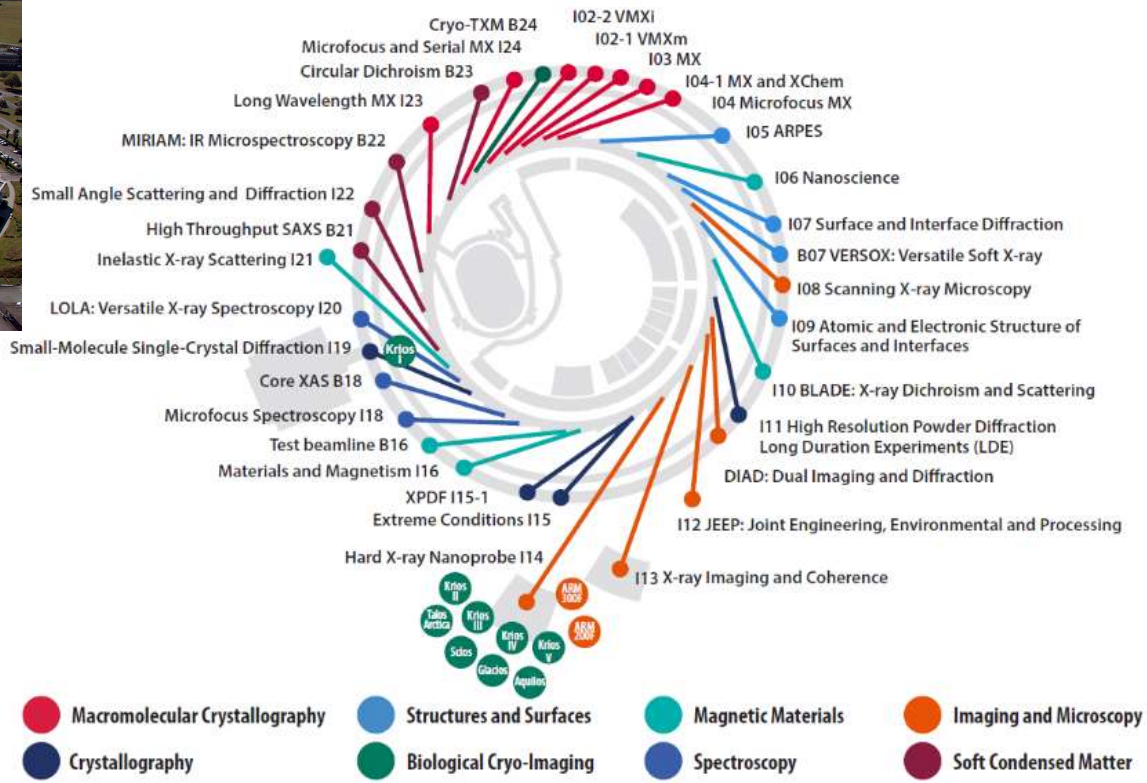
HDF5 workshop

ESRF, Grenoble

17th September 2019

diamond

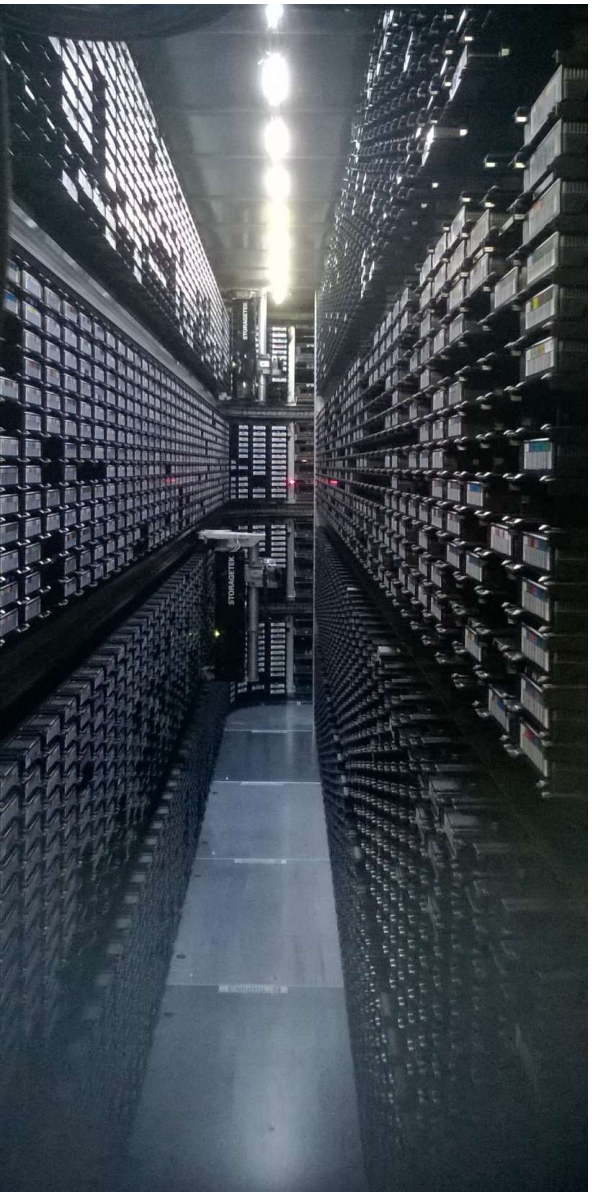
About Diamond



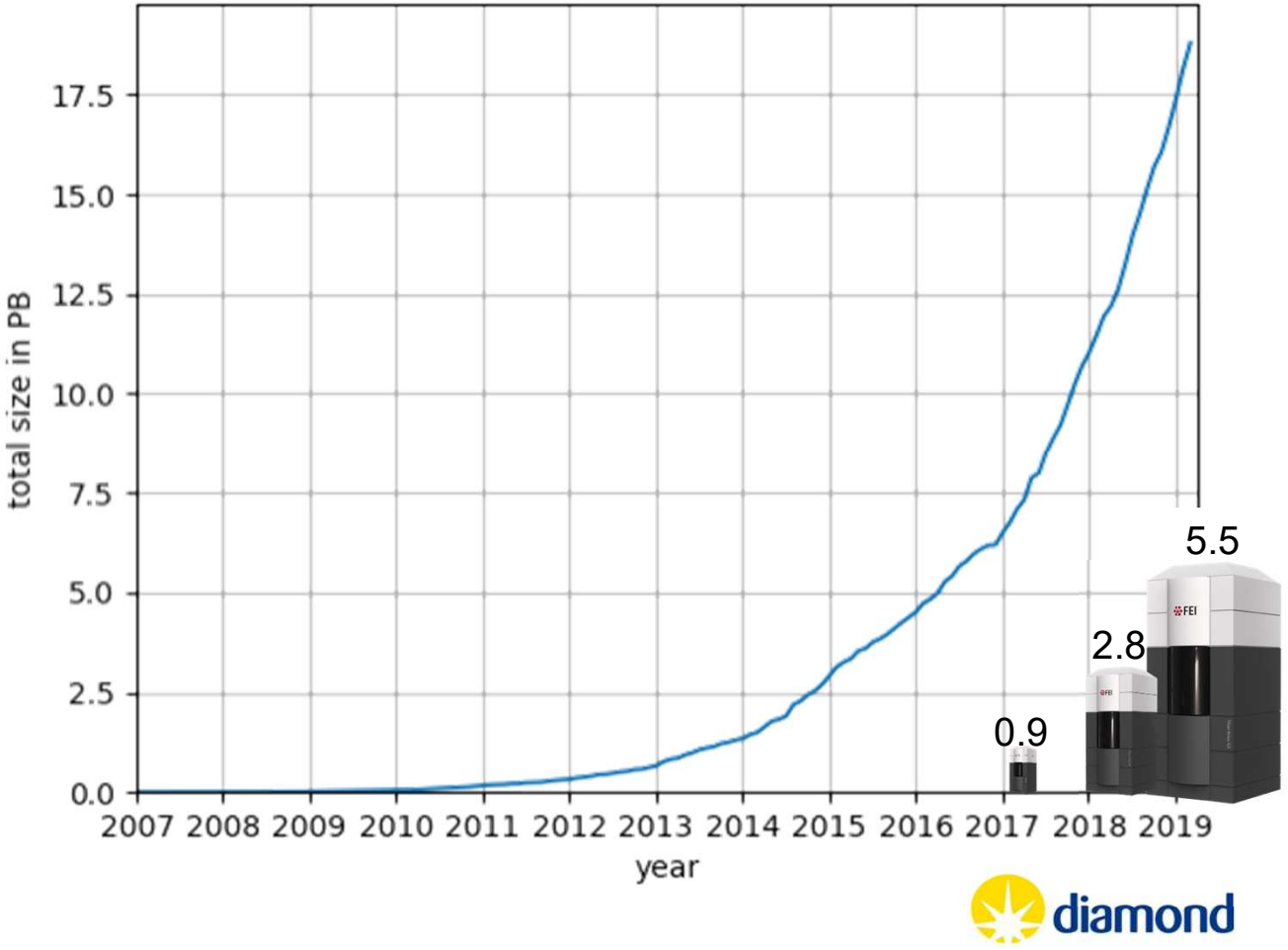
Year to March 2019

