Savu - Tomography Reconstruction and Processing Pipeline

Dr Mark Basham

Dr Nicola Wadeson

Data Analysis Group

Diamond Light Source


Diamond Light Source

Diamonds Aim

- To be a world-leading user facility engaged in synchrotron research and innovation
- To maximise the scientific, economic and societal impact of Diamond
- To ensure the long-term sustainability of Diamond as a national facility
- To engage and inspire the general public through promoting an understanding of and enthusiasm for science
- To continuously plan for Diamond's technical and scientific future

diamond

Diamonds Aim

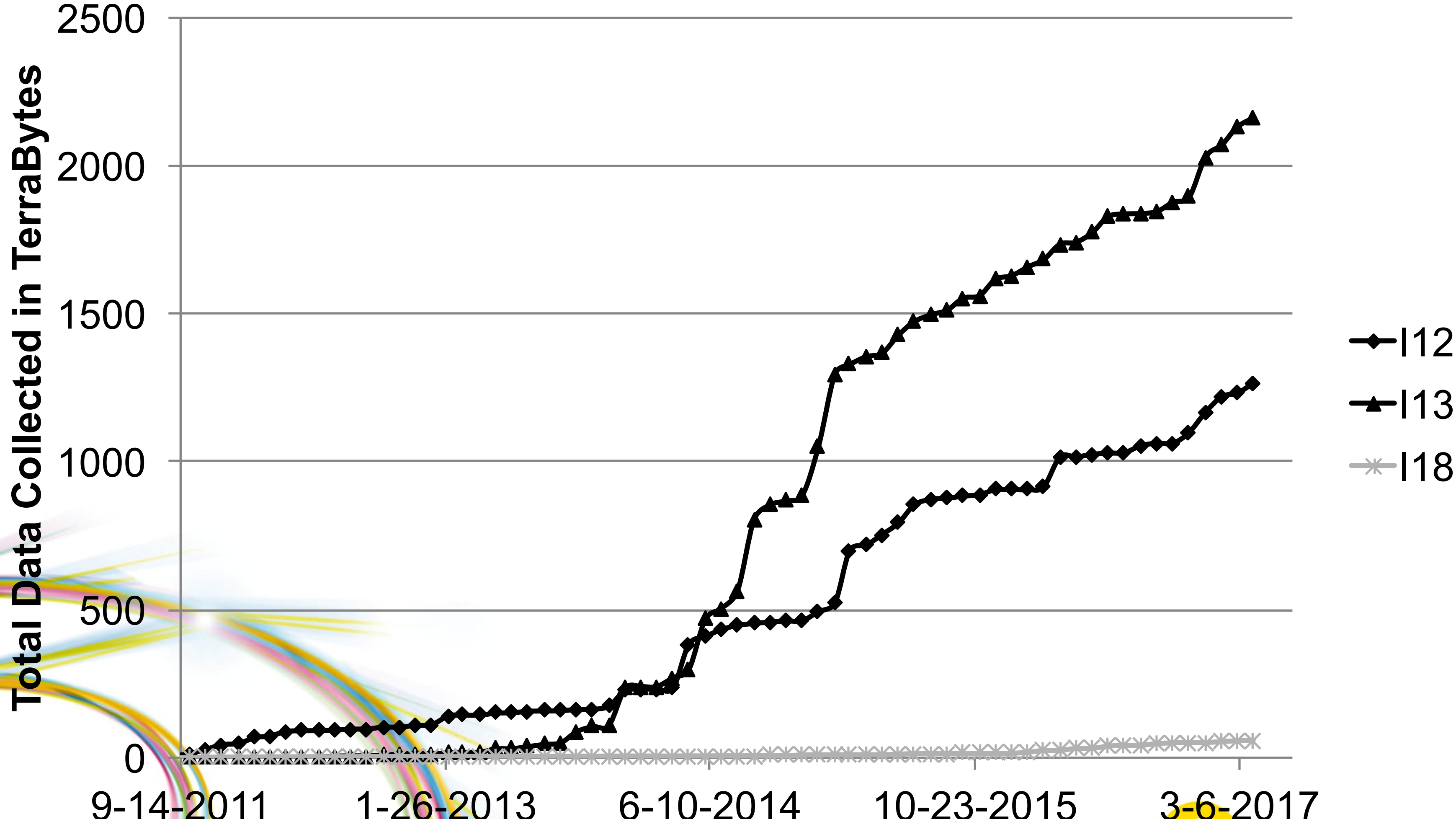
- To be a world-leading user facility engaged in synchrotron research and innovation
- **To maximise the scientific, economic and societal impact of Diamond**
- **To ensure the long-term sustainability of Diamond as a national facility**
- To engage and inspire the general public through promoting an understanding of and enthusiasm for science
- **To continuously plan for Diamond's technical and scientific future**

diamond

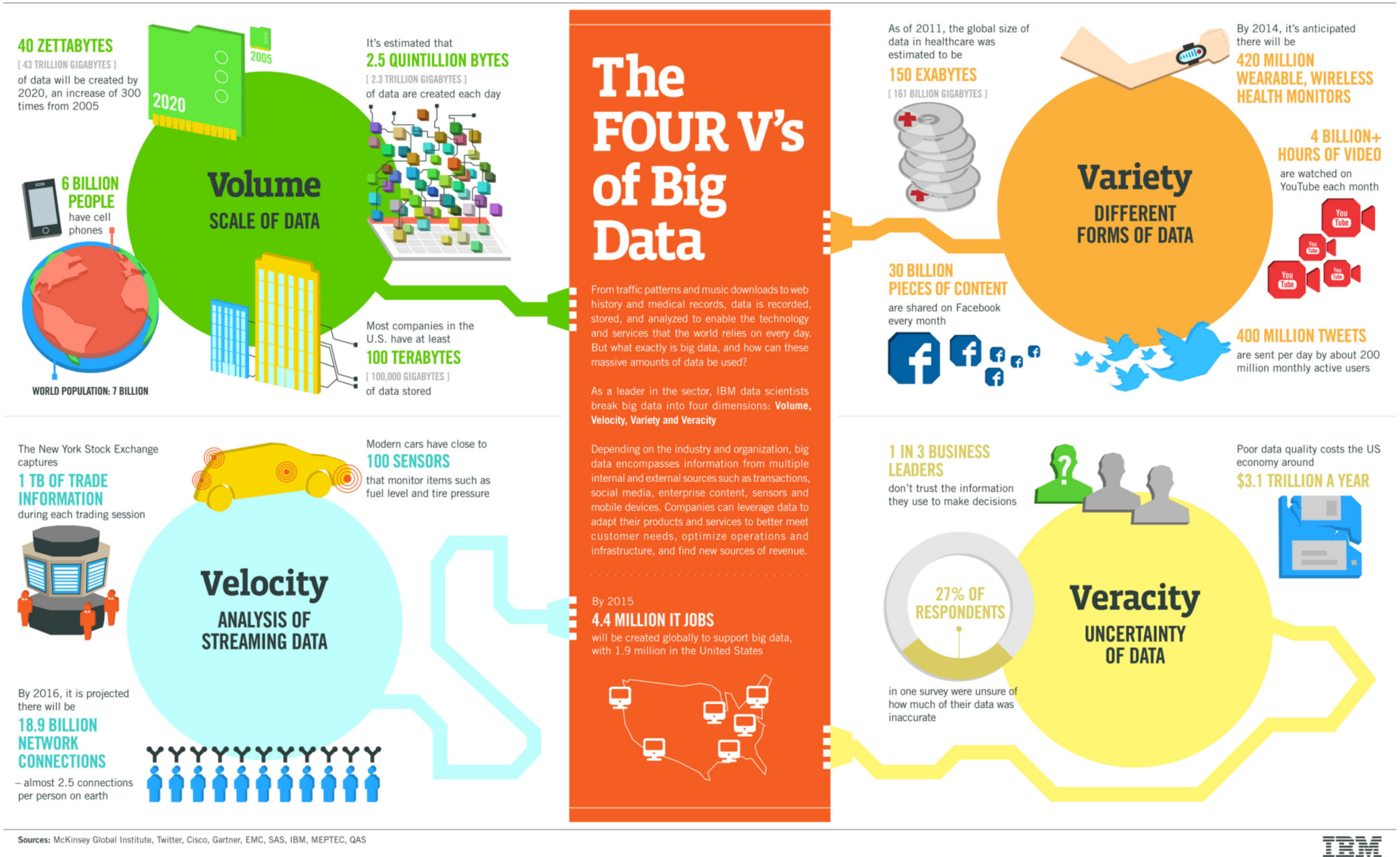
Data Analysis Aims

- Our users should be able to make the most of their beamtime.
- There should be no reason for the creation of “Dark Data”

