What Every Experimentalist Needs to Know about Recording Essential Metadata of Primary (i.e. Raw) Diffraction Data

Herbert J. Bernstein
Rochester Institute of Technology

Work Supported in Part by Dectris

HDRMX

R·I·T

BROOKHAVEN
NATIONAL LABORATORY

What Types of Data are Needed in MX?

- There are several major types of MX Data:
  - Raw diffraction images, usually recorded as arrays of pixel intensities
  - Corrected diffraction images (geometrical corrections, count rate corrections, flat field corrections, ...)
  - Derived data, e.g. structure factors, fitted models, ...
  - Metadata – data that provides information about other data. Metadata may be raw, corrected or derived
- What we call metadata and what metadata is essential depends on the experiment, e.g. for simple rotation exoeriments.
  - We need beam center, detector distance and wavelength to make any sense of diffraction data
  - We need rotation angles to assemble images
  - We need exposure times to estimate radiation damage
  - etc. …
-

R·I·T

BROOKHAVEN
NATIONAL LABORATORY

Why Fuss About Metadata?
Isn’t my lab notebook sufficient?

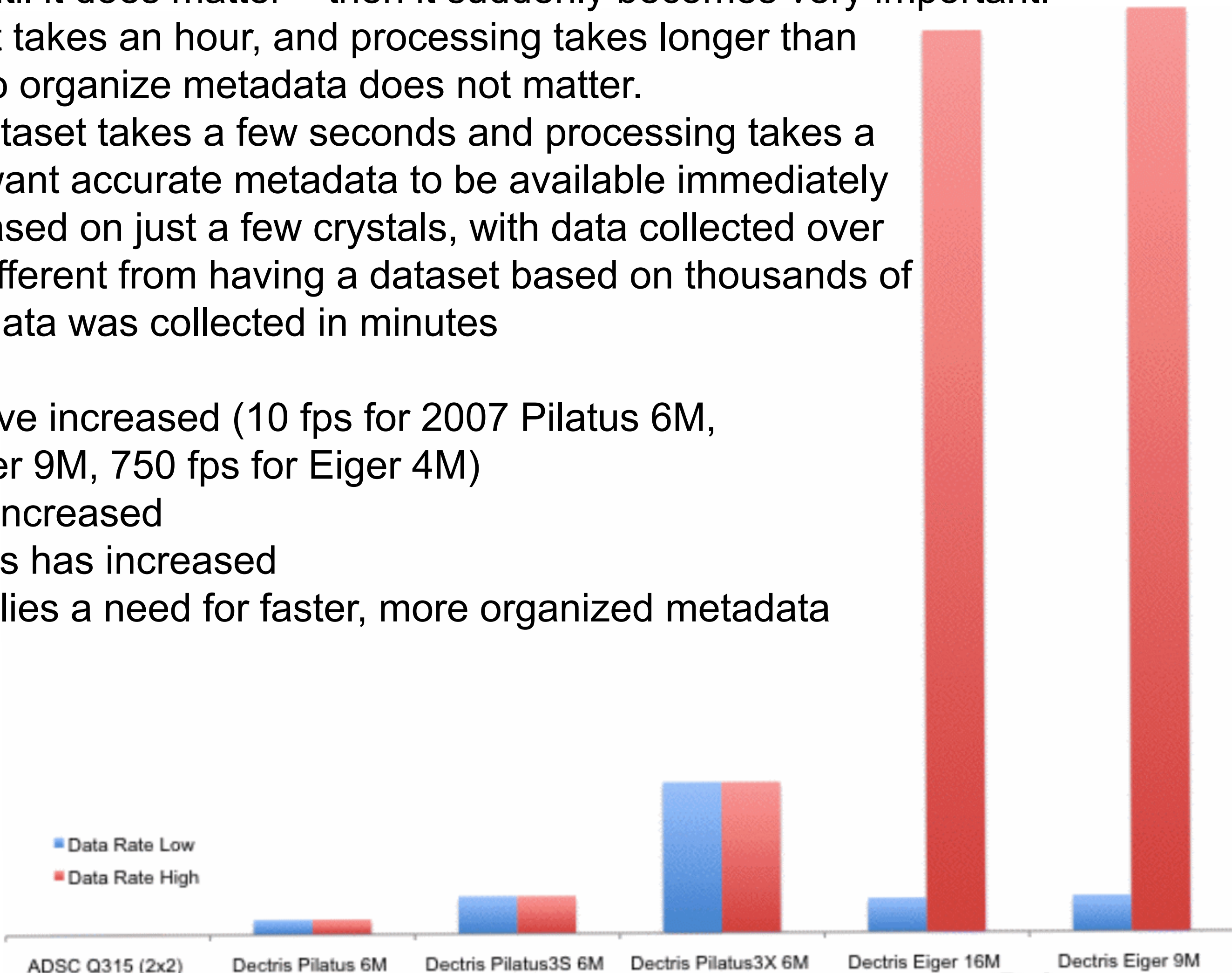
- Data without metadata is easily and often misinterpreted.
  - We need beam center, detector distance and wavelength to make any sense of diffraction data
  - We need rotation angles to assemble images
  - We need exposure times to estimate radiation damage
  - etc. …
- Lab notebooks are great for processing your own data, but may get decoupled from data, especially when the data is archived or deposited
- As data collection rates increase, manual methods for managing metadata become less and less practical and more error-prone
- Different data-processing protocols may require different sets of metadata.