30+ beamlines
11 electron microscopes
12000+ experimental shifts
6300+ visiting users
4400+ remote users

diamond



Total data archived from Diamond



diamond

Recent Motivation/Drivers/Priority

- Increased data volumes
- Live visualisation
- Increased automation

diamond

Data Acquisition with HDF5

Detectors

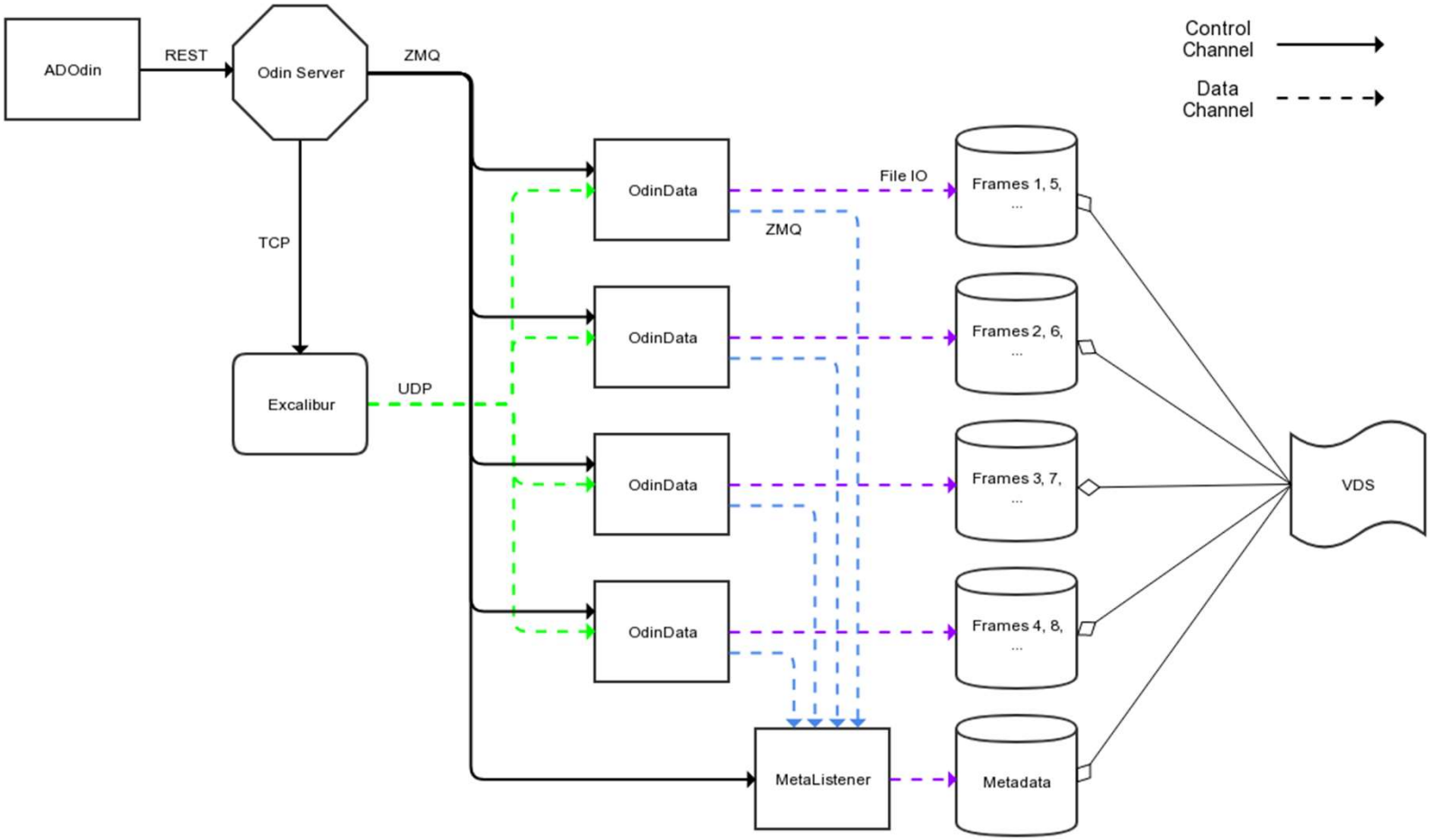
- Excalibur 3M (~250Hz)
- Eiger2 4M, 16M (~500Hz)
- Tristan 10M event data

