Metadata Issues on the I16 Beamline at DLS

S P Collins
Diamond Light Source

Metadata Issues on the I16 Beamline at DLS

S P Collins
Diamond Light Source

I16 is a beamline for:

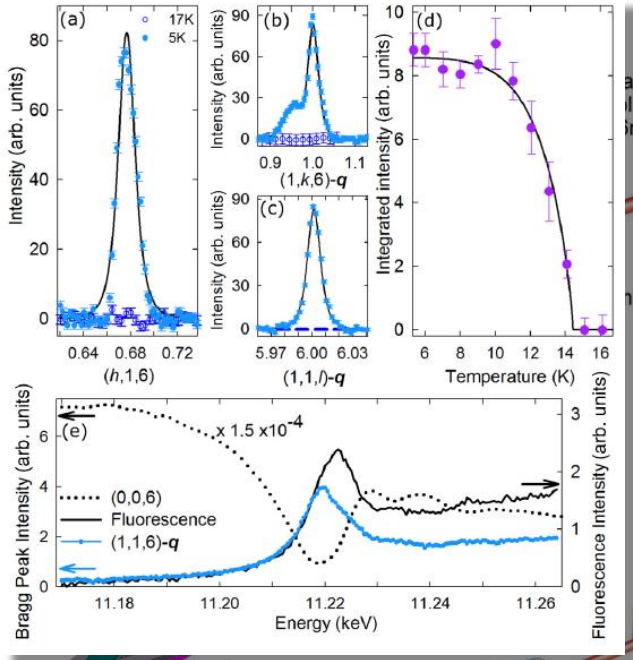
- Resonant/magnetic single crystal x-ray diffraction
- General x-ray physics

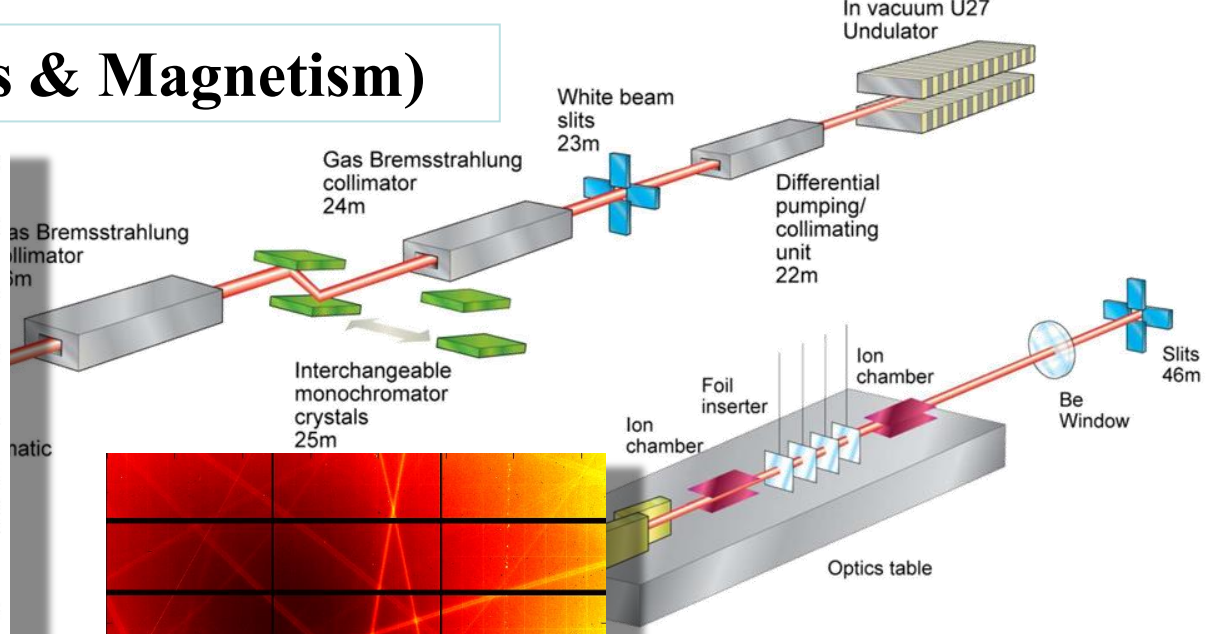
Current NeXus file-writing is rudimentary and not fit-for-purpose

