Synchrotron data in the CSD

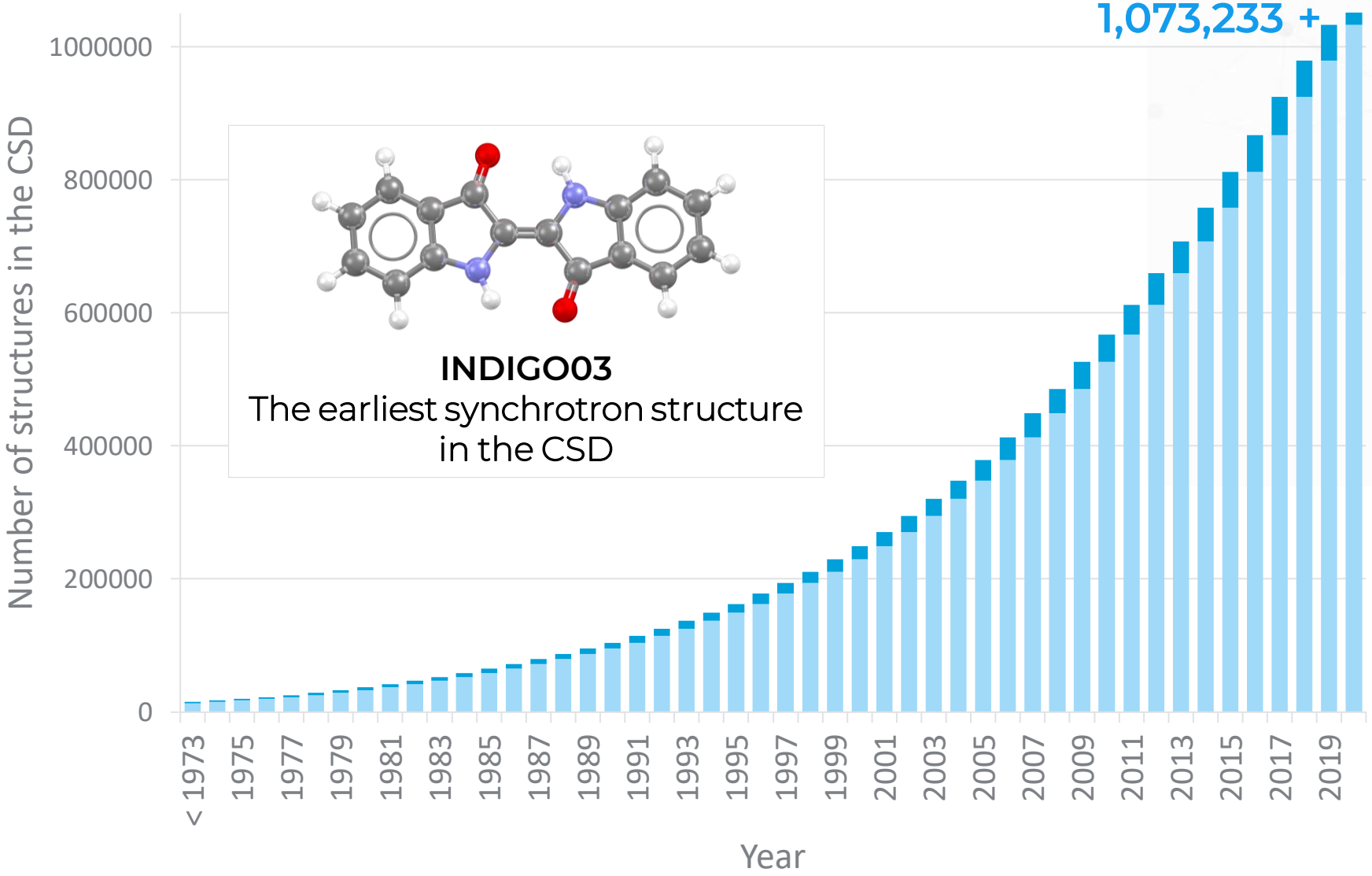
Natalie Johnson

Data Integrity Research Scientist

22nd August 2020

advancing structural science

The Cambridge Structural Database (CSD)



INDIGO03
The earliest synchrotron structure in the CSD

The CSD is a database of small molecule organic and metal-organic crystal structures.