The complexity – this is now a big data problem



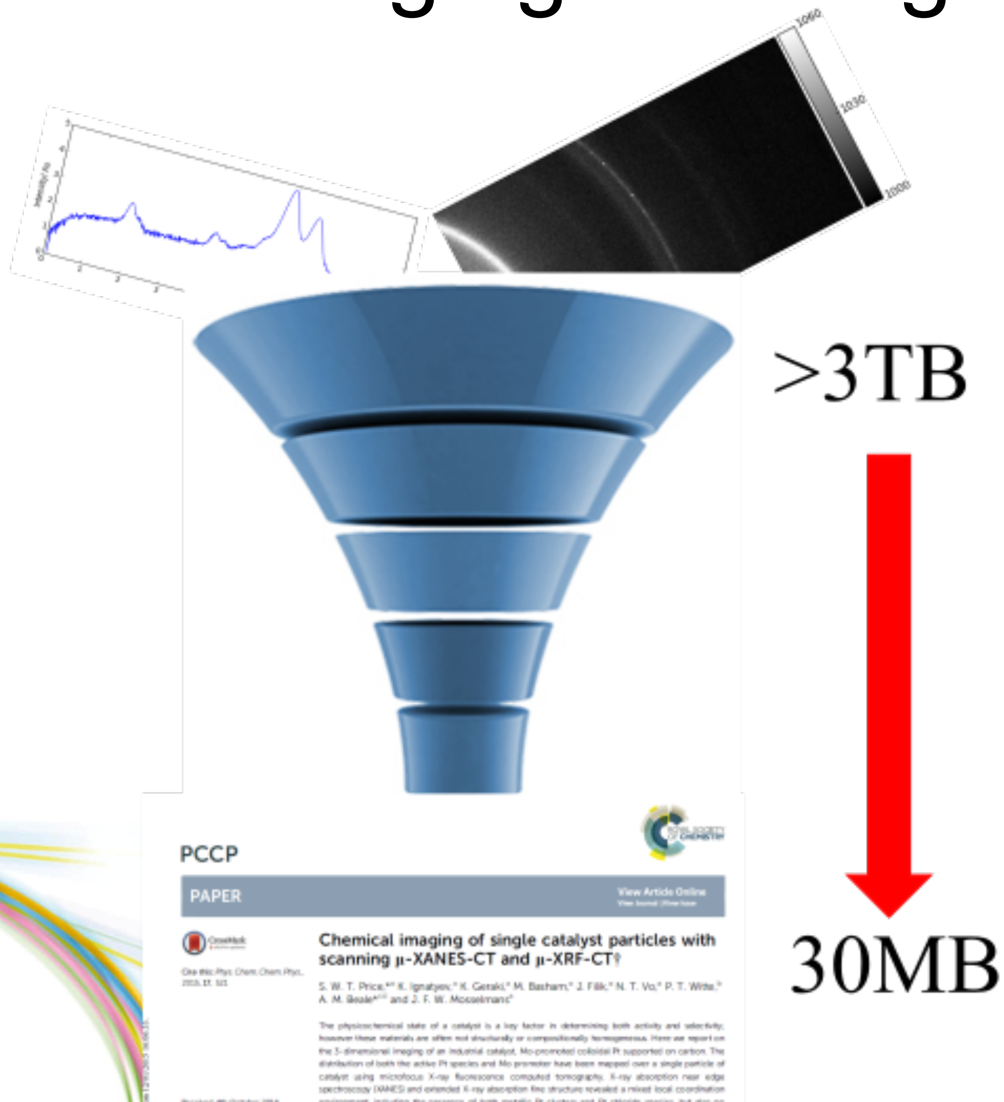
The complexity – this is now a big data problem



Accelerating factors

- Faster Data Collection (Less down time per experiment)
  - 4D studies
  - Mapping Project
  - Sample Automation
- Faster Detectors (~Data rate doubling every 9 months)
  - Increased per tile performance
  - More tiles per detector
- Source Upgrade projects(i.e. Diamond 2).
  - 10x brightness.

The emerging challenge.

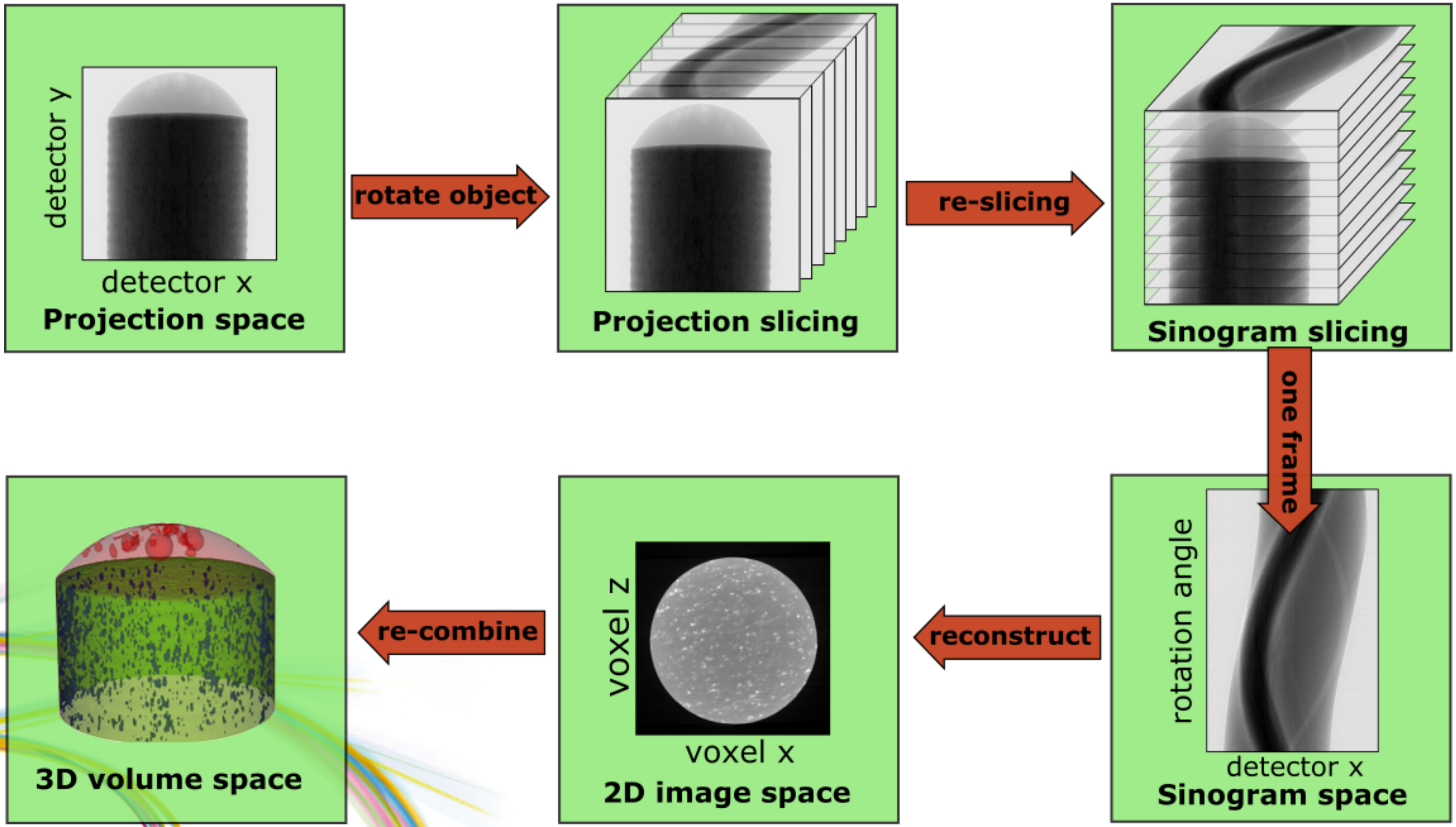


diamond

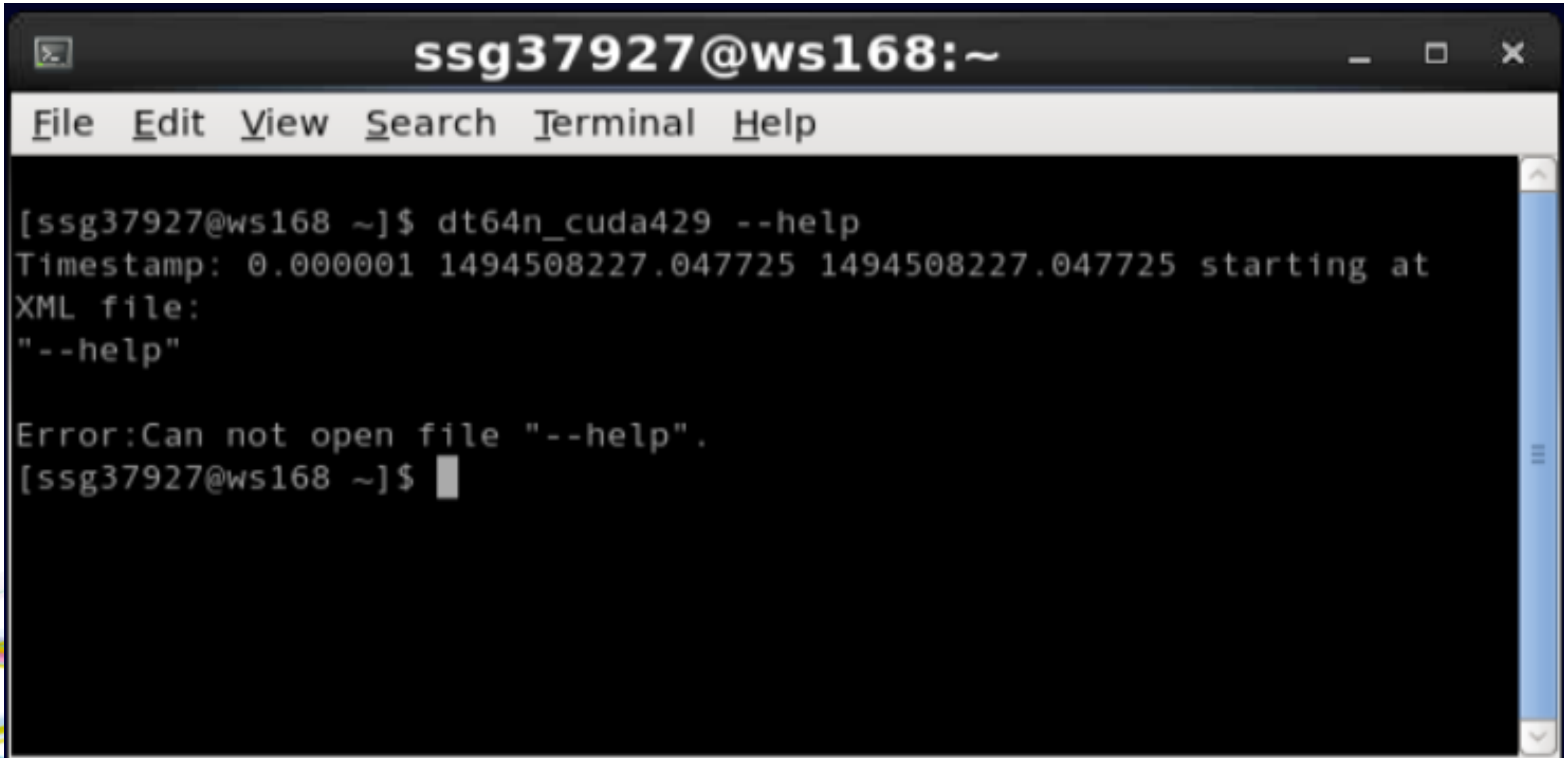
How do we address this?

- Not really an emerging issue, it has emerged, and we are starting to try to address it.
- But it wasn’t 10 years ago, what have we done to change over this time to try to adapt.

Full-field tomography data processing



In the beginning (~2008)



diamond

User Experience!