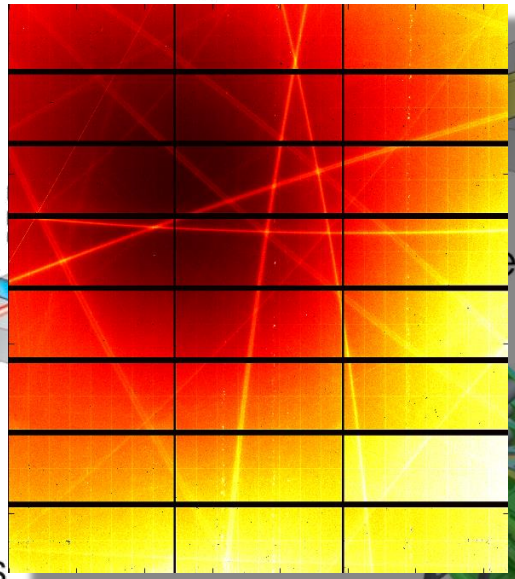
I16 is quite a good model to develop NeXus writing/metadata for Diamond beamlines as part of the wider neutron and synchrotron science community

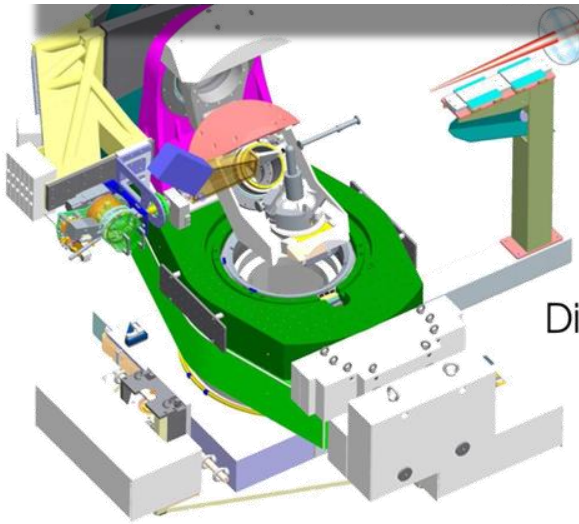
diamond

Diamond I16 (Materials & Magnetism)



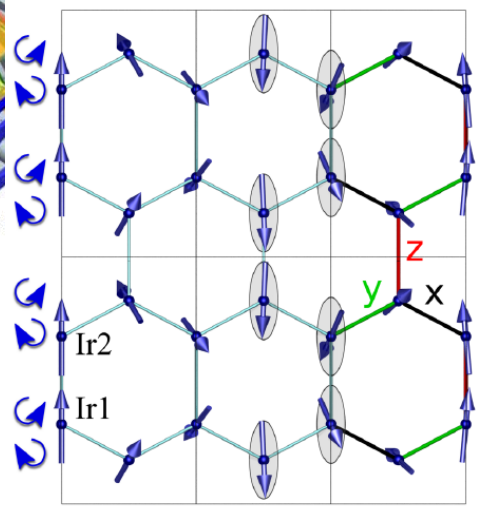






Six-axis
kappa
Diffractometer

Detector
arm
assembly



I16 general requirements (a personal view!):

1. Carry out measurements/scans using traditional methodology; display/manipulate parameters using local vocabulary (epics names *etc*); integrate with notebook functionality (*e.g*. Jupyter)
2. Provide raw data with agreed **metadata** (parameters and vocabulary) within multiple scientific domains/disciplines (NeXus Application Definitions?)
3. Data to be optimized for standard data pipelines/workflows, allowing shared/contributed processing software
4. Metadata automatically stored in standard database with programmatic queries
5. Data storage, transfer, workflows, VMs, containers, database functions, machine learning, implemented using common cloud technologies.
6. Ultimate goal: users of p&n facilities across Europe/World to manipulate data via a common framework, feed into a global database of processed data (ultimately knowledge) with provenance trail.

diamond

1. Carry out measurements/scans using traditional methodology; display/manipulate parameters using local vocabulary (epics names *etc*); integrate with notebook functionality (*e.g.* Jupyter)

Home | tmp_nexus_demo_1 | Confidentiality and pre-publicat

localhost:8888/notebooks/Dropbox/python_i16/steve/tmp_nexus_demo_1.ipynb

Most Visited | Getting Started | Web Slice Gallery | Amazon.co.uk Sign In

jupyter tmp_nexus_demo_1 Last Checkpoint: 03/08/2018 (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