Odin framework (DLS/STFC)

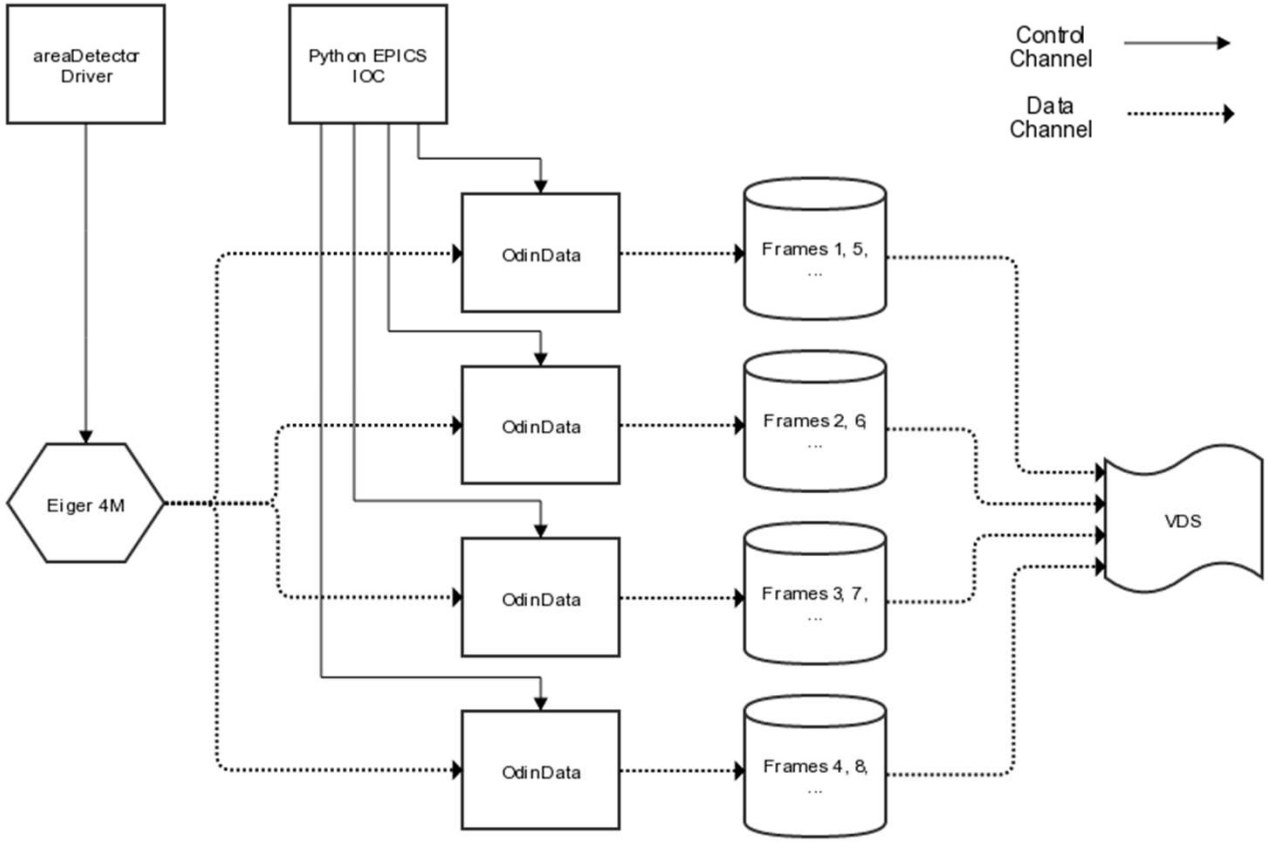
- Control via EPICS areaDetector
- Fan-out to multiple OdinData nodes for light processing and writing HDF5 files