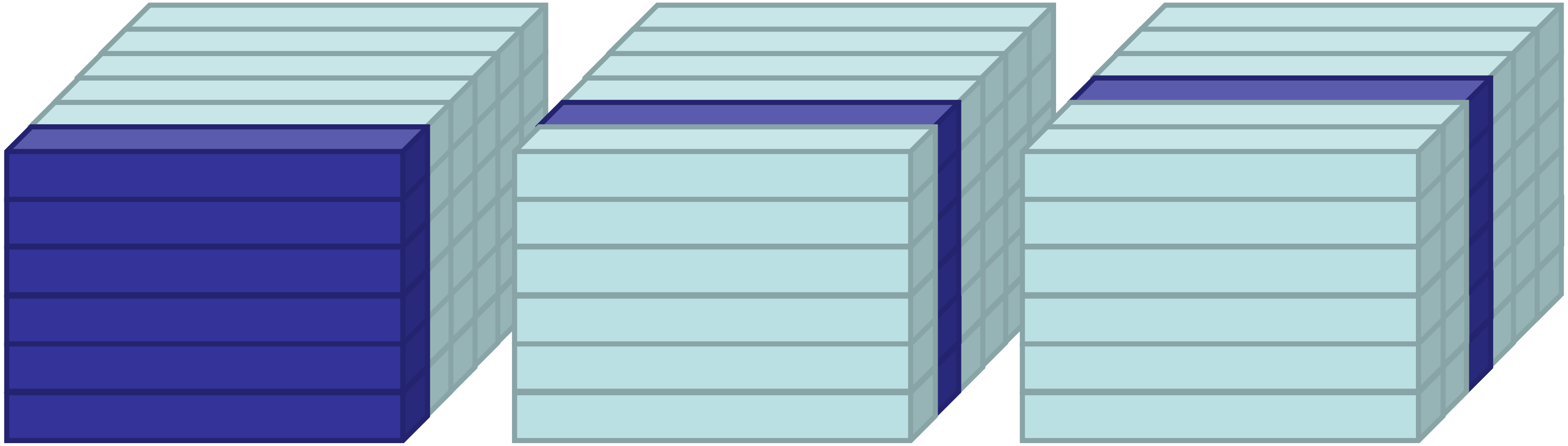
- Costly conversion to sinograms
- Black box experience
- Difficult to use, command line only with XML (Not really XML) which had many foibles.
- Single Machine, single GPU implementation.

Step 1 – Remove / parallelize the Sinogram generation step.

The Sinogram Assumption:

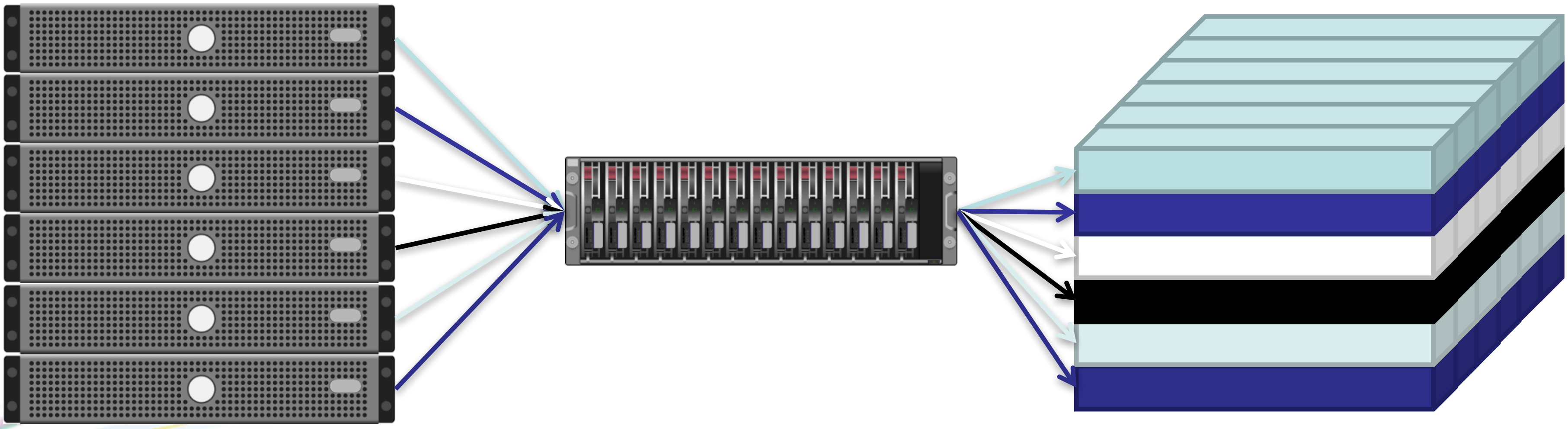
“Tomographic Reconstructions of parallel beam data can be pleasingly parallelised on a per sinogram basis”

Projections to Sinograms (HDF5)



Projections to Sinograms (HDF5)

Parallel Processing of 1 HDF file in Singogram space.



Writing HDF5 Files in the first place.

Area Detector

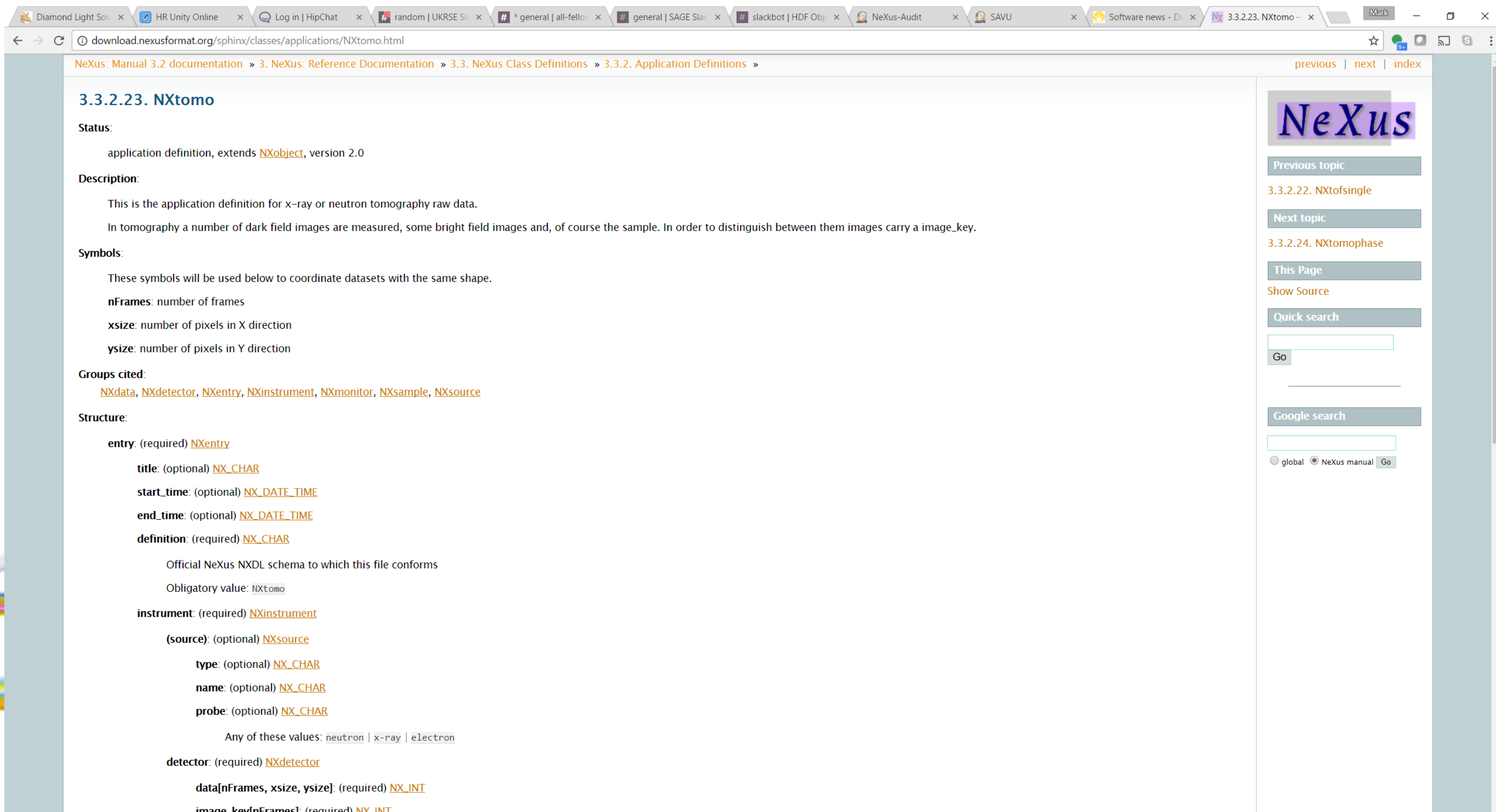
Issues with Buffers

- Area Detector Buffer
- Camera Buffer

- Darkfield, Flatfield and Projection data
- Required flushing the buffer at every step
  - (Not a problem now, but it was in ~2010)

- Solution
  - Collect as one block, with a key, telling you what each frame is.

HDF5 and NeXus



NeXus: Manual 3.2 documentation » 3. NeXus: Reference Documentation » 3.3. NeXus Class Definitions » 3.3.2. Application Definitions » previous | next | index

3.3.2.23. NXtomo

Status:

application definition, extends NXobject, version 2.0

Description:

This is the application definition for x-ray or neutron tomography raw data.

In tomography a number of dark field images are measured, some bright field images and, of course the sample. In order to distinguish between them images carry a image_key.