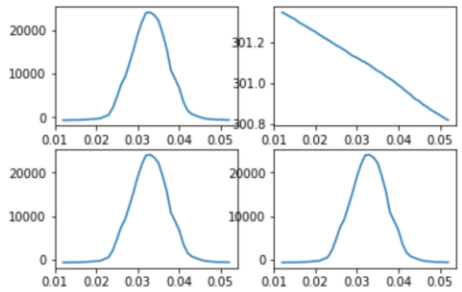
Not Trusted | Python 2

Run | Code

Out[11]:
```
0     0.0060
1     0.0065
2     0.0070
3     0.0075
       ...
37    0.0245
38    0.0250
39    0.0255
40    0.0260
Name: sgome, Length: 41, dtype: float64
```

In [12]:
```
plt.figure();
plt.subplot(2,2,1); plt.plot(n.sgome, n['sum'])
plt.subplot(2,2,2); plt.plot(n.sgome, n.rc)
plt.subplot(2,2,3); plt.plot(n.sgome, n['sum'])
plt.subplot(2,2,4); plt.plot(n.sgome, n['sum'])
```

Out[12]:
```
[<matplotlib.lines.Line2D at 0x7fb4b0e3fe90>]
```



In [ ]:

Screenshot Added
A screenshot was added to your Dropbox.

2. Provide raw data with agreed **metadata** (parameters and vocabulary) within multiple scientific domains/disciplines (NeXus Application Definitions?)

General requirements:

Data processing pipelines to accommodate multiple ***measurement types*** on each beamline

Metadata for each type using vocabulary agreed within scientific domain/community

Questions:

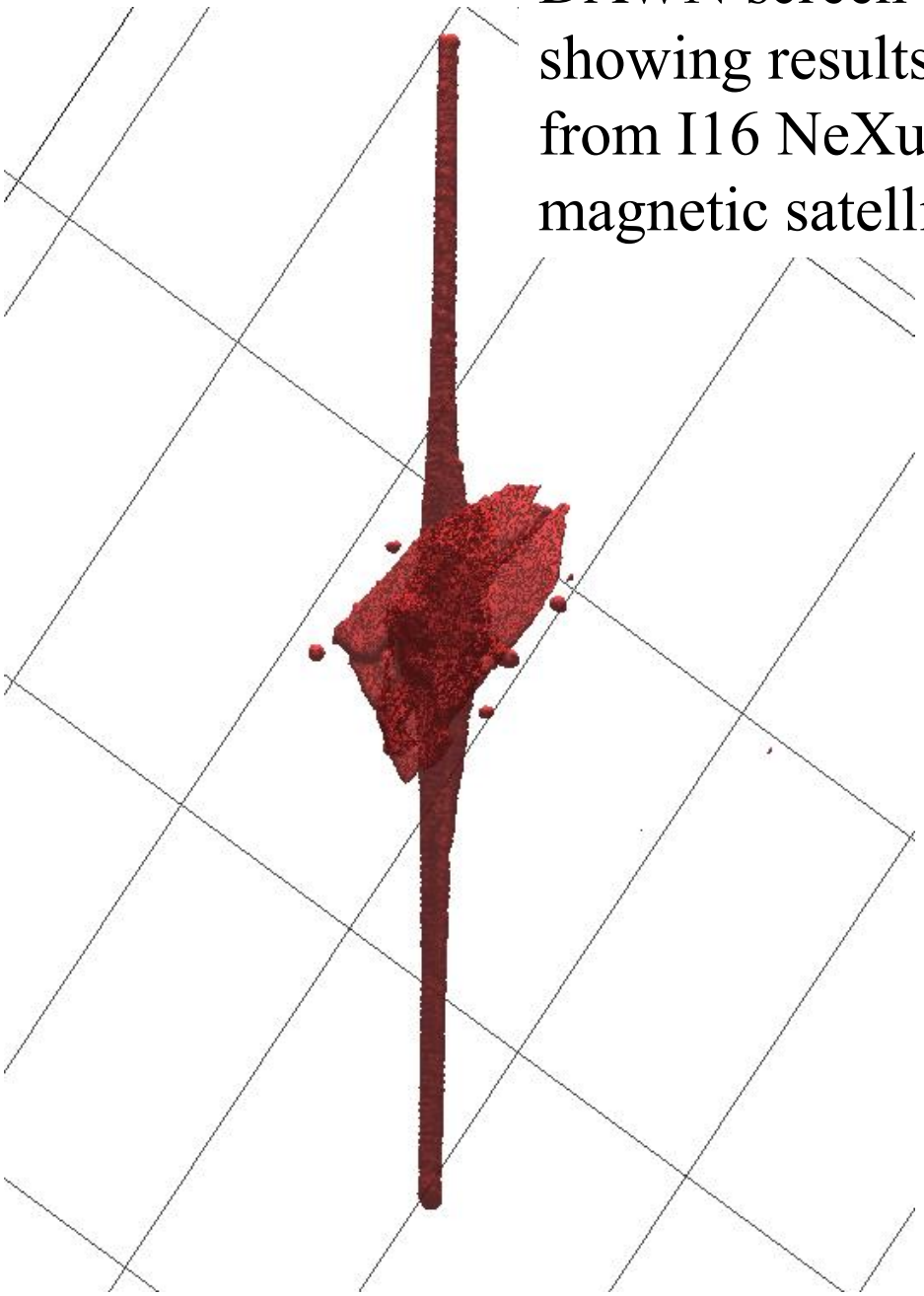
Do Application Definitions map onto measurement types?