diamond

Excalibur


ADOdin
REST
Odin Server
ZMQ
TCP
Excalibur
UDP
OdinData
OdinData
OdinData
OdinData
File IO
ZMQ
Frames 1, 5, ...
Frames 2, 6, ...
Frames 3, 7, ...
Frames 4, 8, ...
VDS
MetaListener
Metadata
Control Channel
Data Channel

Eiger 4M


areaDetector Driver
Python EPICS IOC
Control Channel
Data Channel
OdinData
OdinData
OdinData
OdinData
Frames 1, 5, ...
Frames 2, 6, ...
Frames 3, 7, ...
Frames 4, 8, ...
Eiger 4M
VDS

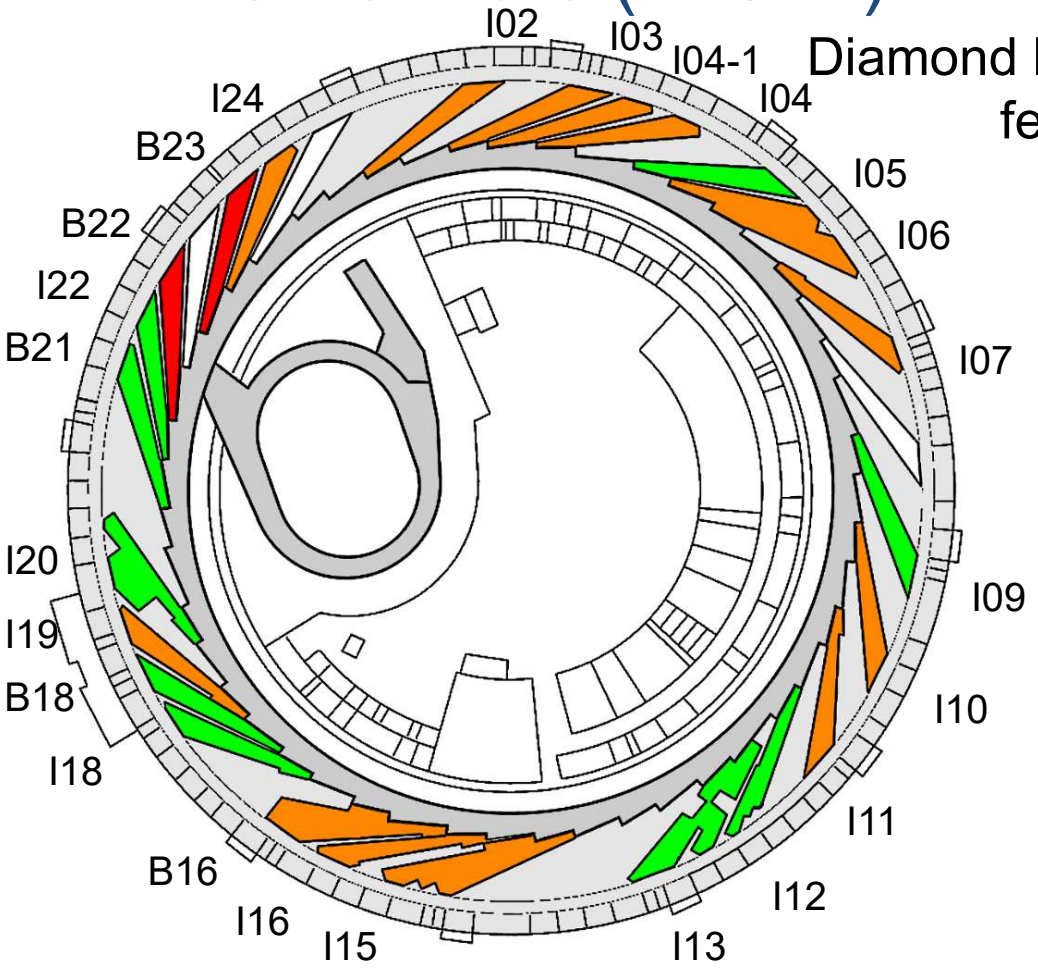
Odin FrameProcessor

Framewriter plugin uses features of HDF5

- Compression
- Direct chunk writing
- VDS
- SWMR

diamond

File Formats (>2012)



Diamond has a policy of, where feasible, to standardise on file formats, the choice being NeXus/HDF5

Green: predominantly using NeXus.