R·I·T

BROOKHAVEN NATIONAL LABORATORY

Why Does Speed Matter?

- Speed does not matter until it does matter – then it suddenly becomes very important.
- When collecting a dataset takes an hour, and processing takes longer than
  - that, a few minutes to organize metadata does not matter.
  - When collecting a dataset takes a few seconds and processing takes a
  - minute or less, you want accurate metadata to be available immediately
  - When a dataset is based on just a few crystals, with data collected over
  - hours or days, it is different from having a dataset based on thousands of
  - crystals from which data was collected in minutes
  - etc. …
  - Dectector speeds have increased (10 fps for 2007 Pilatus 6M, 238 fps for 2017 Eiger 9M, 750 fps for Eiger 4M)
- Processing speeds have increased
- Use of many small crystals has increased
- Faster data collection implies a need for faster, more organized metadata
- collection



R·I·T

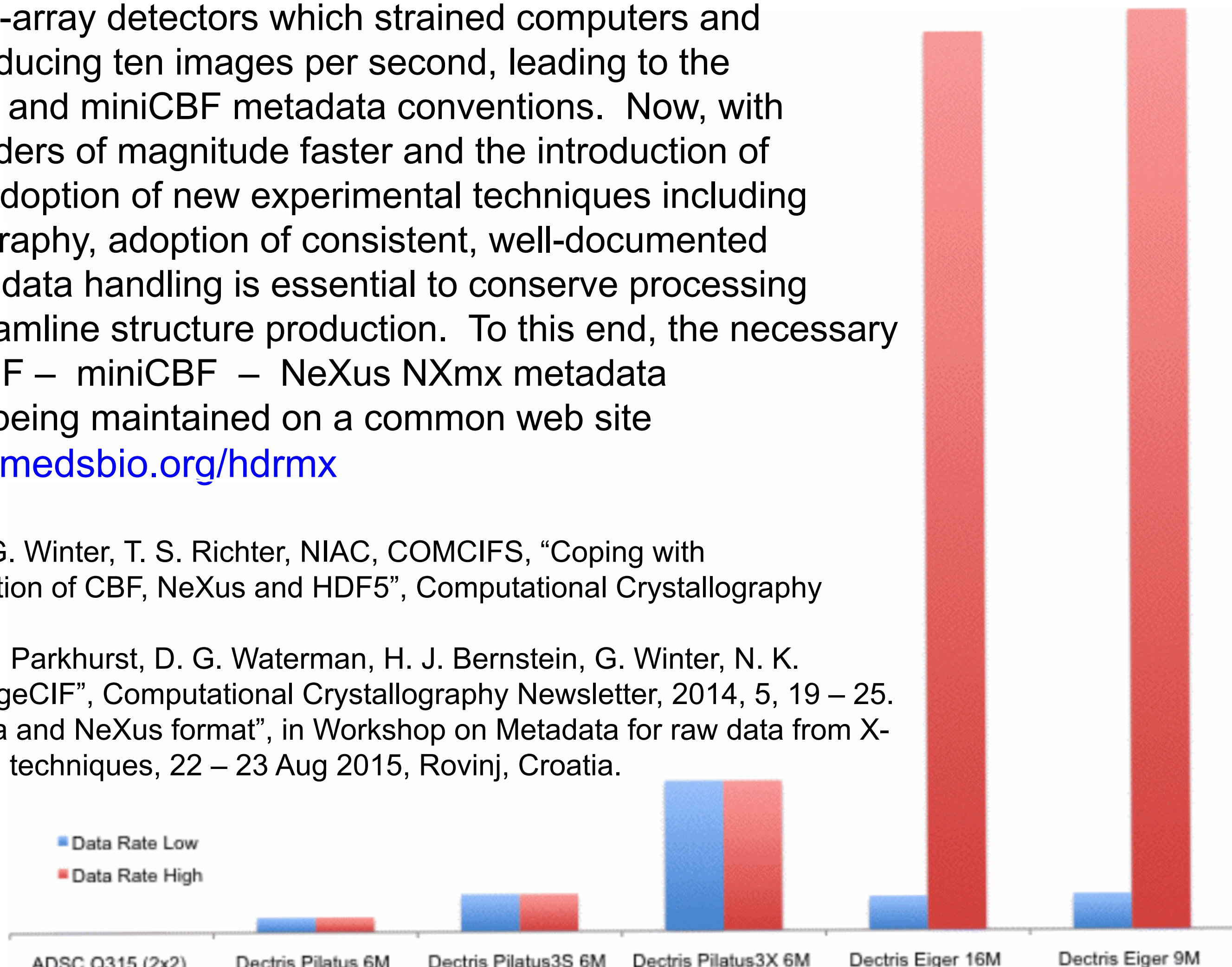
BROOKHAVEN NATIONAL LABORATORY

Why Does Speed Matter (continued)