Is it acceptable to use a large number of Application Definitions?

Do Application Definitions have a future within the NeXus community? What about 'Features'?

diamond

DAWN screen dump showing results of hkl map from I16 NeXus file (four magnetic satellites)



hkl map from CuMnAs thin film (mayavi)

NeXus User Manual: Application Definitions

NeXus application definitions define the minimum required information necessary to satisfy data analysis or other data processing.

Another way to look at a NeXus application definition is as a contract between a file producer (writer) and a file consumer (reader).

The contract reads:

If you write your files following a particular NeXus application definition, I can process these files with my software.

Have we given up on this?

diamond

Useful Application Definitions for Diamond Beamline I16 (raw data)?

NXmx (crytal diffraction, referes to: NXdetector, NXinstrument etc etc)

NXfluo

NXxas

NXscan (generic nD scan in scalar/vector space (base class?))

NXarchive (??)

Missing information for core activities (resonant and magnetic x-ray scattering)

Stokes parameters
Polarization analyser settings
Azimuthal angle
Absolute diffraction intensities

New application definitions citing new base classes

diamond

Possible structure of Diamond/I16 NeXus files

Traditional scan data (e.g. spec)

Complete beamline description; allows automatic ray-tracing/simulation with no additional parameters;

Flat dump of local parameters (meaningless without beamline manual)

Scan object (base class instance); basic scan output, plotting…

Multiple Application Definitions; links to instrument

For automatic archiving of all NeXus tree to database

```
entry:NXentry
    @default = data
    user:NXuser
    sample:NXsample
    instrument:NXinstument
        SASdet:NXdetector
            data:[,]
        fluordet:NXdetector
            data:[,]
        large_area:NXdetector
            data:[,]

    beamline_local:NXcollection
        slit1v=
        slit1h=
        eta=
        delta=
        bla
        bla

    scan:NXsubentry
        definition = "NXscan"
        scan_command="scan th 10 20 .1 pil 1 roi1"
        called_by="myscript.py"
        start_time=
        success=1
        scan_data:NXdata
            th --> /entry/instrument/theta
            pil --> /entry/instrument/detector/pil100k/data
            sum --> /entry/instrument/detector/pil100k/sum
            roi1 --> /entry/instrument/detector/pil100k/roi

    SAS:NXsubentry
        definition = "NXsas"
        instrument--> /entry/instrument

    mx:NXsubentry
        definition = "NXmx"
        instrument--> /entry/instrument

    archive:NXsubentry
        definition = "NXarchive"
        bla
        bla
```

diamond

NeXus metadata/experiment description:
What's still missing?

Intent (formal or informal)

- What was the purpose of the measurement/scan?
- Was is just an alignment scan?
- Which Application Definition(s) are expected to be applied for data analysis?

Success (post scan/experiment)

- Is the dataset likely to be of interest in the future?
- Was the measurement successful?
- Was it superseded by a better one?