Symbols:

These symbols will be used below to coordinate datasets with the same shape.

nFrames: number of frames

xsize: number of pixels in X direction

ysize: number of pixels in Y direction

Groups cited:

NXdata, NXdetector, NXentry, NXinstrument, NXmonitor, NXsample, NXsource

Structure:

- **entry**: (required) NXentry
  - **title**: (optional) NX_CHAR
  - **start_time**: (optional) NX_DATE_TIME
  - **end_time**: (optional) NX_DATE_TIME
  - **definition**: (required) NX_CHAR
    - Official NeXus NXDL schema to which this file conforms
    - Obligatory value: `NXtomo`
  - **instrument**: (required) NXinstrument
    - **(source)**: (optional) NXsource
      - **type**: (optional) NX_CHAR
      - **name**: (optional) NX_CHAR
      - **probe**: (optional) NX_CHAR
        - Any of these values: `neutron` | `x-ray` | `electron`
    - **detector**: (required) NXdetector
      - **data[nFrames, xsize, ysize]**: (required) NX_INT
      - **image_key[nFrames]**: (required) NX_INT

NeXus

Previous topic

3.3.2.22. NXtofsingle

Next topic

3.3.2.24. NXtomophase

This Page

Show Source

Quick search

Go

Google search

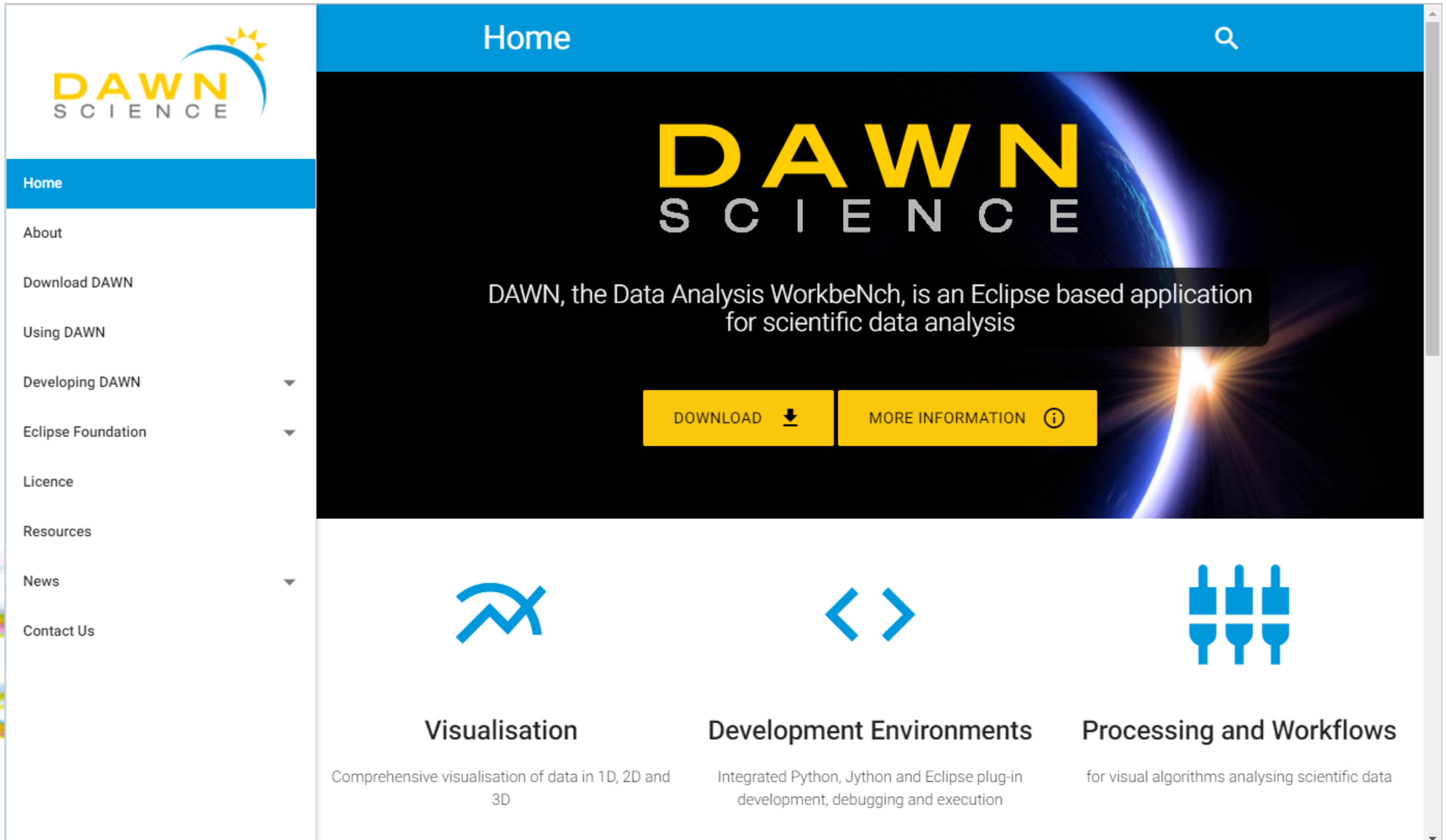
global NeXus manual Go

Step 2 :Drive forward parallelisation and user experience

- Tomo-recon.py
  - Command line wrapper around dt64n
    - Handles all the XML
  - Runs across the whole SGE cluster + multiple cards
    - Extracts data from NeXus files
      - Writes local tiff images, already dark and flatfield corrected
    - Outputs a stack of Tiff images

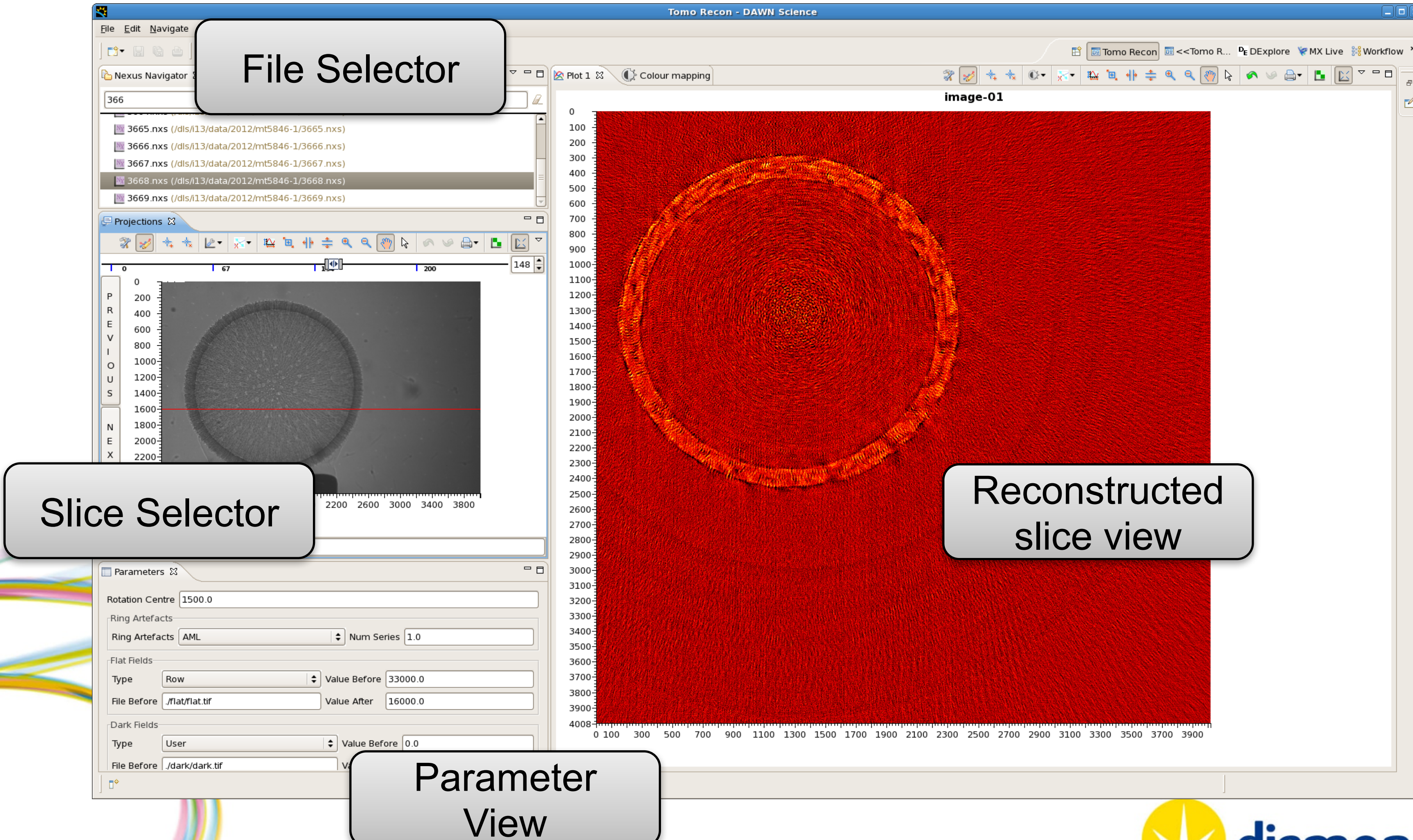
diamond

Big improvements, but GUI was required. Dawn?



diamond

Step 3: Dawn Tomo Recon UI



diamond

Move forward to 2014

- **Is there a problem with the Sinogram assumption?**

- **Is there a problem with assuming 3D only.**

Move forward to 2014

- **Is there a problem with the Sinogram assumption?**
  - Much processing occurs in projection space, which we have deliberately discounted for speed and convenience:
    - Paganin filter.
    - Distortion correction.
    - Frame stitching.

- **Is there a problem with assuming 3D only.**
  - XRD, XRF tomography,
  - Time resolved becoming more prevalent.

Step 4: We re-evaluated our approach.

PHILOSOPHICAL TRANSACTIONS A

rsta.royalsocietypublishing.org

Research

CrossMark
click for updates

Cite this article: Atwood RC, Bodey AJ, Price SWT, Basham M, Drakopoulos M. 2015 A high-throughput system for high-quality tomographic reconstruction of large datasets at Diamond Light Source. *Phil. Trans. R. Soc. A* **373**: 20140398.
http://dx.doi.org/10.1098/rsta.2014.0398

Accepted: 10 February 2015

One contribution of 11 to a theme issue 'X-ray tomographic reconstruction for materials science'.

A high-throughput system for high-quality tomographic reconstruction of large datasets at Diamond Light Source

Robert C. Atwood, Andrew J. Bodey, Stephen W. T. Price, Mark Basham and Michael Drakopoulos

Diamond Light Source Ltd, Harwell Science and Innovation Campus, Didcot OX11 0QX, UK

Tomographic datasets collected at synchrotrons are becoming very large and complex, and, therefore, need to be managed efficiently. Raw images may have high pixel counts, and each pixel can be multidimensional and associated with additional data such as those derived from spectroscopy. In time-resolved studies, hundreds of tomographic datasets can be collected in sequence, yielding terabytes of

diamond

We investigated the existing software

- Existing tomography packages
- Streaming technologies (Spark, …)

- But none were able to deal with some of the use cases we needed to cover

- So reluctantly, we decided to write some new software.

A tomography pipeline called Savu.