The last time our community faced a similar speed-constrained transition was with the Dectris Pilatus pixel-array detectors which strained computers and networks of that time by producing ten images per second, leading to the adoption of the imgCIF/CBF and miniCBF metadata conventions. Now, with data arriving one to three orders of magnitude faster and the introduction of NeXus/HDF5 images, and adoption of new experimental techniques including serial synchrotron crystallography, adoption of consistent, well-documented crystallographic-image metadata handling is essential to conserve processing resources and maximize beamline structure production. To this end, the necessary concordances of imgCIF/CBF – miniCBF – NeXus NXmx metadata specifications [1] [2] [3] are being maintained on a common web site

http://www.medsbio.org/hdrmx

[1] H. J. Bernstein, J. M. Sloan, G. Winter, T. S. Richter, NIAC, COMCIFS, “Coping with BIG DATA image formats: integration of CBF, NeXus and HDF5”, Computational Crystallography Newsletter, 2014, 5, 12 – 18.
[2] A. S. Brewster, J. Hattne, J. M. Parkhurst, D. G. Waterman, H. J. Bernstein, G. Winter, N. K. Sauter, “XFEL Detectors and ImageCIF”, Computational Crystallography Newsletter, 2014, 5, 19 – 25.
[3] M. Mueller, “EIGER HDF5 data and NeXus format”, in Workshop on Metadata for raw data from X-ray diffraction and other structural techniques, 22 – 23 Aug 2015, Rovinj, Croatia.



