Metadata and checkCIF for raw diffraction data and their role in realising crystallographic science objectivity

L.M.J Kroon-Batenburg

Structural Biochemistry, Faculty of Science, Utrecht University, Padualaan 99, 3584 CG Utrecht, The Netherlands

l.m.j.kroon-batenburg@uu.nl

The 'Open Science' model is based on open access to published scientific results and on sharing scientific data according to the FAIR (Findable, Accessible, Interoperable and Re-usable) principles. The importance of archiving raw data underpinning crystallographic research is well appreciated in recent years as demonstrated and emphasized by the report of the IUCr working group Diffraction Data Deposition Working Group (DDDWG) [1] and has now become feasible as the technology of disk storage has advanced enormously [2]. Crystallography as a research community has always been at the forefront of data sharing, firstly with atomic coordinates (derived data) and secondly with processed diffraction data. The next step, archiving of raw data, has the challenge of providing adequate metadata and the need for community agreed checks, similar to the checkCIF approach, to ensure adherence to the FAIR principles. Several generic and specific to crystallography raw data archives are available. IUCr Journals are now encouraging authors to provide a doi for their deposited original raw diffraction data when they submit an article describing a new structure or a new method tested on unpublished diffraction data. The Protein Data Bank (PDB) also asks for the DOI (digital object identifier), when available, for raw data and metadata for raw data during a deposition. Additionally, the Protein Data Bank Japan has set up the X-ray Diffraction Archive XRD-Arc where authors can submit their raw diffraction data corresponding to PDB entries. The current situation with respect to metadata, important for the future (re-)use of raw diffraction data, will be scrutinised in the light of the FAIR principles. A recent notable effort is the HDRMX community has agreed on a Gold Standard specification for high-rate diffraction data [3]. A leap forward in checking the completeness and validity of metadata by a CheckCIF approach will be discussed.

[1] Final report of the DDDWG https://forums.iucr.org/viewtopic.php?f=21&t=396

[2] L.M.J. Kroon-Batenburg, J.R. Helliwell, B. McMahon, T.C. Terwilliger, IUCrJ (2017). 4, 87–99, https://doi.org/10.1107/S2052252516018315

[3] H. J. Bernstein, A. Förster, A. Bhowmick, A. S. Brewster, S. Brockhauser, L. Gelisio, D. R. Hall, F. Leonarski, V. Mariani, G. Santoni, C. Vonrhein and G. Winter, IUCrJ (2020). 7, 784–792, https://doi.org/10.1107/S2052252520008672

Keywords: raw diffraction data; metadata; FAIR