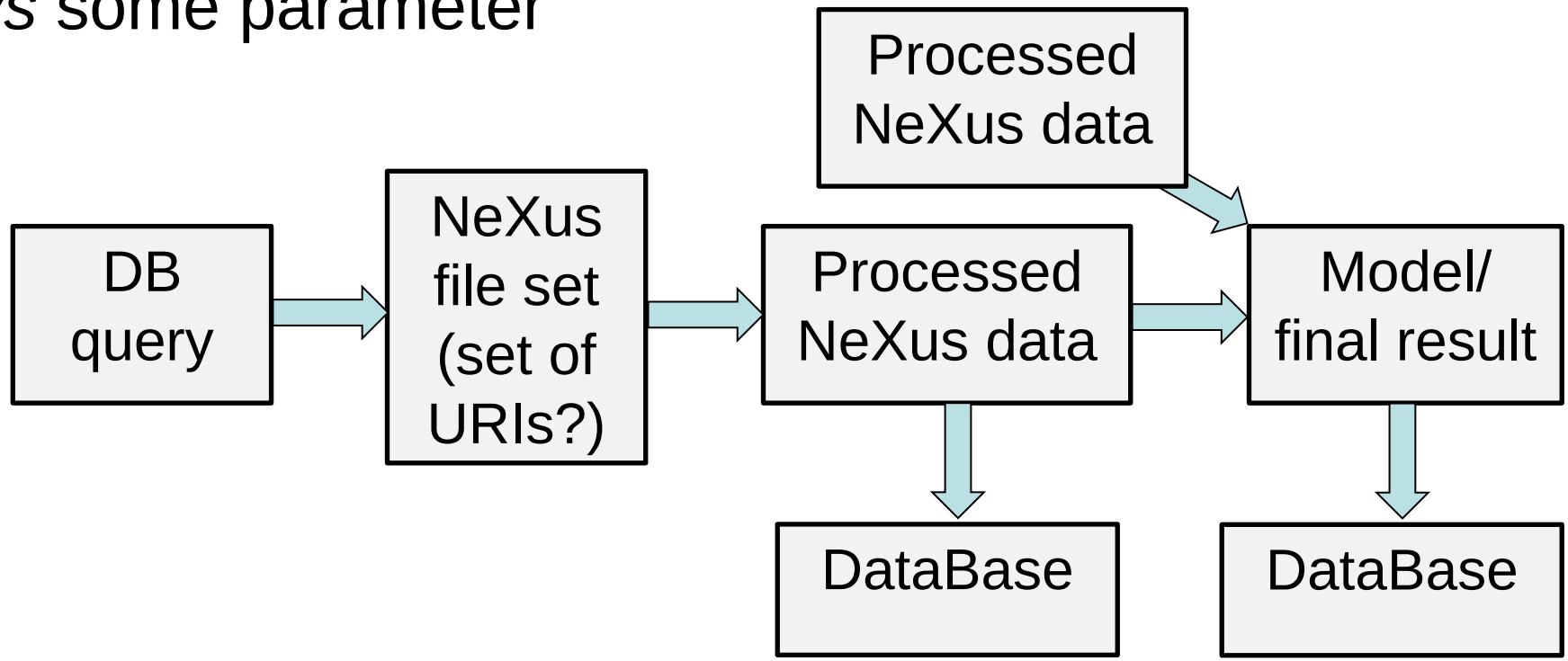
Uncertainties/errors

- Errors in all key parameters (energy, flux calibration etc)

With these fields, and appropriate NeXus files, it should be possible to process an entire experiment automatically, making quantitative comparisons with processed/derived data from the database.

diamond

Possible example workflow: sets of ‘theta’ scans *vs* some parameter



Find theta scans collected by abc.py on Beamline I16 5/11/2017	{URI1, URI2,…} Name: ‘theta scan *vs* T’ Sample: ABO3	Peak area, width, peak count-rate, peak reflectivity… *vs* T	Unit cell volume *vs* T
Bla bla	Bla bla	Bla bla	Magnetic moment vs T

Metadata Issues on the I16 Beamline at DLS

The way forwards (a personal view):

Now

1. **Agree** common (NeXus) structures with other light sources, for the benefit of science communities.
2. Confirm future of/commitment to **Nexus Application Definitions** (and/or Features *etc*) with other light sources, for the benefit of science communities.
3. Agree standards in order to share workload and not duplicate effort as this is **damaging to the progress of science**

Medium term

4. Agree interaction with databases.

Longer term

5. Agree common approaches to workflows, databases for processed data, knowledge base, common cloud framework…

diamond