R·I·T

BROOKHAVEN
NATIONAL LABORATORY

Current Metadata Practices

- Some applications do not require any metadata with images and draw their metadata from separate control files.
- Some applications require metadata just sufficient to process the data as a single axis rotation experiment or as stills, e.g. using a small set of minicbf or equivalent set of NeXus tags:
-

  # Detector: Dectris Eiger 16M, S/N E-32-0101
  # Pixel_size 75e-6 m x 75e-6 m
  # Silicon sensor, thickness 0.450000 m
  # Exposure_time 0.049990 s
  # Exposure_period 0.050000 s
  # Count_cutoff 125019 counts
  # Wavelength 0.978943 A
  # Detector_distance 0.180001 m
  # Beam_xy (2012, 2420) pixels
  # Start_angle 0.800000 deg.
  # Angle_increment 0.100000 deg.

- Some applications require much more complete metadata, especially to describe complex axis configurations
- There are many MX data image formats, but two of them are worthy of special note in High Data Rate MX: the imgCIF/CBF format and metadata used for Dectris Pilatus detectors and the NeXus/HDF5 format and metadata used for Dectris Eiger detectors. Most major MX applications have ways to accept CBF images directly and an increasing number have ways to accept HDF5 images directly.

Current Metadata Practices (continued)

- imgCIF/CBF is managed by the IUCr Committee on the Maintenance of the CIF Standard (COMCIFS)
- NeXus/HDF5 NXmx is managed by the NeXus International Advisory Committee (NIAC)
- Since 2012, COMCIFS and NIAC have been collaborating on aligning the necessary metadata for MX in support of the Dectris Eiger in particular, but with an eye on supporting data from high performance detectors at synchrotrons and XFELs.
- The objective is to make one coherent interoperable system for support of diffraction images in either NeXus/HDF5 Nxmx or imgCIF/CBF as needed.

Where to Find Useful Raw MX Metadata Tags

- See http://www.hdrmx.medsbio.org/ at the Eiger Data Tags tab

Eiger Data Tags

- In this space we will be providing information on the data and metadata tags used in handling Eiger images. There is a gradually evolving set of documents available on the Dectris website www.dectris.com. Dectri given permission for us to consolidate and summarize some of that information here for conventience. This is a work in progress. [more/less ...]

This is a table of the Dectris Software Releases of " EIGER Detector HDF5 NeXus Format" through release 1.3.0. Partial information from the NeXus NXmx application definition has been added. You may sort on by clicking on the heading. Please report comments and corrections to the hdrmx-bb list.

Version	NeXus Path	Name	Type	Units	Description	Full CBF	mini CE
< 1.2.0	/(entry):NXentry /(instrument):NXinstrument /(beam):NXbeam/wavelength	wavelength	NX_FLOAT	A	wavelength of the beam in the case of monochromatic beam	_diffrn_radiation_wavelength.wavelength 0.7085	Wavelength 0.7085 A
< 1.2.0	/(entry):NXentry /(instrument):NXinstrument /(detector):NXdetector/geometry /orientation	orientation	NX_FLOAT		To be described by Dectris		
< 1.2.0	/(entry):NXentry /(instrument):NXinstrument /(detector):NXdetector/geometry /translation	translation	NX_FLOAT		To be described by Dectris		
≥ 1.2.0	/(entry):NXentry /(instrument):NXinstrument /(beam):NXbeam /incident_wavelength	incident_wavelength	NX_FLOAT	A	Wavelength of the beam in the case of monochromatic beam	_diffrn_radiation_wavelength.wavelength	Wavelength
	/(entry):NXentry	beam_center_x	NX_FLOAT		Beam center in x in pixels.	_diffrn_detector_element.reference_center_fast	Beam_x

See http://download.nexusformat.org/sphinx/classes/applications/NXmx.html

- and especially
- http://download.nexusformat.org/sphinx/classes/base_classes/NXtransformations.html
- Think about what is there and what else is needed for your experiments.