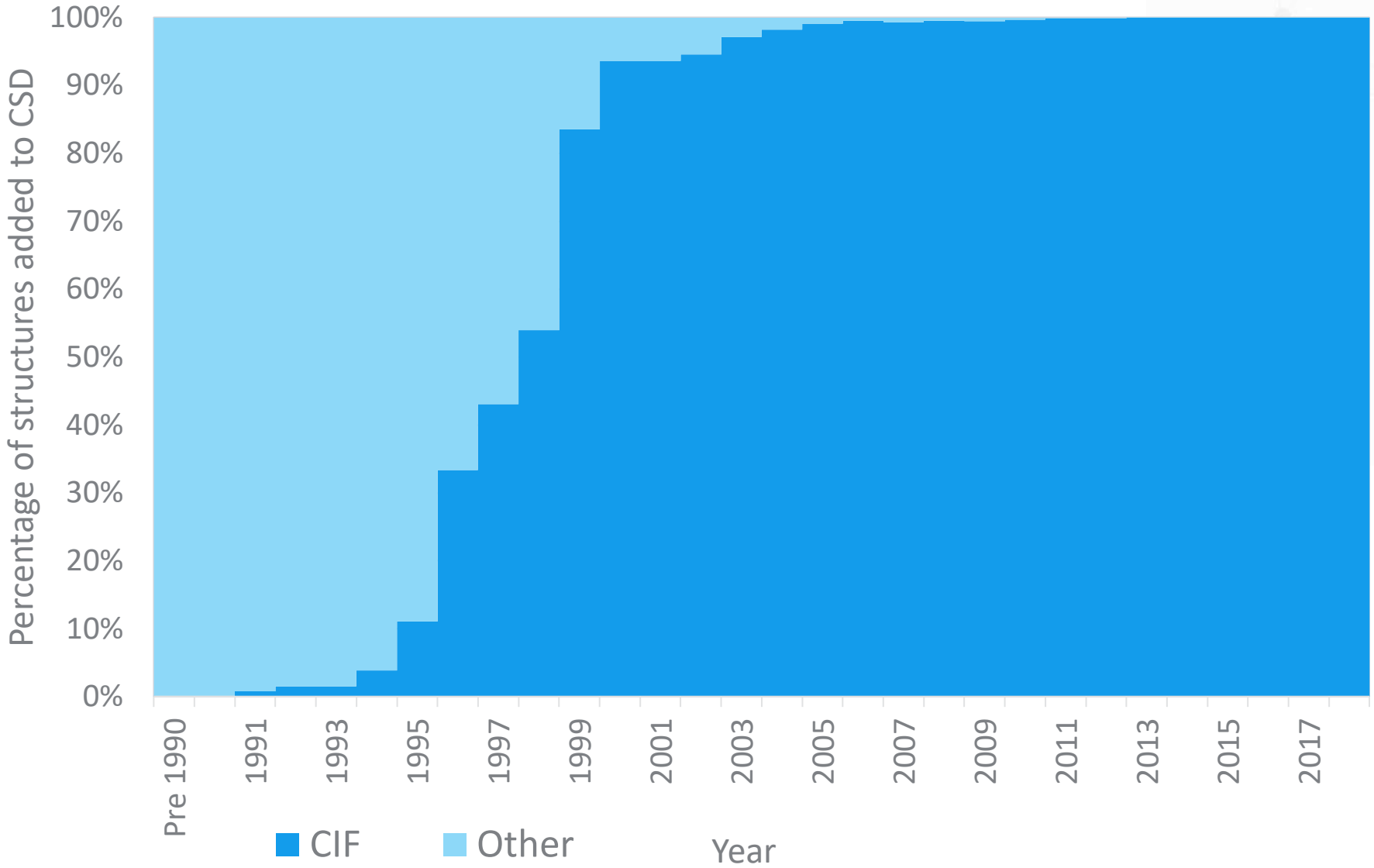
Every entry enriched and annotated by experts.

Every published structure:

- Inc. ASAP & early view
- *CSD Communications*
- Patents
- University repositories

CCDC

The Cambridge Structural Database (CSD)



Data can be deposited to CCDC through an online deposition platform or via email.

Pre-CIF, entries were created by typing data from tables in papers.

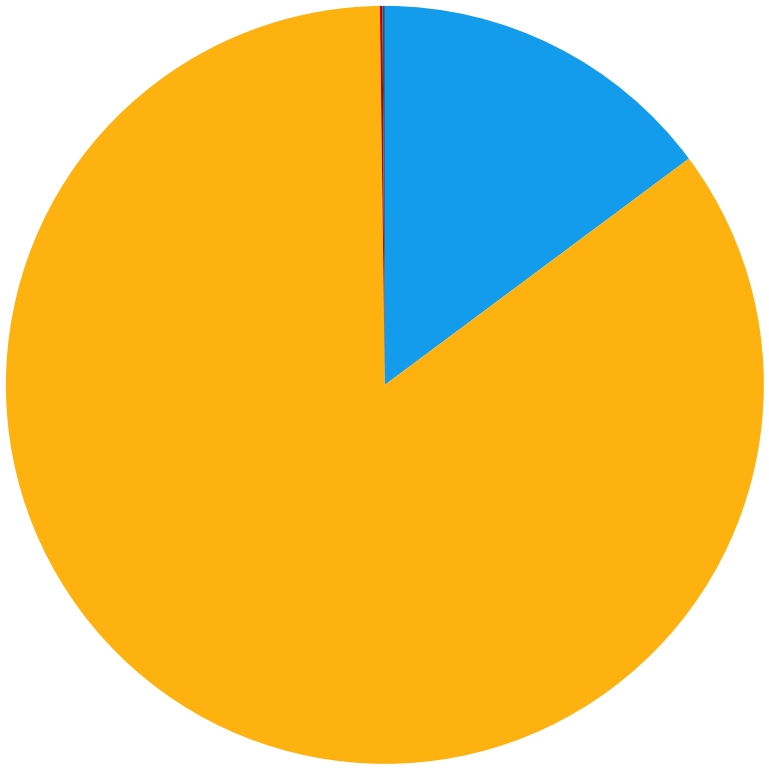
When CIF is not available, CCDC manually creates a CIF from the accessible information.

365 structures (< 0.6%) added to the CSD with a manually created CIF in 2019.

