**INTRODUCTION**

The Cambridge Structural Database (CSD) is a fully validated database of organic and metal-organic crystal structures with > 550,000 entries. Alongside curating new structures, each year existing entries are enhanced with additional information or to improve the consistency of search terms. One planned enhancement is to ensure proper categorisation of synchrotron studies. To do this, data deposited to the CSD in Crystallographic Information File (CIF) format (~ 750,000 structures) was investigated using the CSD Python API for “synchrotron identifying information”.

As well as detecting additional synchrotron studies, the possibility of attributing these studies to particular facilities was investigated. Large facilities desire the ability to locate data measured on-site after it has been published, as research outcomes are not always reported back to them. There have been several projects in recent years focusing on the tracing of data in published research.\(^2\)

### Identification of synchrotron studies

A total of 10.095 structures were identified using the API script, including ~ 500 that were not previously flagged within the CSD.\(^3\) The field that contained “synchrotron identifying information” varied between CIFs; Table 1 shows the number of synchrotron structures that had identifying information within certain attributes. Similar CIF attributes are grouped together: _diffm_measurement_device/_diffm_measurement_device_type and _diffm_source/_diffm_radiation_source.

### Identification of synchrotron facilities

A facility could be identified for 64% of structures, measured on 28 individual synchrotrons from across the world, indicating that the information is present in the majority of deposited CIFs. Figure 1 shows the number of unknown facilities by year. There is a generally increasing trend as the total number of structures increases, however, the % of structures where the facility is unknown is generally constant. An example of statistics which can be created for a facility is shown in Figure 2 – the number of structures appearing in the CSD attributed to the Synchrotron Radiation Source (SRS, Daresbury, UK) by year. The trend in the increase of structures up to 2008 mirrors the uptake of the CIF format, the numbers then decline after the synchrotron was decommissioned. The fields in which “facility identifying information” was most commonly identified are displayed in Table 2. This shows that information can be found in a variety of places.

### Identification of beamlines

Initial estimates show that > 45% of synchrotron studies also reported the beamline that was used. The % of structures with reported beamlines varies for each facility. Table 3 shows statistics for the SRS – where the beamline is reported for 97% of structures.

**CONCLUSIONS**

This poster demonstrates that a search of CIF attributes could be an effective mechanism to identify facilities for data deposited in the CSD. This work additionally highlights the importance of including synchrotron and facility identifying information within CIFs in order to appropriately classify studies and to provide increased traceability for facilities. This could lead to the formation of new CIF guidelines in the synchrotron community.

**ACKNOWLEDGEMENTS**

We thank John Halliwell, Mike Hoyland, Mark Waren, Brian McMahon, Jason Price, Amy Sarjeant and Peter Strickland for helpful discussions relating to this work.

---

3. Using CIF version 5.40 (November 2018)