MS107.O02

Maintaining and obtaining maximum value from a CIF publication archive

Brian McMahon[1], Michael A. Hoyland[1], Simon P. Westrip[2]

[1]*IUCr, Chester, United Kingdom,* [2]*The Walled Garden, Horton Green, Cheshire, United Kingdom*
E-mail: bm@iucr.org

The journals published by the International Union of Crystallography (IUCr) required authors to deposit structure models as Crystallographic Information Files (CIF) [1] soon after the CIF standard was adopted [2]; subsequently structure factors were also required in CIF format. Some IUCr journals accept or mandate the publication of complete research articles, including text and references, as CIF. Although CIF was intended from the outset as an archival format, during the course of the last quarter century there have been inevitable examples of inconsistency or error in the way that individual data items have been interpreted by different authors or software packages. Further, as scientific practice evolves, certain terms take on new meanings, new experimental parameters need to be defined and recorded, and new techniques contribute to a structure determination experiment, sometimes using terminology which conflicts with that in established CIF dictionaries. Dictionaries managed by COMCIFS (the IUCr Committee responsible for the CIF standard) have been developed in a well-disciplined manner, with issue date stamps, internal linking to alternative or deprecated data representations, and registration of reserved namespace components for data names. This presentation will address the ways in which CIF dictionaries can be leveraged to flexibly handle variation in preferred data names over time, and will consider the possibility of using methods in the new generation of enhanced DDLm dictionaries to extract implicit data from CIFs.

[1] Allen, F. H., Bugg, C. E. & Maslen, E. N. (1991). Acta Cryst. C47, 2263-2265.

[2] Hall, S. R., Allen, F. H. & Brown, I. D. (1991). Acta Cryst. A47, 655-685.

Keywords: crystallographic information framework, data integrity, data archive