CCDC

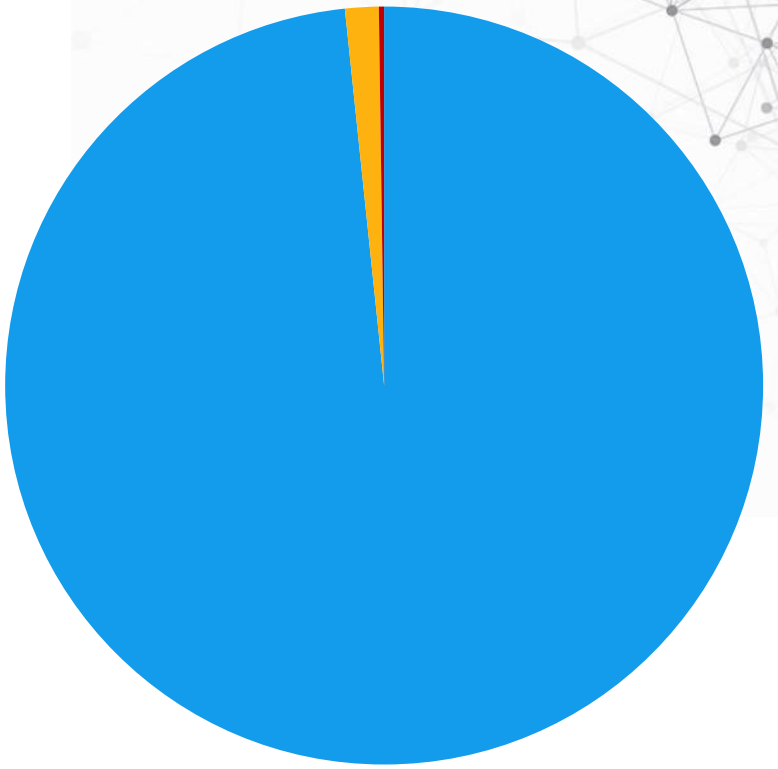
Synchrotron data in the PDB vs CSD

PDB



Lab X-ray Synchrotron Neutron Electron

CSD

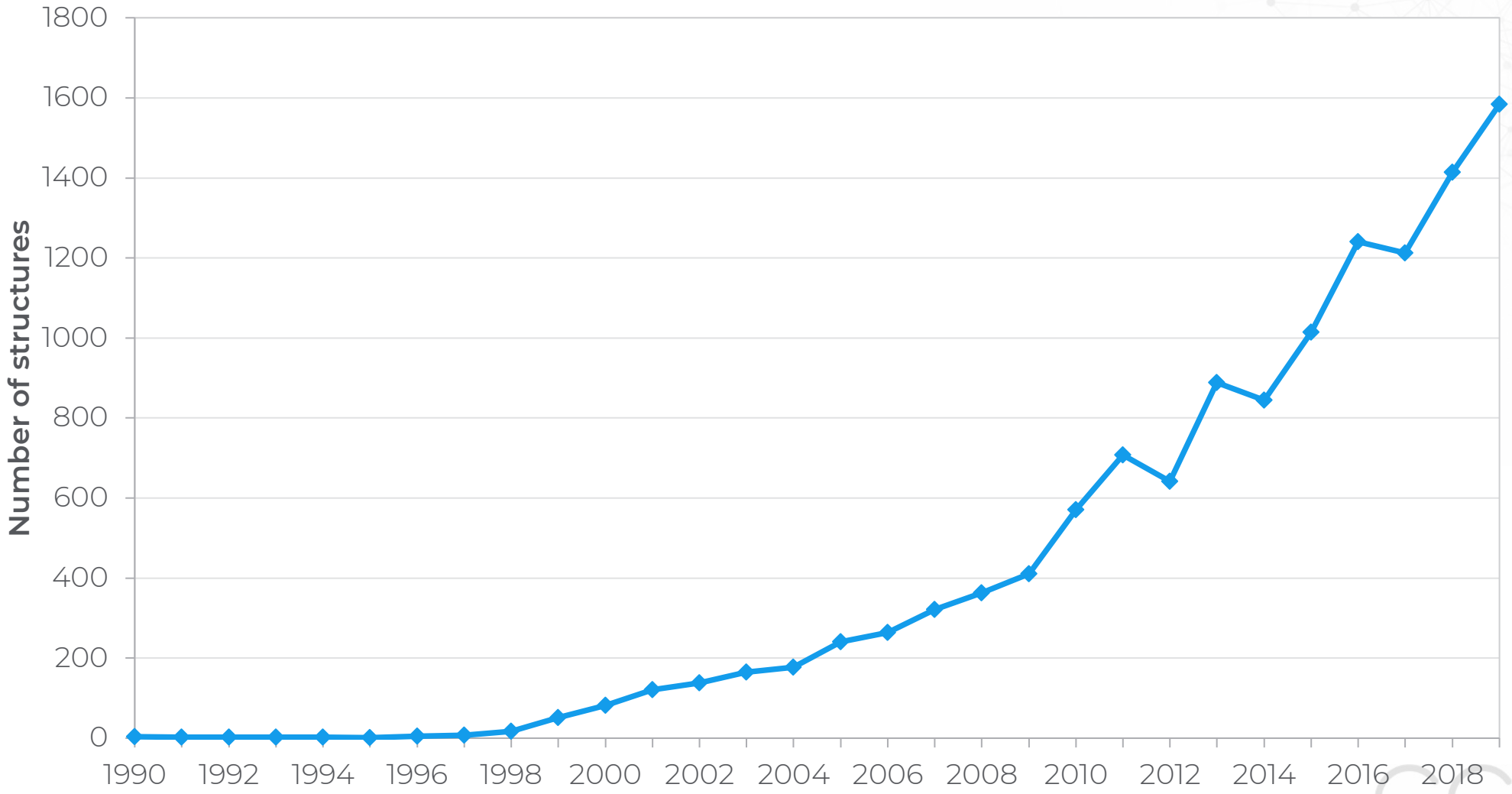


Lab X-ray Synchrotron Neutron Electron

1 PDB Statistics from Biosync (http://biosync.sbkb.org/) and PDB website, accessed 25/06/2020
Diffraction data only.

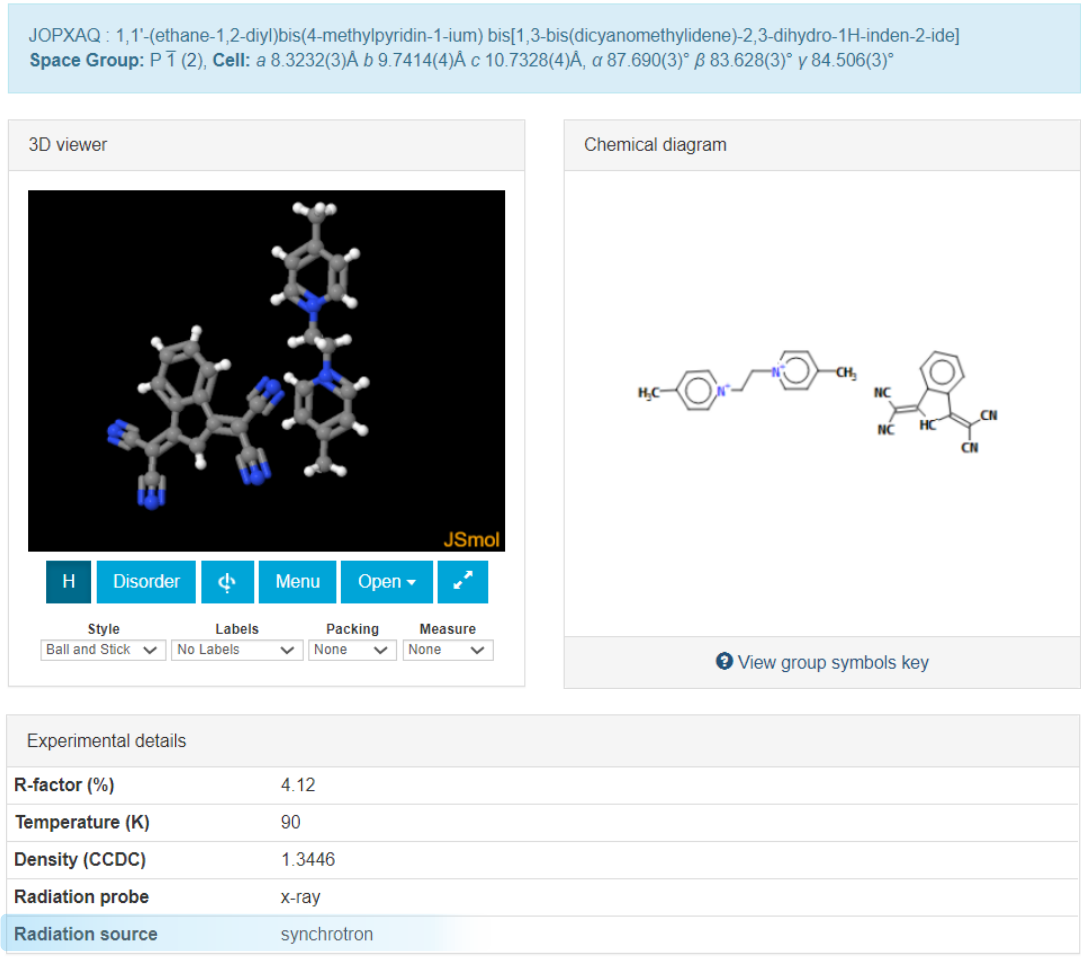
CCDC

Synchrotron data in the CSD

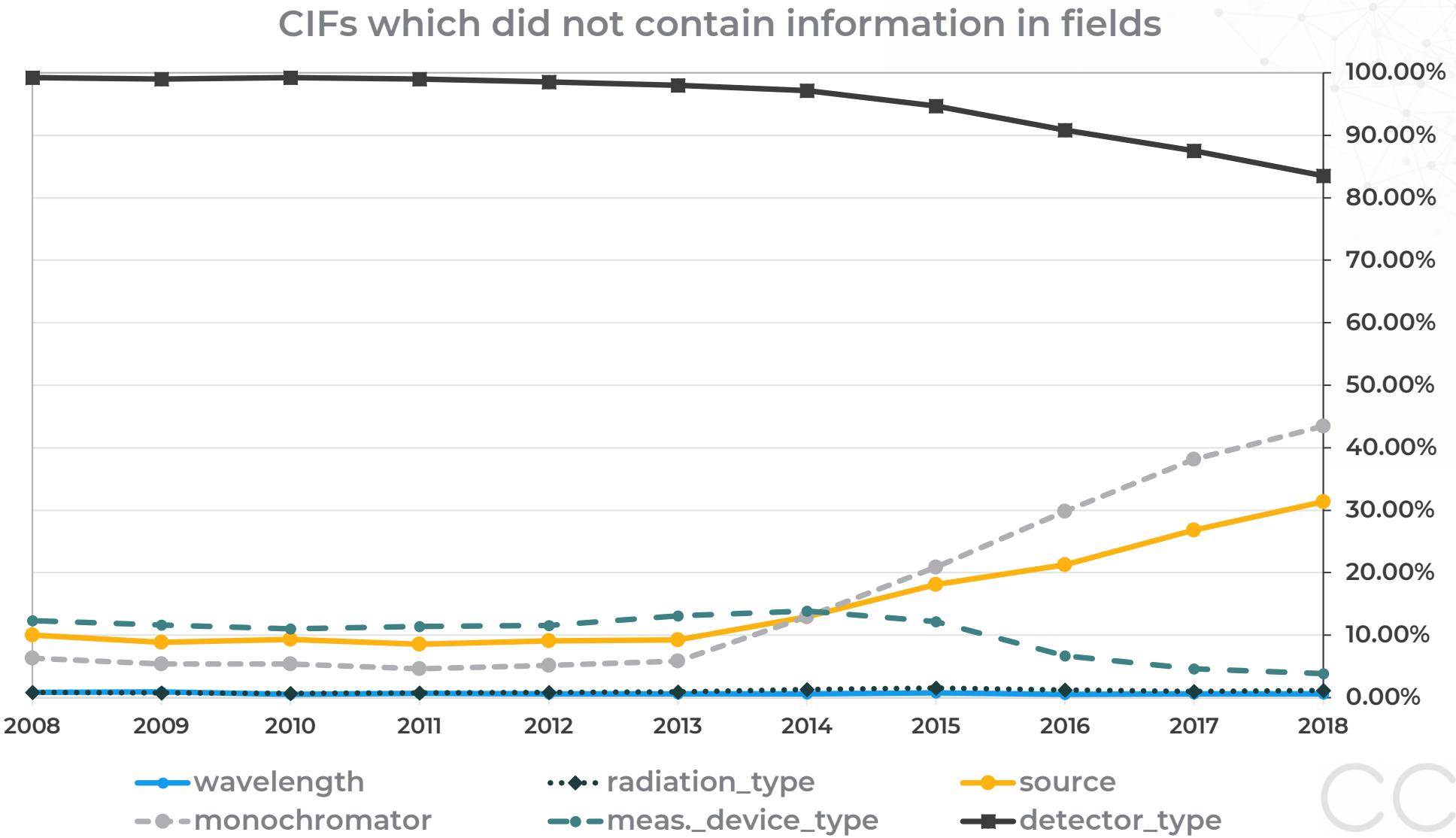


Identification of synchrotron data

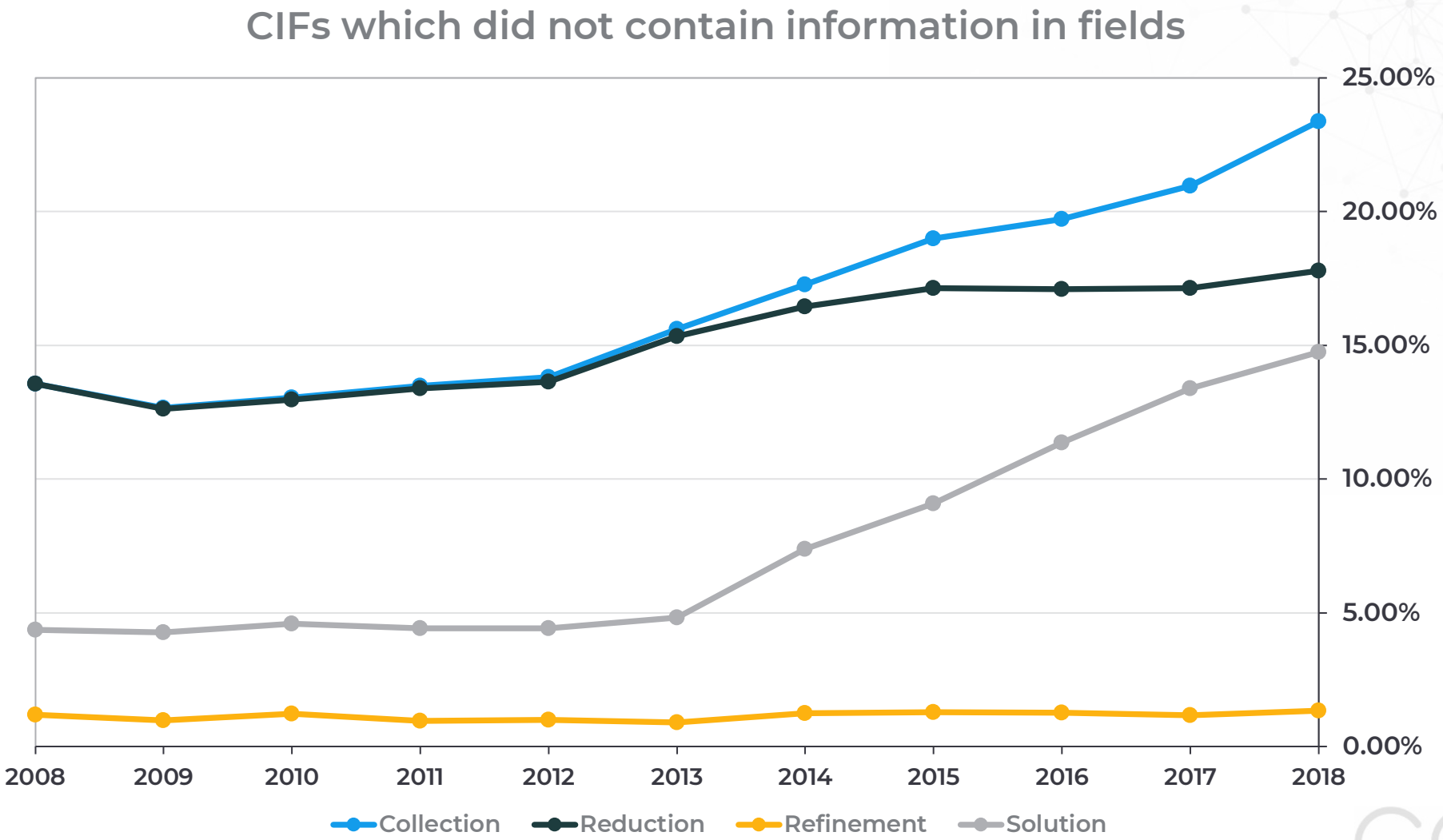
- Structures are flagged with ‘synchrotron’ in the CSD.
- These structures are identified during the data curation process either automatically by the curation software, which checks a select number of CIF fields, or manually labelled by an Editor.