R·I·T

BROOKHAVEN NATIONAL LABORATORY

Where to Get Metadata values for MX Metadata Tags

- If you are collecting data at a synchrotron, be sure to get all the metadata to go with your data. It may come from several different sources
  - Some of the metadata values may be stored in the detector hardware
  - Some of the metadata values may be stored in the beamline control computers
  - Some of the metadata values may be stored in facilities databases
  - Some of the metadata values may be stored in various paper or electronic laboratory notebooks or be written on a whiteboard, or just known to the local beamline scientist.
- Try to get:
  - The beam center both in pixels and in mm
  - The wavelength and polarization
  - The detector distance
  - All positioner settings, especially rotation angles
  - Some pictures of the layout
  - The names of all the files involved
  - Information on which parts of the detector images should not be used: intermodule gaps, beamstop shadows, bad pixels
- What to do with all this?
  - Preferably add it to the data files
  - At the very least – write it all down

R·I·T

BROOKHAVEN NATIONAL LABORATORY

A Taste of imgCIF/CBF

Information is organized into blocks of data
Each block of data is managed essentially in terms of tables
Tables are called “categories” or “loops”
The column headings are called “tags” or “data names”
Some tables have only one row of data
    then each tag can be put with its value
Some tables have multiple rows of data
The permitted categories and tags are documented in IUCr International Tables Volume G.
The latest software is available on github in cbflib

R·I·T

BROOKHAVEN NATIONAL LABORATORY

A Taste of imgCIF/CBF (continued)

```
# category AXIS CBF example of a Kappa goniometer
  loop_
  _axis.id
  _axis.type
  _axis.equipment
  _axis.depends_on
  _axis.vector[1] _axis.vector[2] _axis.vector[3]
  _axis.offset[1] _axis.offset[2] _axis.offset[3]
   GONIOMETER_OMEGA rotation goniometer . 1 0 0 . . .
   GONIOMETER_KAPPA rotation goniometer GONIOMETER_OMEGA 0.64279
   0 0.76604 . . .
   GONIOMETER_PHI   rotation goniometer GONIOMETER_KAPPA 1 0 0  . . .
   SOURCE          general source . 0 0 1 . . .
   GRAVITY         general gravity . 0 -1 0 . . .
   DETECTOR_Z      translation detector . 0 0 1 0 0 0
   DETECTOR_Y      translation detector DETECTOR_Z 0 1 0 0 0 0
   DETECTOR_X      translation detector DETECTOR_Y 1 0 0 0 0 0
   DETECTOR_PITCH  rotation    detector DETECTOR_X 0 1 0 0 0 0
   ELEMENT_X       translation detector DETECTOR_PITCH  1 0 0 172.43 -172.43 0
   ELEMENT_Y       translation detector ELEMENT_X  0 1 0 0 0 0
```

An Introduction to NeXus/HDF5 Nxmx

Where CBF is table oriented, NeXus/HDF5 is tree oriented, but they serve a similar purpose, providing a container into which to put data a metadata. HDF5 is favored by large facilities to handle very large numbers of datasets efficiently.

For MX NeXus/HDF5 allows large datasets to be packaged into a smaller number of files than CBF. It has one master file for all the metadata, and additional data files with multiple images in each data file.

The application definition for MX is in

http://download.nexusformat.org/sphinx/classes/applications/NXmx.html

The equivalent to the CBF axis definitions are in

http://download.nexusformat.org/sphinx/classes/base_classes/NXtransformations.html#nxtransformations

R·I·T

BROOKHAVEN NATIONAL LABORATORY

Implications for Archiving Raw Data

If we adopt interoperable, documented standards for archived data, we increase the chances that

1. Journal referees will be able to make effective use of archived images
2. Methods developers will be able to test new processing algorithms against a large base of archived data
3. Interested researchers will be able to mine archived data for common processing
4. That users will be better able to process or reprocess their own data
5. That this valuable mass of data may still be of value decades later

Questions?

R·I·T

BROOKHAVEN
NATIONAL LABORATORY