Aside : Consider Ophidiophobia when making a logo.

savu

diamond

Feature 1 : Standard and Scalable

- Standard Tools
  - Python
    - Conda install
  - Mpi4py
    - Generic across many clusters
  - H5py
    - With phdf5

Feature 2 : Plugin structure

Core Developer

- Parallelisation
- Performance
- ND compatibility
- Code Structure
- Sustainability

Plugin Developer

- Science case
- Local optimisation
- Compare with others

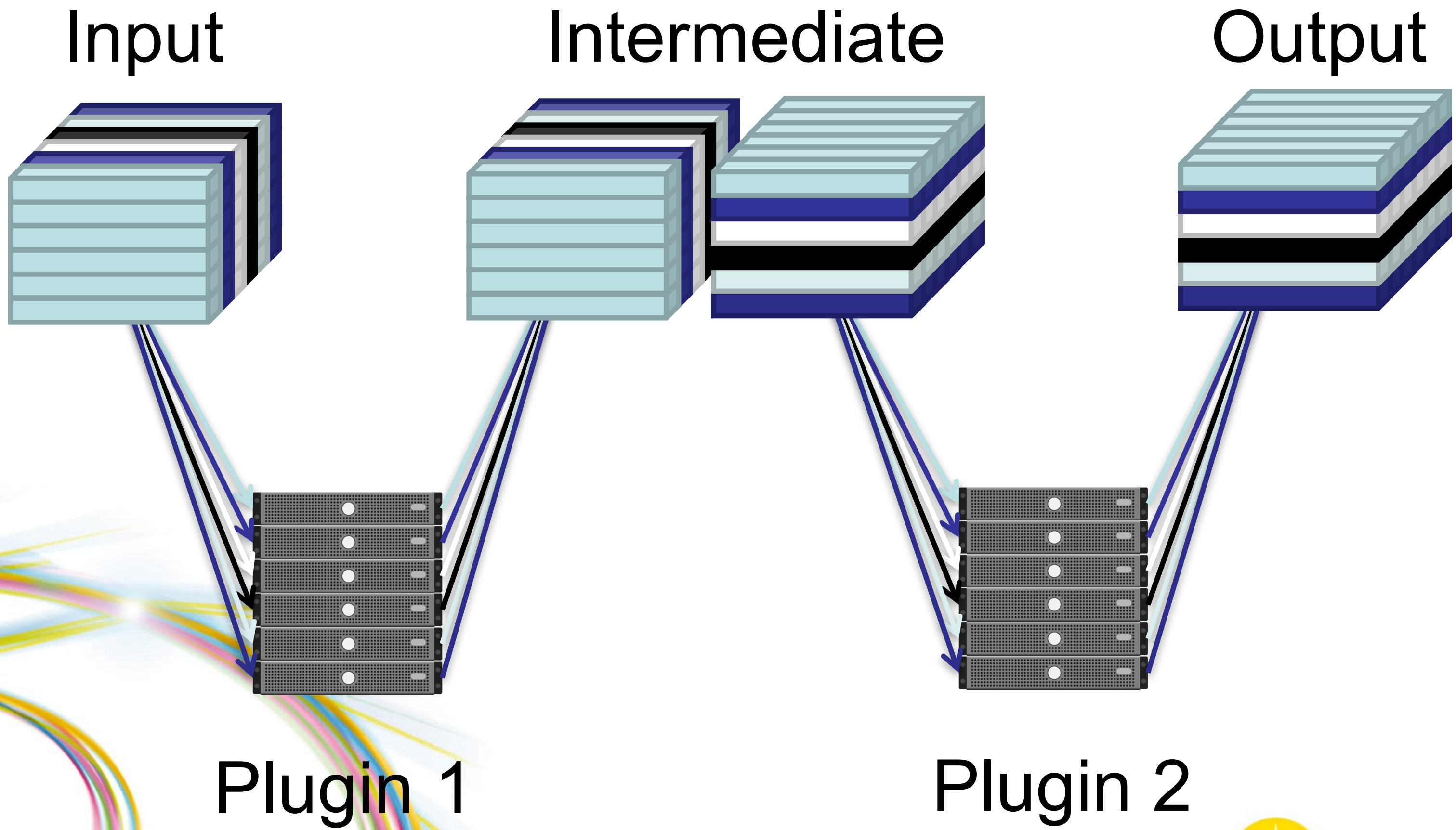
Example Plugin

```python
import numpy as np
import dezing
from savu.plugins.base_filter import BaseFilter
from savu.plugins.driver.cpu_plugin import CpuPlugin

class DezingFilter(BaseFilter, CpuPlugin):
    """
    A plugin to remove zingers.
    :param outlier_mu: Magnitude for detecting outlier. Default: 10.0.
    :param kernel_size: Number of frames included in average. Default: 5.
    """

    def __init__(self):
        super(DezingFilter, self).__init__("DezingFilter")

    def pre_process(self, exp):
        in_data = self.get_data_objects(exp.index, "in_data")[0]
        data_size = in_data.get_shape()
        self.padding = (self.parameters['kernel_size'] - 1) / 2
        dezing.setup_size(data_size, self.parameters['outlier_mu'],
                          self.padding)

    def filter_frame(self, data):
        result = np.empty_like(data[0])
        dezing.run(data[0], result)
        return result

    def post_process(self):
        dezing.cleanup()

    def set_filter_padding(self, in_data, out_data):
        pad = self.padding
        in_data[0].padding = {'pad_multi_frames': pad}
        out_data[0].padding = {'pad_multi_frames': pad}
```

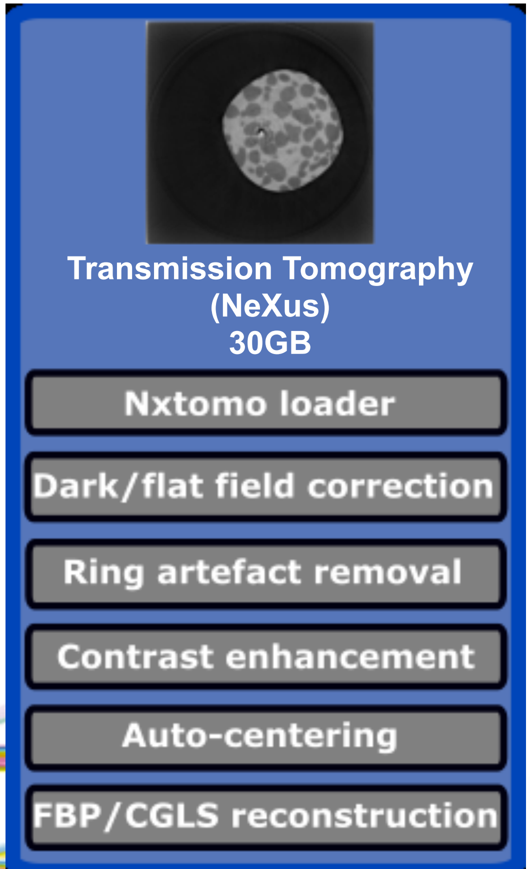
diamond

Feature 3 : Allow Data Transpose

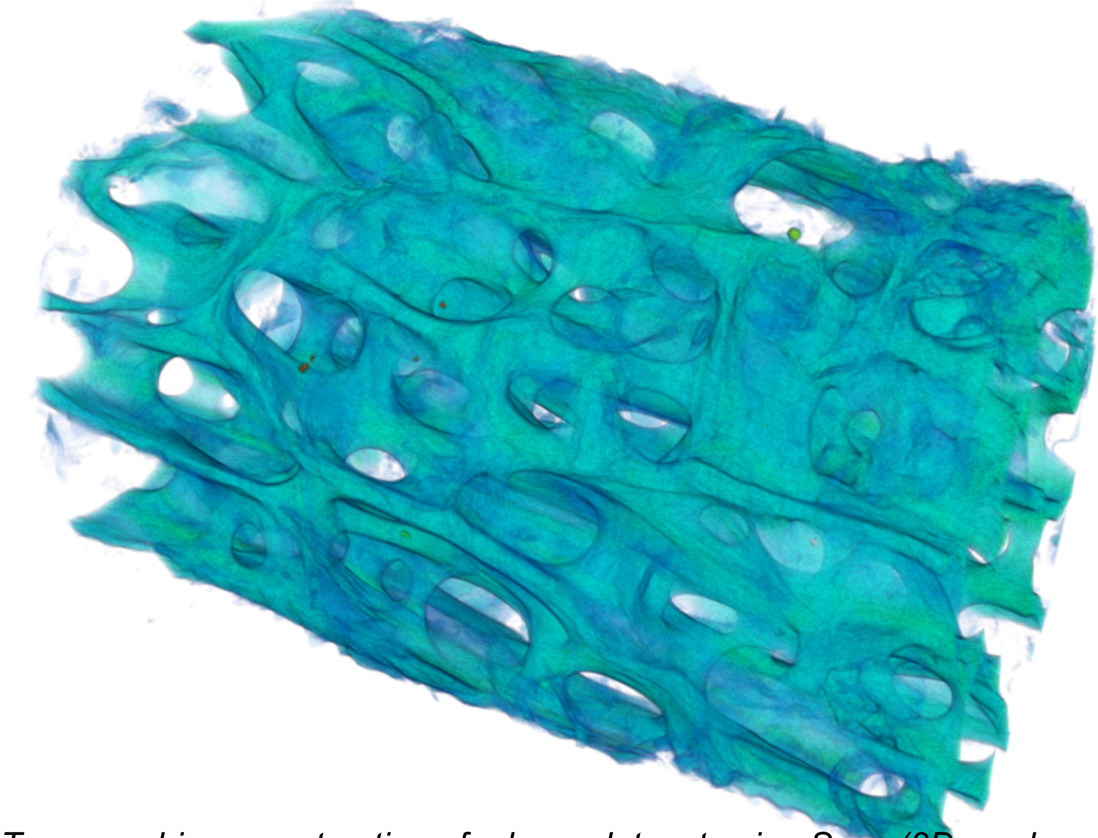
Optimally chunked HDF5 datasets.



Full-field tomography processing with Savu at DLS



A typical process list for reconstructing full-field tomography data.