CIF completeness - Experimental



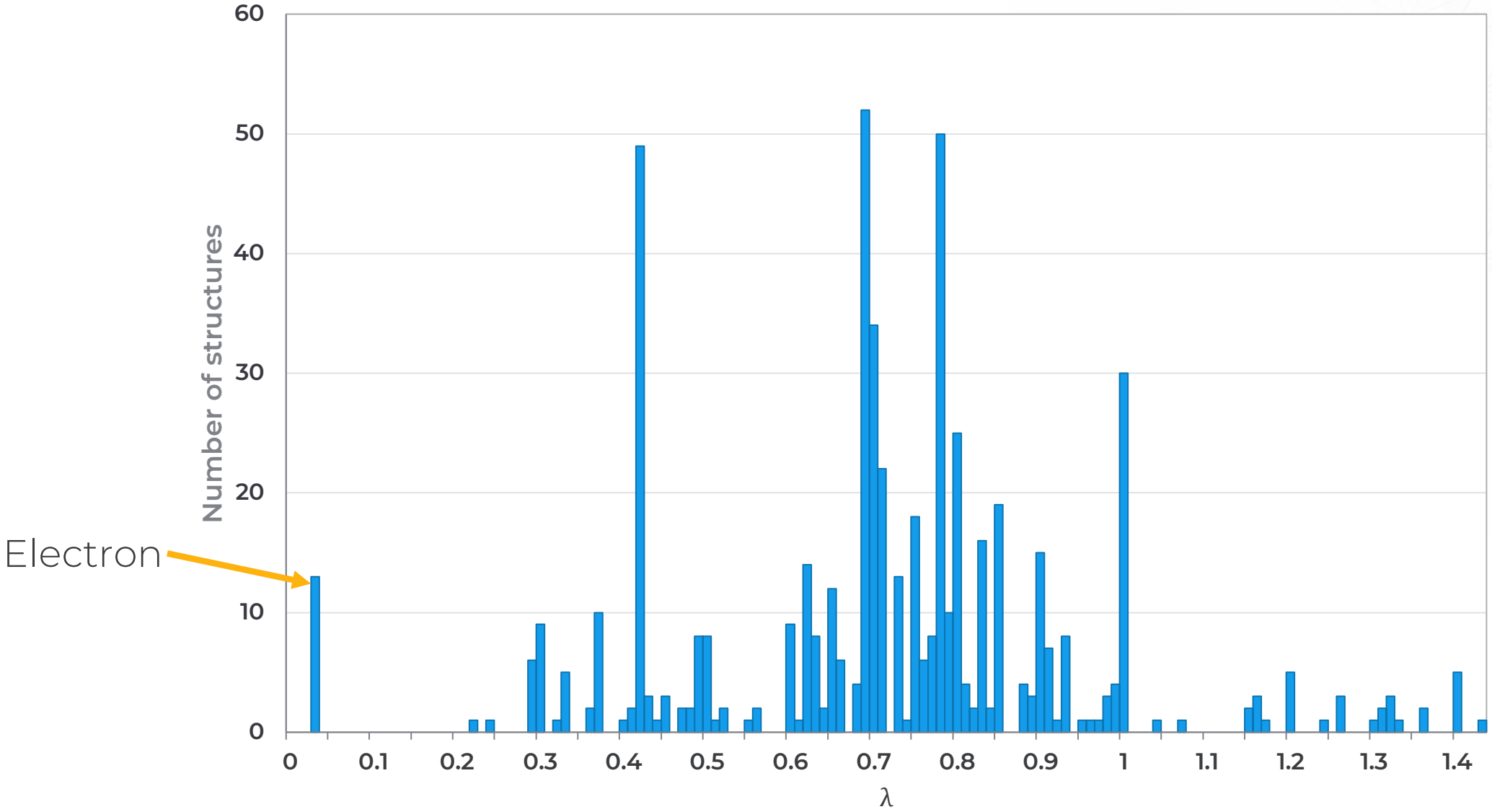
CIF completeness - Computational



CIF completeness – Outcomes

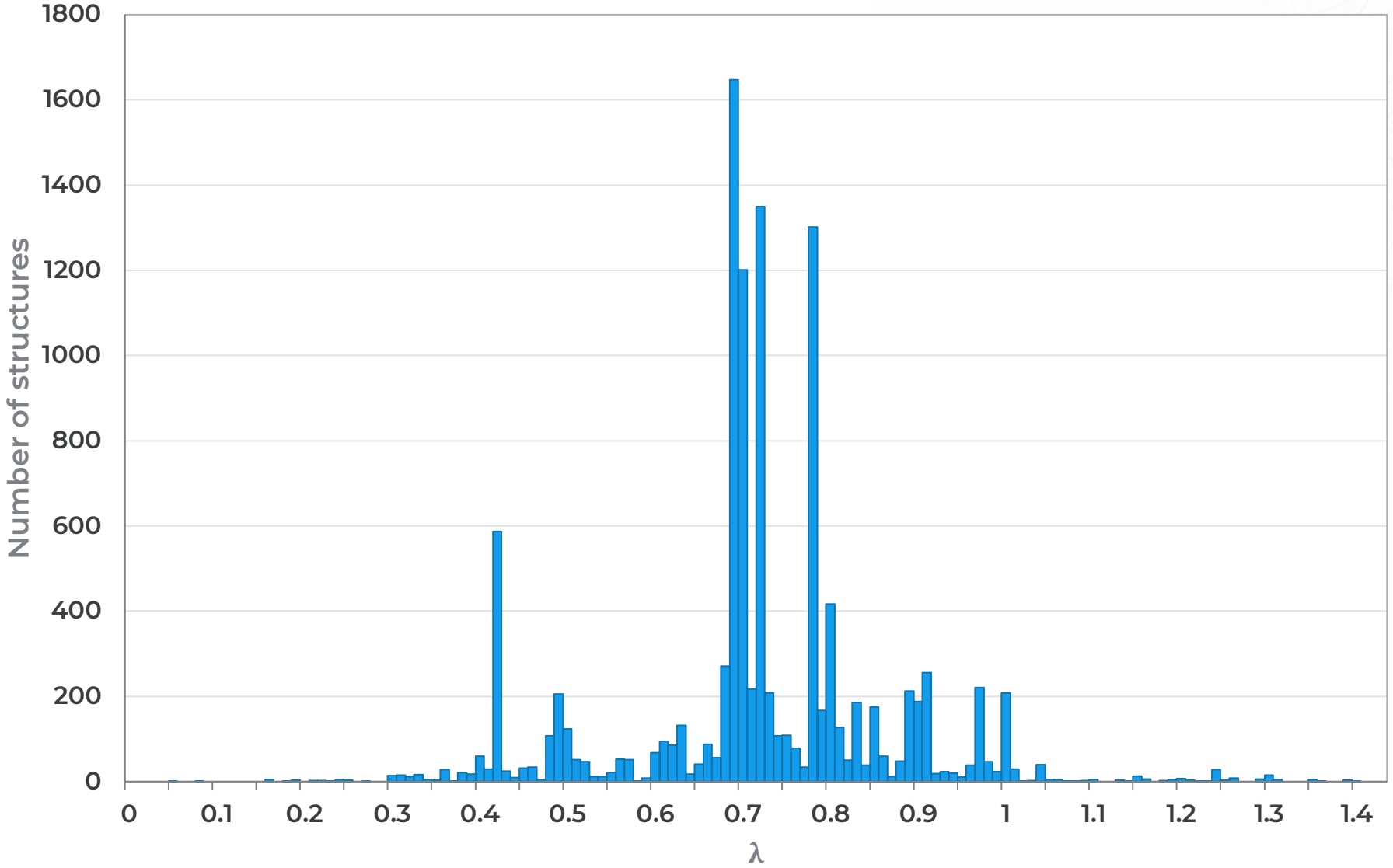
- Some structures were not identifiable from the CIF as it did not contain complete information.
  - 25 structures found by cross-referencing papers associated with the Australian Synchrotron and structures in the CSD.
  - ~ 600 structures report a ‘non-standard’ wavelength, yet contain no synchrotron identifying information.
- Other inconsistencies found in a number of synchrotron CIFs e.g. radiation_type ‘synchrotron’ and _diffrn_source ‘sealed tube’/’rotating anode’

