

MS11-04 | KEEPING THINGS 'N SYNCH: ANALYSING THE CONTENT AND COMPLETENESS OF CIF METADATA IN THE CSD

Johnson, Natalie (Cambridge Crystallographic Data Centre, Cambridge, GBR); Wiggin, Seth (Cambridge Crystallographic Data Centre, Cambridge, GBR); Ward, Suzanna (Cambridge Crystallographic Data Centre, Cambridge, GBR)

Alongside structural information, Crystallographic Information Files (CIF)s can store metadata detailing how the crystal data was collected and processed [1]. Complete and accurate experimental metadata is important as it provides additional insights about the experiment [2] and can be useful in a variety of cases. Databases of crystal structure information could be mined for specific material or to assess trends within a field. One such database is the Cambridge Structural Database (CSD), containing 1 million small-molecule organic and metal organic crystal structures, which has an archive of over 750,000 underlying CIFs. Scripts can be written to utilise the CSD Python API [3] to examine the completeness and content of certain CIF attributes.

CIF metadata has been used to assess the possibility of attributing structures identified as being collected using synchrotron radiation to particular facilities, or even specific beamlines. Data from a number of CIF fields was probed for '*facility identifying information*'. For facilities with at least 100 attributed single crystal structures, over 90% of the structures can be attributed to a named beamline – although percentages varied by facility. This information could be used to provide increased traceability of data measured at synchrotron facilities and help establish community guidelines.

[1] Hall, S R et al., *Acta Cryst.*, 1991, **A47**, 665-685.

[2] Kroon-Batenburg, L M J, et al., *IUCrJ*, 2017, **4**, 87-99.

[3] Groom, C R, et al., *Acta Cryst.*, 2016, **B72**, 171-179.