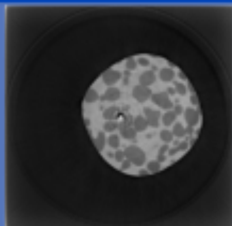
Tomographic reconstruction of a bone dataset using Savu (3D-rendered using VisIt). Courtesy of Gianluca Tozzi, Marta Pena-Fernandez, Rachna Parwani, and Asa H. Barber (2016) from Portsmouth University. Data collected on the Diamond Manchester Imaging Branchline (I13-2) with support from Andrew J. Bodey.

diamond

Feature 4 : Multidimensional Data

- Deals with ND data as plugins use patterns
- Sample Patterns
  - Projection
  - Sinogram

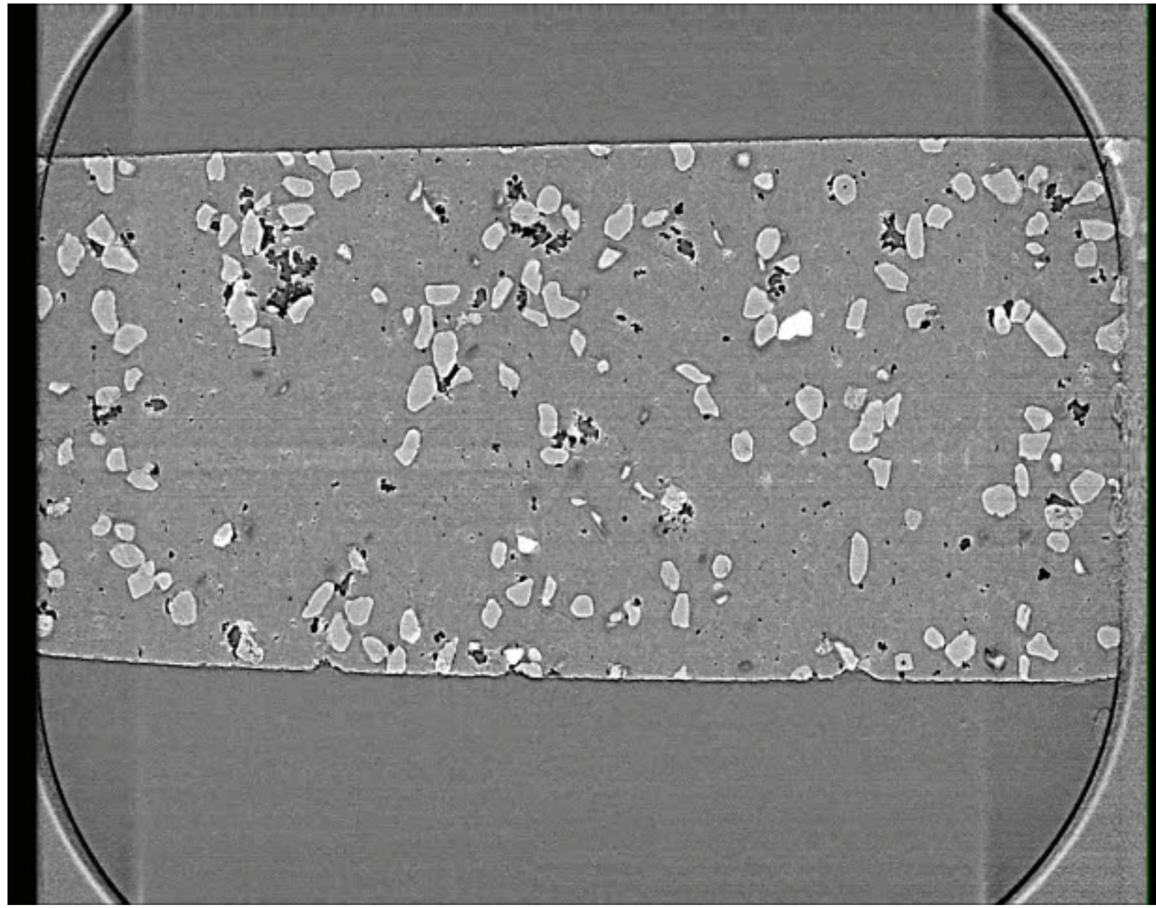
Full-field tomography processing of 4D data with Savu at DLS



Time Resolved Tomography (cine)
1500GB

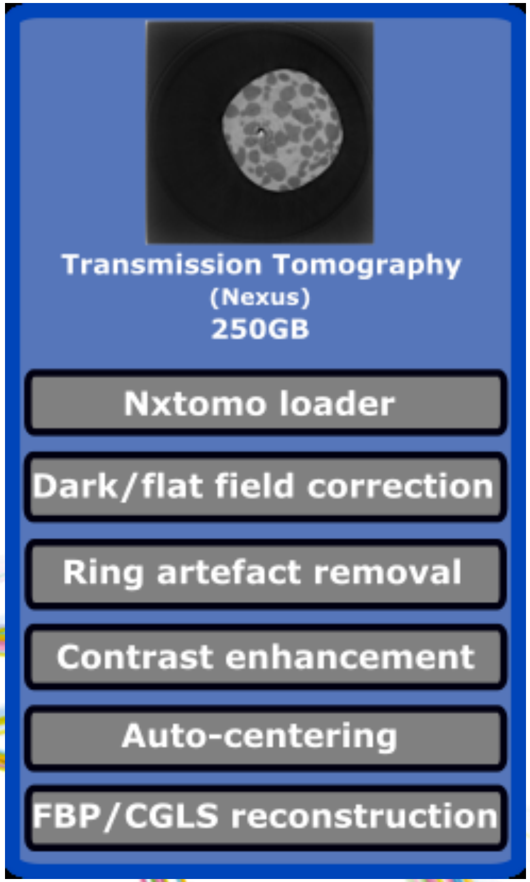
- Nxtomo loader
- Dark/flat field correction
- Ring artefact removal
- Contrast enhancement
- Auto-centering
- FBP/CGLS reconstruction

A typical process list for reconstructing full-field tomography data.



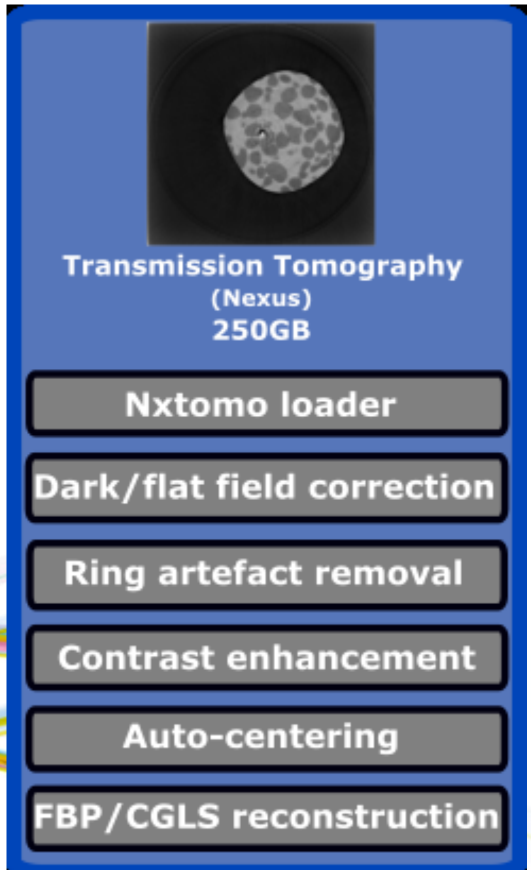
diamond

Feature 5 : Parameter Tuning



- Set any parameter to a range or sequence.
- Increases the dimension of the result at that point.
- This allows you to mix and match parameters across different plugins.

Feature 5 : Parameter Tuning



- Set any parameter to a range or sequence.
- Increases the dimension of the result at that point.

