Correct recording of metadata: towards archiving and re-use of raw diffraction images in crystallography.

L.M.J. Kroon-Batenburg
Crystal and Structural Chemistry
Utrecht University
l.m.j.kroon-batenburg@uu.nl

In recent years scientists and policy makers made major steps toward Open Science. The incentive is, to allow validation and falsification of the research based on the data and to allow its re-use, as the aquisition of the data is mostly funded by the tax payer. New methods and technologies can be developed with the availiability of large data bases covering diverse types of experiments. In this framework the IUCr established a Diffraction Data Deposition Working Group (DDDWG) with the aim of developing standards for the representation of raw diffraction data in crystallography. Two key issues play a role: the importance of persistent identifiers and the full recording of meta data. Whilst discussions are vividly going on about what data to archive, only those related to published papers or also of incomplete or unsuccessful research that could be particularly interesting for the development of new science, the field should prepare itself for depositing fully self-contained data. A recent review (Kroon-Batenburg et al. 2017) summarizes the ongoing developments. Ideally, metadata should comprise the following: identification of the image format, number of pixels, pixel sizes, byte-storage architecture, baseline offset and handling of overflows, information on the corrections that are applied (dark current, distortion correction, non-uniformity correction), detector gain, goniometer axes orientations and rotation directions, and information on the experiment such as exposure time, number of repeats, oscillation axis and range, wavelength used, beam polarization, detector position (or beam position) and offsets. Details and the importance of such information will be discussed. The necessity to use a structured language (DDL) that defines data names (tags) in data formats like CIF or Nexus defines data names (tags) in data formats like CIF or Nexus (Bernstein, 2015) to ensure unambiguous interpretation, will be demonstrated. Awareness of detector manufacturers and experimentalists of recording sufficient metadata is essential, and guide lines for these are under way.

Kroon-Batenburg, L.M.J., Helliwell, J.R., McMahon, B & Terwilleger, T.C. (2017). *IUCrJ* **4**, 1-13.

Bernstein, H.J., DDDWG Workshop (2015). http://www.iucr.org/resources/data/dddwg/rovinj-workshop