Introduction to the DDDWG 2017 Workshop on Research Data Management

John R Helliwell[1] and Brian McMahon[2]

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

The IUCr Executive Committee established a Diffraction Data Deposition Working Group (DDDWG), to define the need for and practicalities of routine deposition of primary experimental data in X-ray diffraction and related experiments. Since the Working Group's first Workshop in Bergen, Norway (August 2012), important strides have been taken to make routine deposition of raw data a reality. The major facilitator for this has been the establishment of digital data storage repositories registered to issue persistent unique Digital Object Identifiers (DOIs) for a raw dataset. Such repositories include universities (*e.g.* University of Manchester), the EU's Zenodo initiative, and several centralised neutron, synchrotron and X-ray laser facilities. As stressed by John Westbrook of the PDB (http://www.iucr.org/resources/data/dddwg/bergen-workshop) metadata descriptors for raw data are vital for its effective re-use. The PDB has extensive experience of specifying metadata descriptors for structure factors, coordinates and *B* factors, as well as for cryoEM and bioNMR data depositions. Kroon-Batenburg and Helliwell [1] provided an example of appropriate metadata, critically including a picture of their diffractometer, for their local raw diffraction data archive. This archive has seen successful examples of raw data re-use such as by Wladek Minor and collaborators [2]. A second DDDWG workshop on 'Metadata for Raw Data' (Rovinj, Croatia, August 2015) brought together another wide range of global experts (http://www.iucr.org/resources/data/dddwg/rovinj-workshop), including the Chair of the IUCr Committee for the Maintenance of the CIF Standard (James Hester), who has vast experience of metadata descriptors for processed and derived data. An outcome of the second Workshop was '*checkCIF* for raw diffraction data', a notional service akin to the existing IUCr *checkCIF* for processed structure factors and derived atomic coordinates data (http://checkcif.iucr.org/). This third Workshop at ACA 2017, New Orleans, broadly titled 'Research Data Management', includes the charge to Workshop participants to focus on metadata (including their experiences with processed structure factors and derived atomic coordinates data), and help to define as closely as possible the optimum metadata for raw diffraction data to guide the raw data archives listed above. Re-use of raw data leveraged upon metadata descriptions has already been shown to be viable [1,2]. This should now be built on more energetically by the single-crystal diffraction community (including chemical crystallography), as well as by the various scattering, diffraction, imaging and spectroscopy techniques represented in the various IUCr Commissions. Excellent headway has been made in defining SAXS and EXAFS metadata, for example. For an overview of raw diffraction data preservation and re-use including an update on practicalities and metadata requirements see the very recent publication by Kroon-Batenburg et al 2017 [3].

1 *Kroon-Batenburg, L. M. J. & Helliwell, J. R. (2014). Acta Cryst. D70, 2502–2509.*
2 *Shabalin, I., Dauter, Z., Jaskolski, M., Minor, W. & Wlodawer, A. (2015). Acta Cryst. D71, 1965–1979.*
3 *Kroon-Batenburg, L. M. J. & Helliwell, J. R., McMahon, B. and Terwilliger, T.C. IUCrJ (2017). 4, 87–99.*