Improved access to raw diffraction data and its impact on crystallographic education and teaching

John R. Helliwell[1], Brian McMahon[2]

1. School of Chemistry, University of Manchester, Manchester, M13 9PL, UK
2. IUCr 5 Abbey Square, Chester CH1 2HU, UK

email: john.helliwell@manchester.ac.uk

Raw diffraction images are expected to be useful for improving numerous research applications that crystallographers engage with. There are two main aspects to this 'raw data revolution' relevant to education. Firstly, researchers will learn the protocols associated with archived, open, raw diffraction data, to accompany their processed diffraction data and derived coordinates that are expected to be found in the curated databases. Also funding agencies are looking at an Open Science protocol to improve speed of discovery for societal challenges, from the start of a funded project. Secondly routine access to raw data will impact crystallographic teaching such as for our important European Crystallography School. For the archiving of our raw data the IUCr global Diffraction Data Deposition Working Group has examined the issues and prospects for linking raw diffraction data sets to publications. Important strides have been made in the structural biology area and by the ICDD for powder diffraction data. Long time pioneers of raw diffraction data archiving are at the National Crystallography Service at Southampton University UK. At the neutron and synchrotron facilities major pioneering efforts include assigning digital object identifiers (DOIs) to all data sets by the facility, across all techniques; this is being done at ISIS and now also at the ESRF. At the PDB the submission of an entry now routinely includes the opportunity for depositors to provide details of raw data DOIs. Another major development is that IUCr Journals (e.g. IUCrJ, J. Appl. Cryst., Acta Crystallographica D, F) have started linking their publications with raw diffraction data sets in the data repositories. In our own research, as an example, some thirty crystal structure studies on the binding of anti-cancer compounds (platins) to histidine in a protein (Tanley et al. (2012). Acta Cryst. F68, 1300-1306 have the associated raw data deposited at the University of Manchester eScholar repository (doi:10.15127/1.215887, https://www.escholar.manchester.ac.uk/uk-ac-man-scw:215887) so that there is Gold Open Access to the publications, the PDB files and the raw data. These form a complete set of our research in this area. Other colleagues in this specific medically oriented research topic, and who are providing raw diffraction data on the platins binding to proteins and nucleic acids, have been invited to contribute to this data archive.

Keywords: Raw diffraction data; Data archiving; Open Science; Open Data.