3D

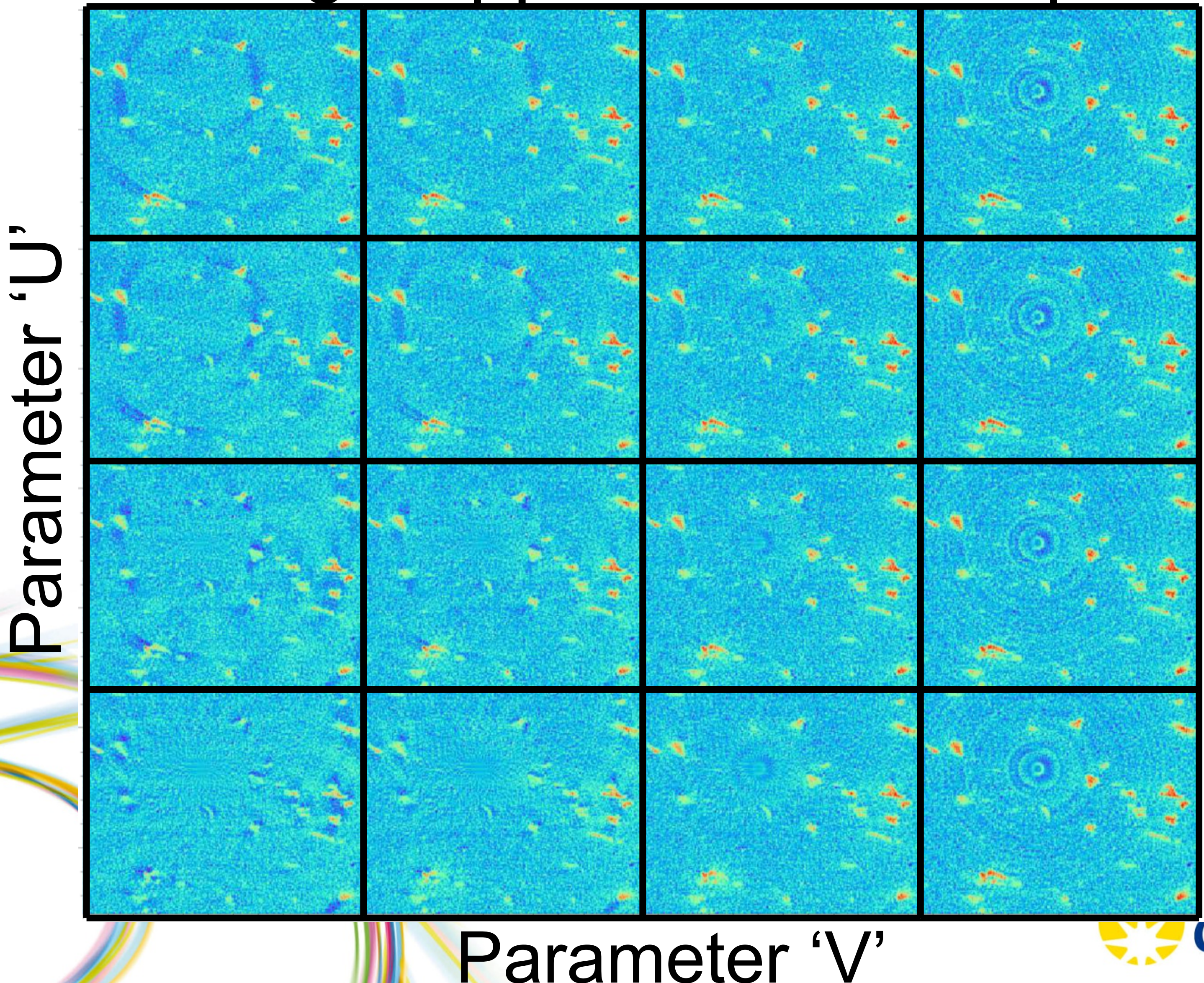
3D

3D -> 5D

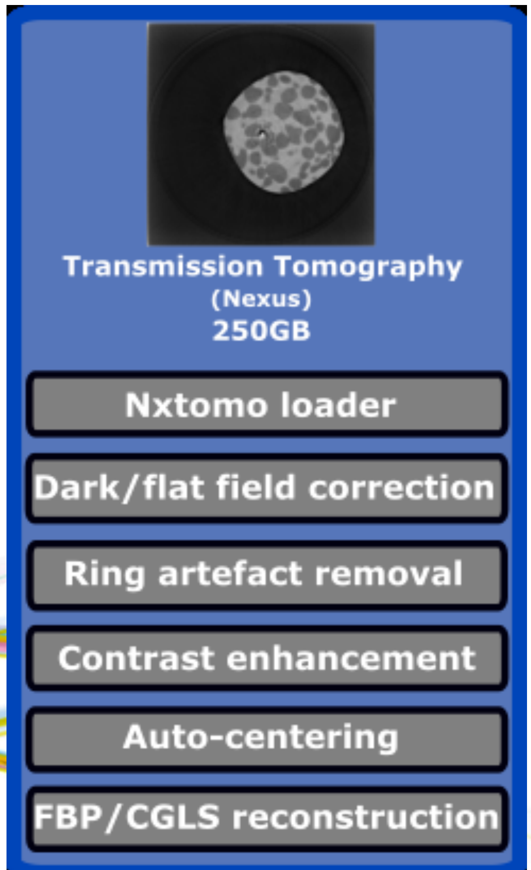
5D

5D

Ring Suppression Example



Feature 5 : Parameter Tuning



- Set any parameter to a range or sequence.
- Increases the dimension of the result at that point.

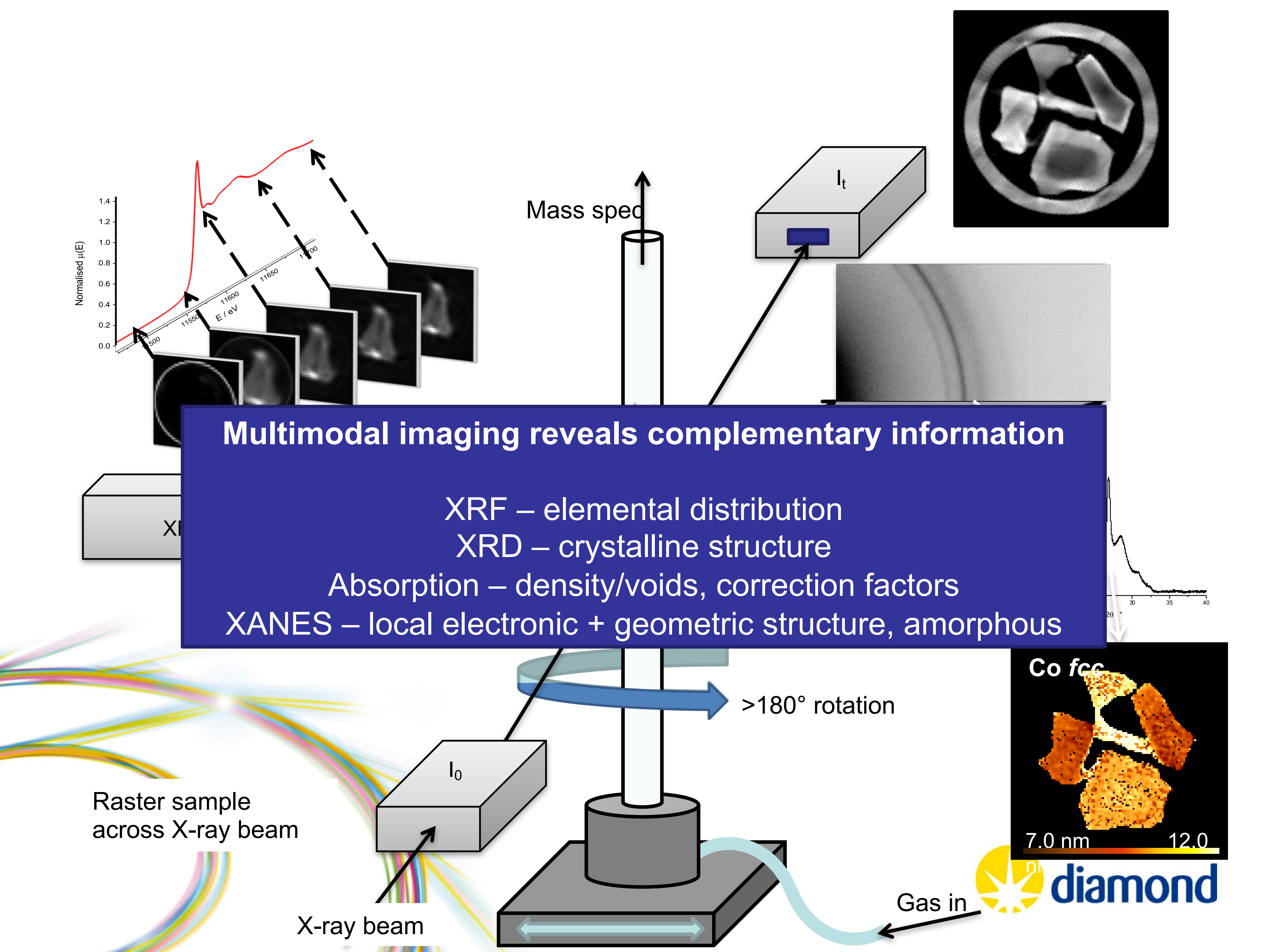
3D	3D
3D	3D
3D -> 5D	3D -> 4D
5D	4D -> 6D
5D	6D

diamond

Feature 6 : Multimodal Data

- Specifically data collected at the same time, with the same reference frame

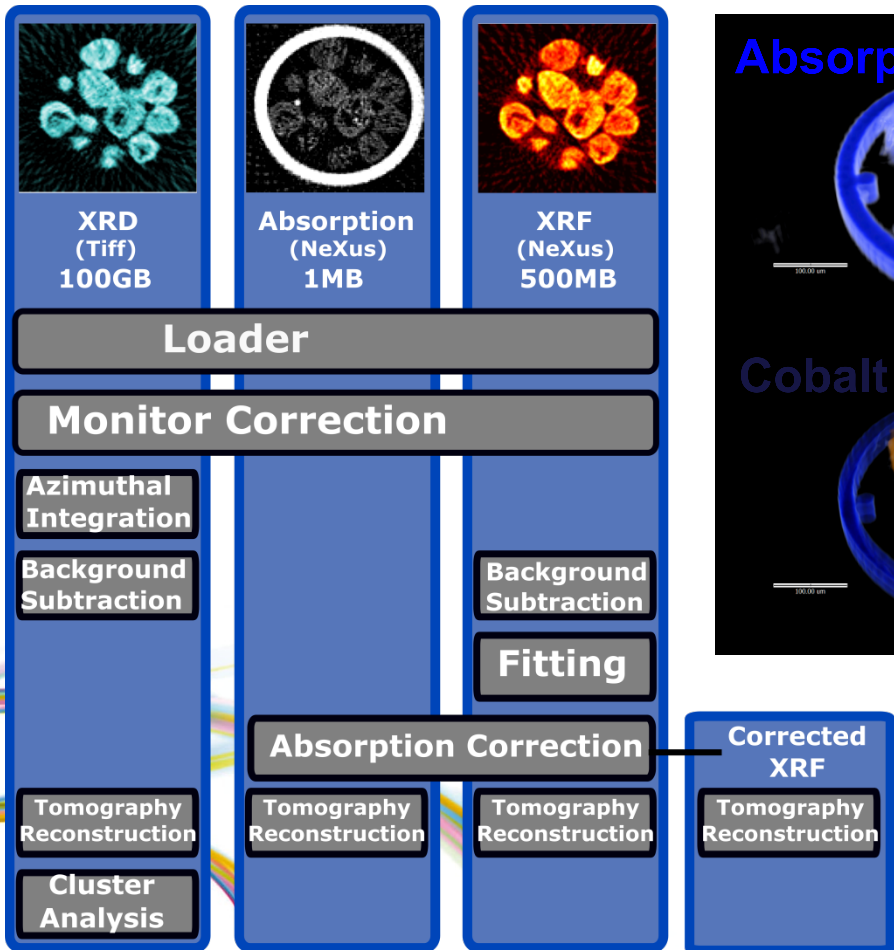
- Savu does not deal currently with data collected with different reference frame, but there is nothing to stop it from doing so!