Orange: Mixed NeXus and other formats or considering NeXus in the next 12 months.

Files can be generated by Detector, EPICS or Data Acquisition

Data formats

38 beamlines currently active:

- 28 write NeXus
- 10 = 3 CBF, 3 TIFF, 3 proprietary, 1 ASCII
- 13/28 are grid scan/mapping beamlines so write better NeXus

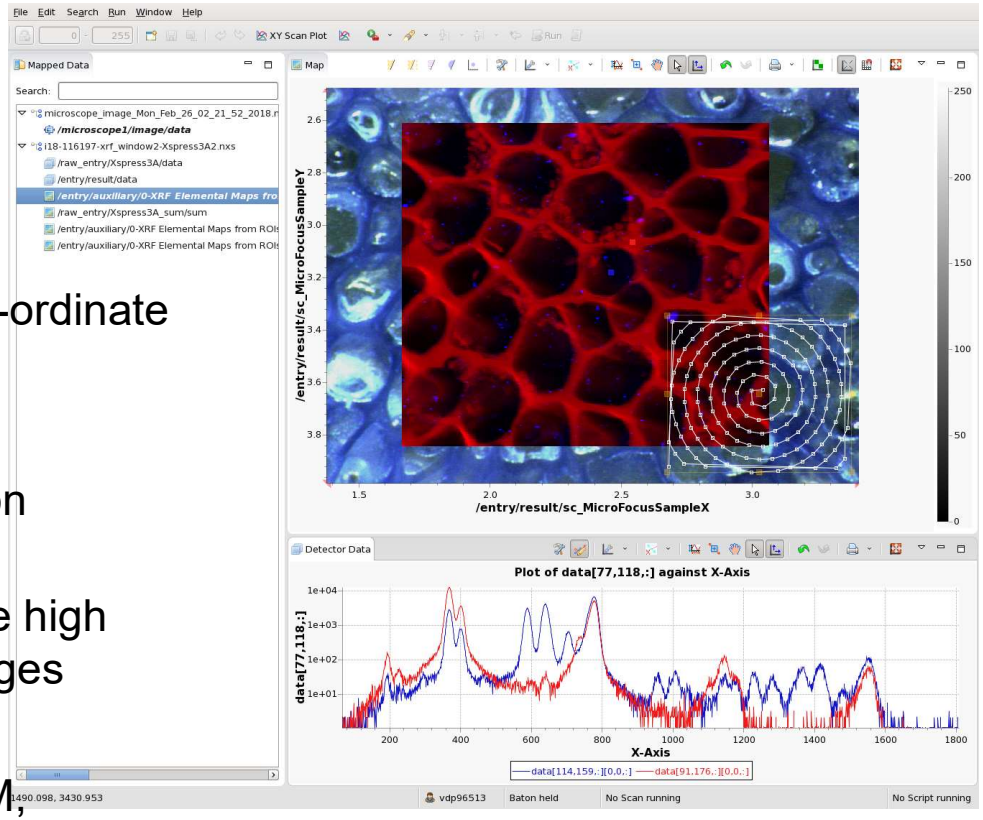
Improve quality of metadata by including application definitions with Nexus template engine

- New NXentry groups that comprise links to dynamic datasets
- Additional static datasets to supplement definitions