Structures with a non-standard wavelength



CCDC

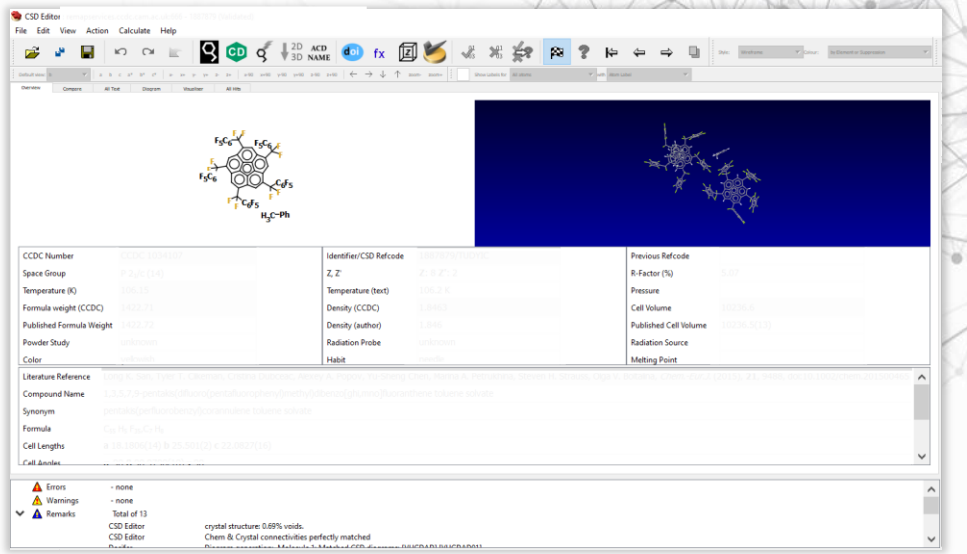
Wavelength of synchrotron structures



CCDC

FAIR and FACT small molecule data

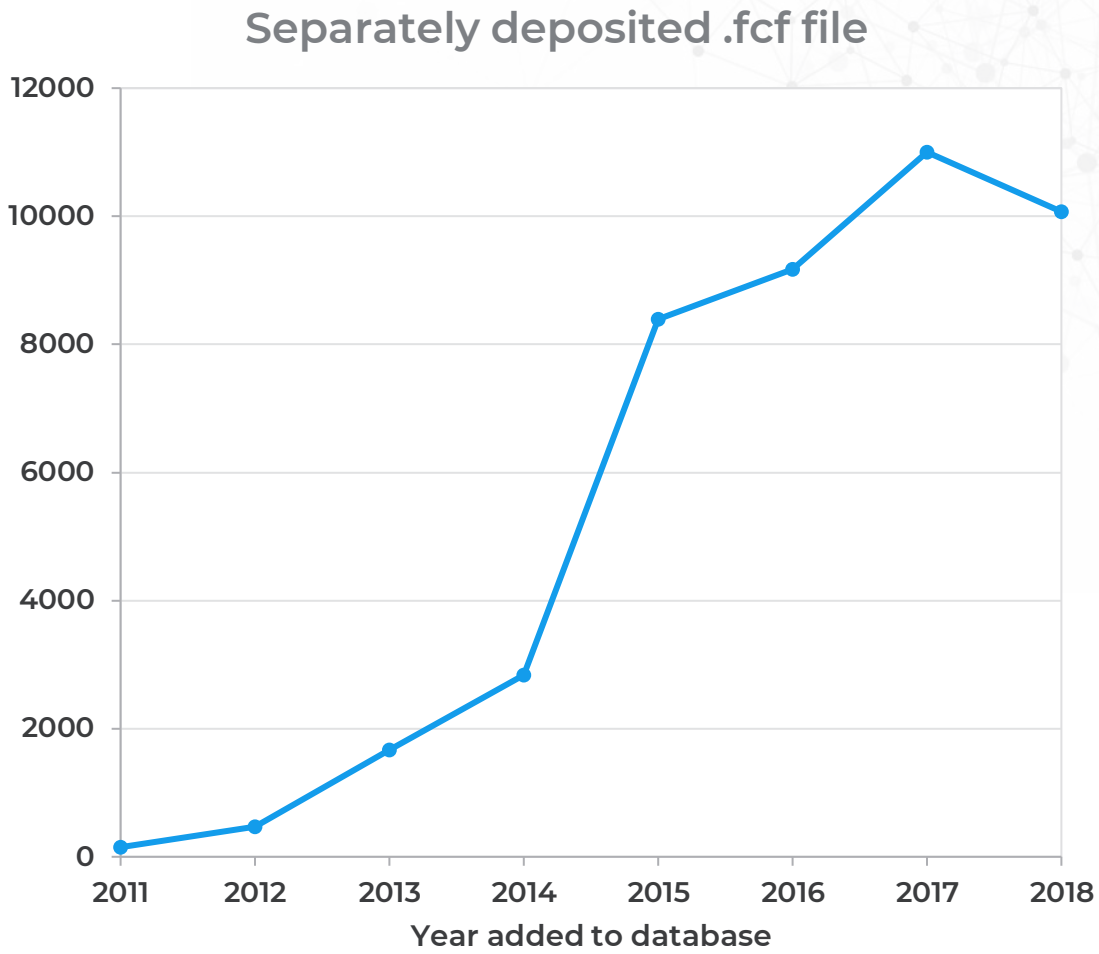
- Curation to ensure structures are labelled and presented consistently so are **findable** within the CSD.
- Option to download additional information associated with entry in WebCSD/Access Structures.



