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

Herbert J. Bernstein[1]
[1]Rochester Institute of Technology, Rochester, NY, USA;

As the rate of production of diffraction images rises to several hundred datasets per day per beamline, it is becoming increasingly important to record essential metadata in an efficiently retrievable form. It is impractical to expect to refer to laboratory notebooks and do manual metadata entry in such an environment. Indeed, as data rates increase further it will become impractical to handle the same images multiple times in order to transform metadata from one convention to another. 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. In this talk we review compromises between a common minimal set of metadata to allow for processing of simple rotation data and richer sets of metadata needed for more demanding experiments. We also consider the implications of these choices for future reprocessing of archived datasets.

[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.