diamond

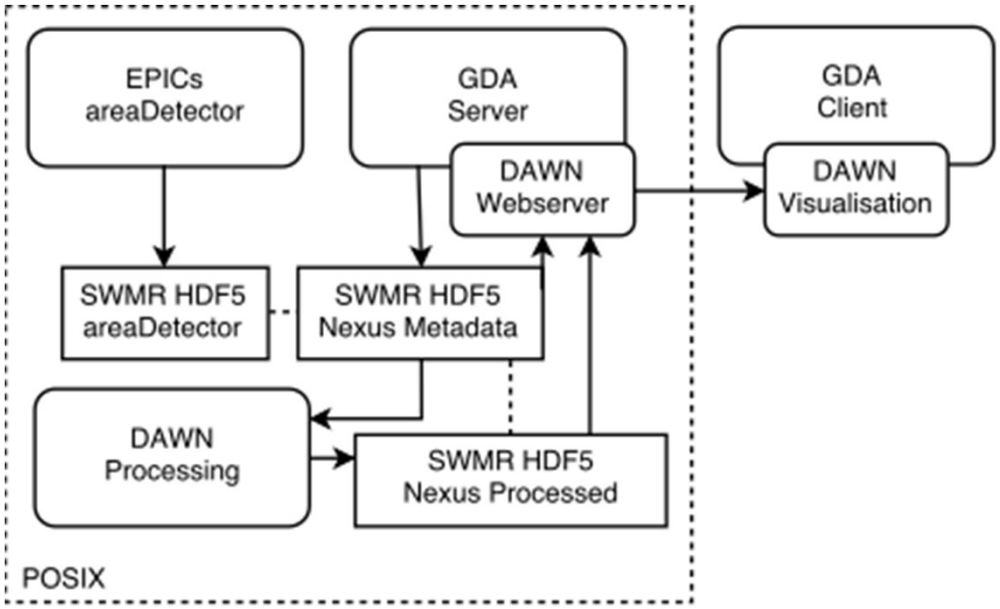
GDA Live Grid Scan Visualisation

- HDF5 SWMR allows visualisation during scan
  - Raw and processed data
- Nexus tagging used to identify:
  - Appropriate data for visualisation
  - Dimensions of data to visualise in sample co-ordinate
  - Complex scan trajectories – spirals
  - RGB images from optical microscopes
- Since sample stage co-ordinate used visualisation independent of scan resolution
  - Overlay – Coarse sample location scans, fine high resolution scans and optical microscope images
- Required tags independent of experiment
  - Same UI used for – XRF, XRD, SAXS, STXM, Ptychography, ARPES, FTIR…
  - Consistent sample stage coordinates.
  - No application definition.



Live Grid Scan Visualisation - SWMR

- Not so simple as “SWMR lets you read the data as its being written”
- Has to be across POSIX compliant mounts
- Cluster and control machines all GPFS – SWMR works
- Visualisation clients mount file system as NFS – SWMR doesn’t work
- Webserver on control machine sends requested visualisation datasets (small) back to clients



Dawn processing

DAWN analysis workbench has visualization and automated processing for many X-ray techniques

- SAXS, WAXS
- XRF, XRD, XANES
- Reciprocal space remapping

These rely on NeXus application definitions

Other analysis programs

Tomographic reconstruction and processing pipeline

- Uses parallel HDF5
- Reads application definitions (NXtomo, NXfluoro)
- Translates non-NeXus with yaml-dictionaries
- Has plugins for PyFAI, PyMCA, TomoPy, Astra, etc

Python framework for ptychography

- Uses NXcxi_ptycho (adapted from CXI - coherent x-ray imaging standard on HDF5)

diamond

ctd…

Macromolecular crystallography (MX)

- DIALS/Xia2 can use NXmx application definition
- SWMR and VDS not used as HDF5 v1.10 not supported by many third-party software
- Dectris's master.h5 links the multiple .h5 files written by their Eiger detector software as well as by OdinData

HDF5 and NeXus issues

Problems and blockers to further adoption:

- Third-party analysis software
- Multi-threaded reads not supported
- MPI-I/O tuning difficulties on cluster filesystems
- Chunk size optimization
- Hyperslab reading (with non-unit strides/steps) can be quicker when broken down to smaller slabs (chunk cache?)
- VDS reports of slow reads
- SWMR NFS issues

diamond

Future

NeXus working group

- Add more metadata
- Populate ISPyB
- FAIR data access

NeXus/HDF5 usage

- Adding 3D shadow masks for goniometer arms
- More data acquisitions will need VDS. Examples: non-contiguous rotation scans in MX to equalize radiation damage, logical mapping of images from ptychography scans

diamond

Thanks

- Gary Yendell, Ulrich Pedersen – Controls group
- DLS NeXus working group led by Steve Collins
- Jake Filik, Nicola Wadeson, Aaron Parson, Graeme Winter, and Alun Ashton – Scientific Software team

diamond