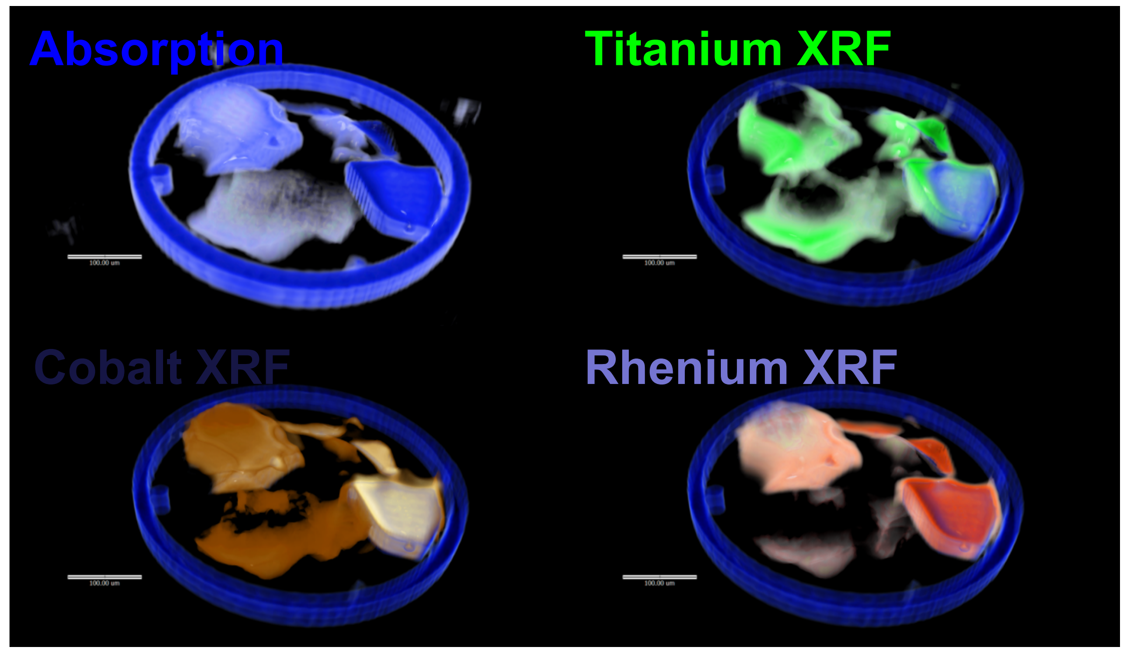


Multimodal imaging reveals complementary information

XRF – elemental distribution
XRD – crystalline structure
Absorption – density/voids, correction factors
XANES – local electronic + geometric structure, amorphous

Multi-modal tomography processing with Savu at DLS

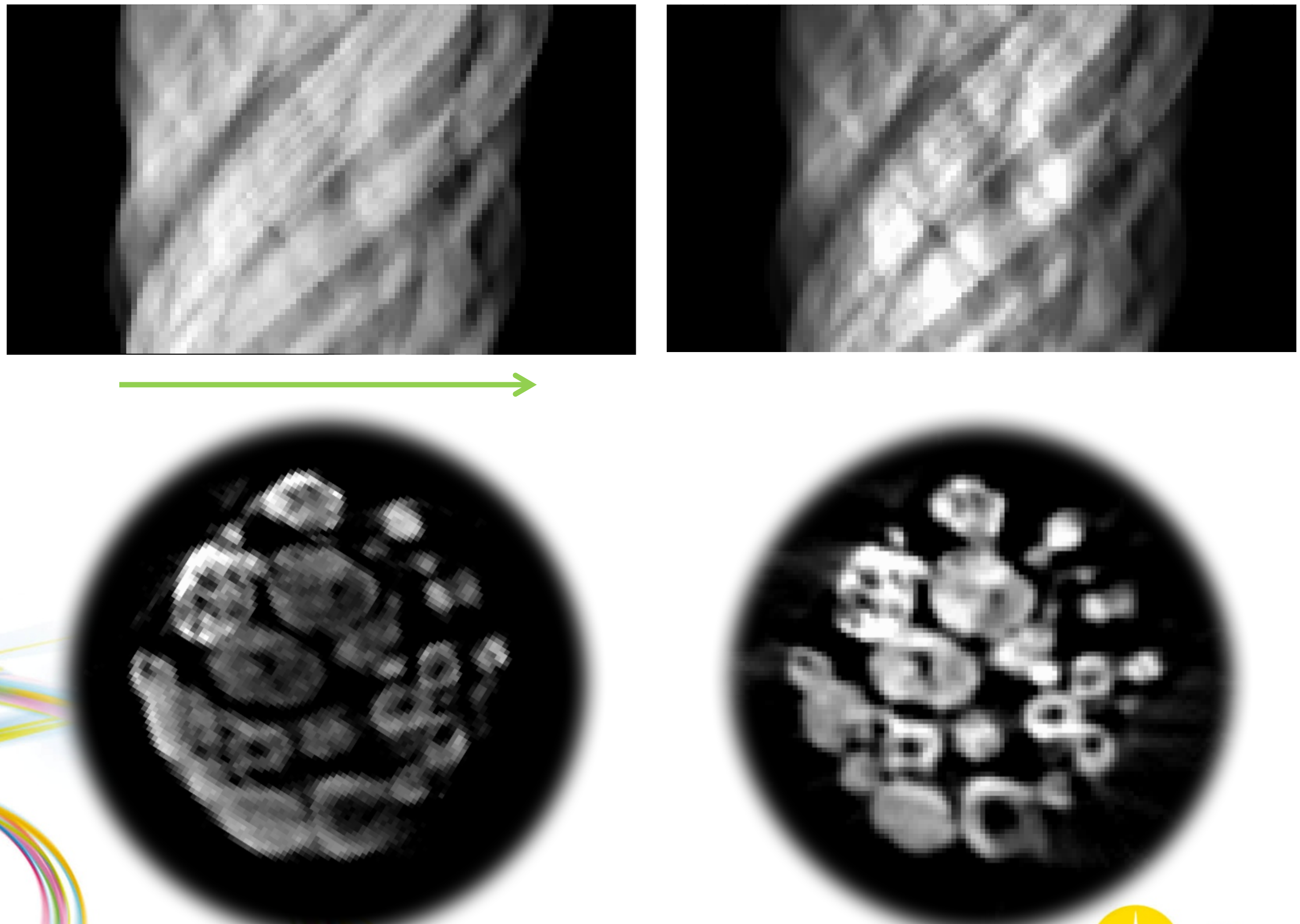




Multimodal chemical tomography: absorption, XRF, XRD (not shown) of Re and Ti promoted Co/SiO_2 catalyst for Fischer Tropsch Synthesis collected at room temperature.

550 x 550 x 50 µm volume imaged
3 x 3 µm spot, raster scan 5 µm translation, 3 ° rotation
13 hr collection. Processing/reconstruction: 2 min absorption, 10 min XRF, 1 hr XRD (5-10 minutes/slice)
1.3 GB absorption/XRF → ca. 10 MB reconstructed volume (fitted XRF)
942 GB XRD → ca. 2 GB reconstructed volume

Absorption correction



Feature 7 : Citation Information

- All plugins can contain citation information.
  - Doi
  - Bibtex/endnote info
  - Description

- citation_extractor.py
  - Pulls all this out, and creates a “method” section for a paper.

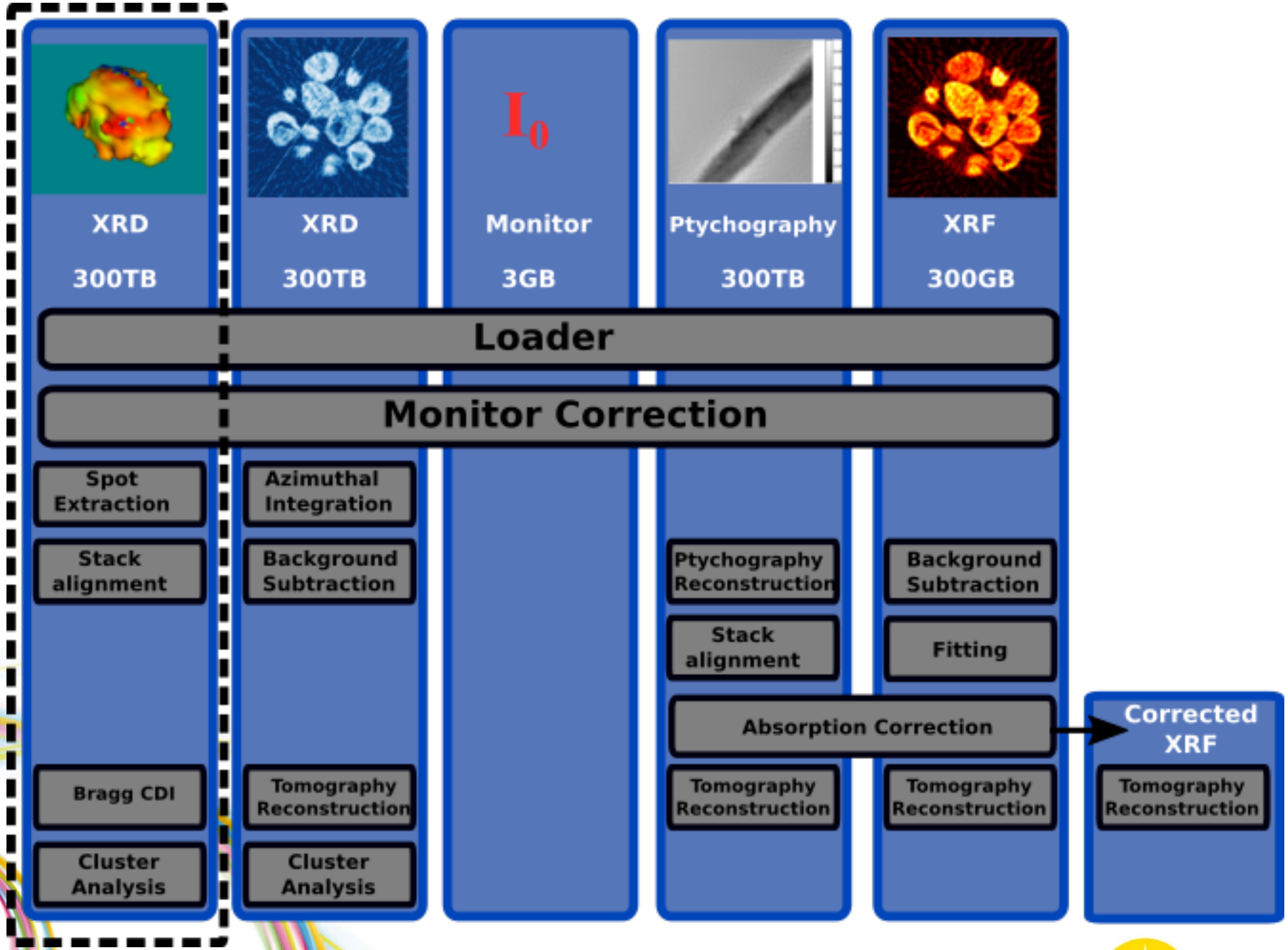
Where Next

- Savu Core
  - New Backends
    - SWMR + VDS HDF5
    - Object Stores (Ceph, etc)
- Savu Plugins
  - Alignment
- GUI of some description
  - Update the Dawn UI to drive Savu?
    - It can already run some savu plugins.
  - Web UI?
  - …

Where Next

- Things to integrate
  - Alignment (Enables the Nanoscale community)
  - Xdesign (Phantom generation and quality comparisons, for benchmarking)
  - Tigre (Iterative reconstruction, being ported to python, more reconstruction methods)
  - Hyperspy (Electron Energy Loss Spectroscopy methods)
  - Yours?

Science Driver : I14 Use Case



Thank you!

The Team

Alun Ashton

Nicola Wadeson

Aaron Parsons

Nghia Vo

mark.basham@diamond.ac.uk

@basham_mark

diamond