Helliwell, J. (2019). *Structural Dynamics*. 6, 054306., DOI:10.1063/1.5124439

Reflection intensities

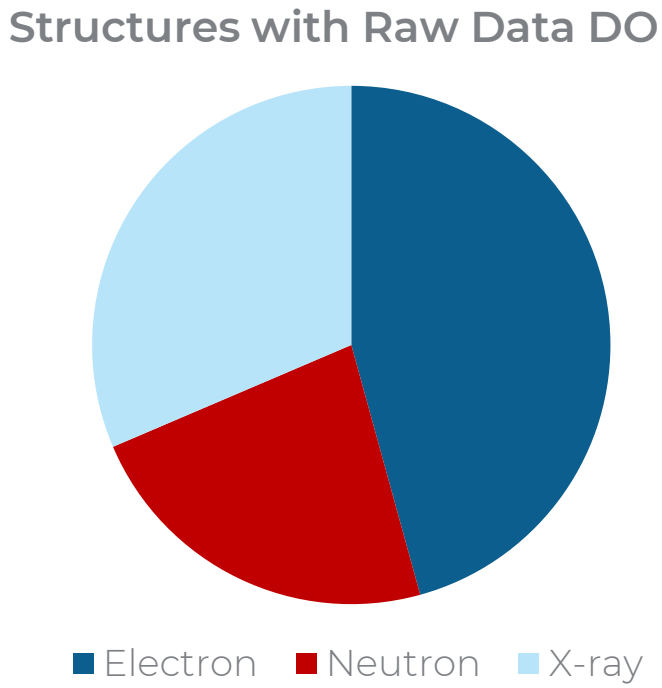
- The deposition of structure factors is not explicitly required by CCDC.
- The number of separately deposited structure factor files is tending to increase.
- Many programs automatically include hkl, res and even structure factor information within the CIF.

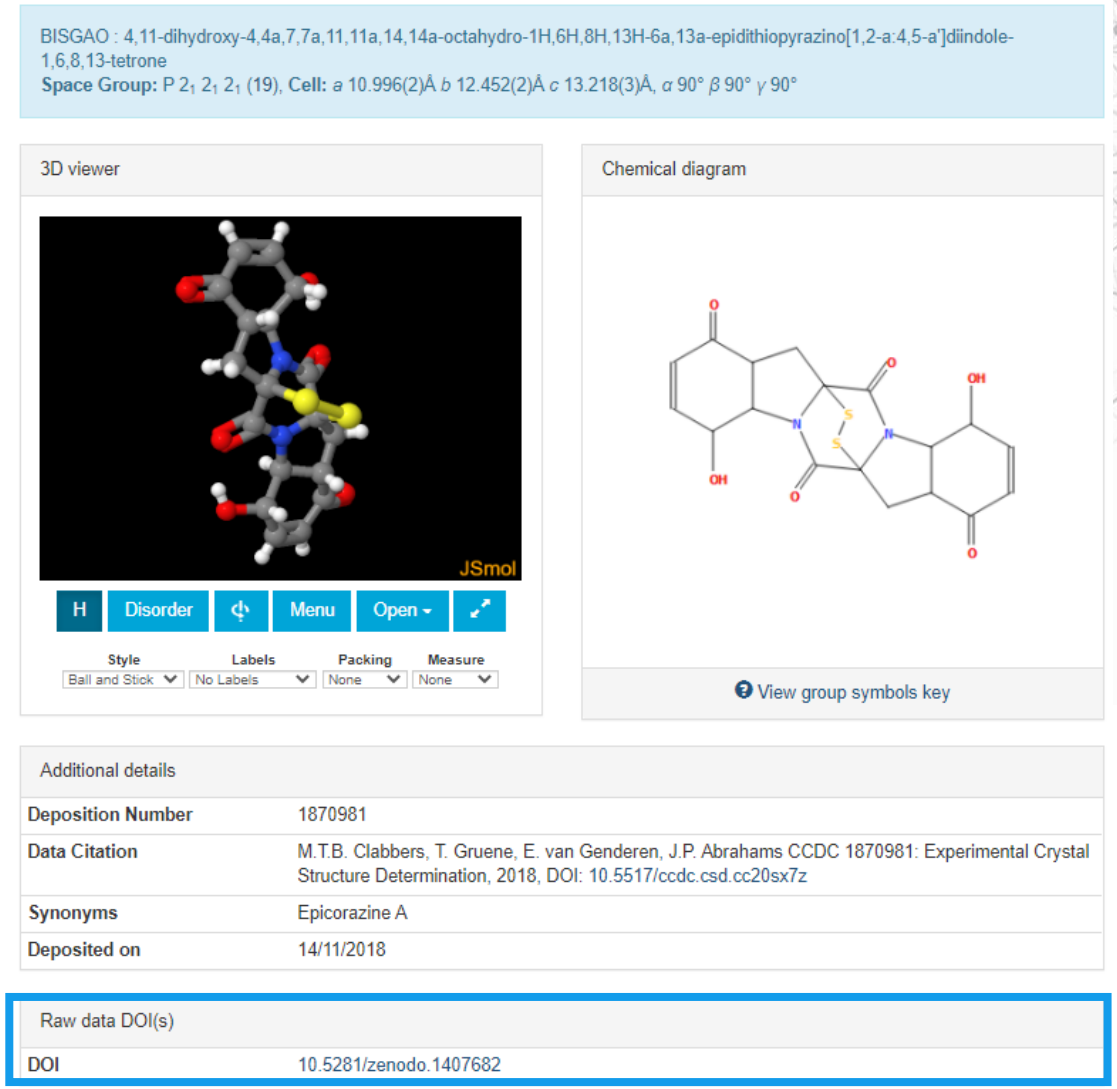


CCDC

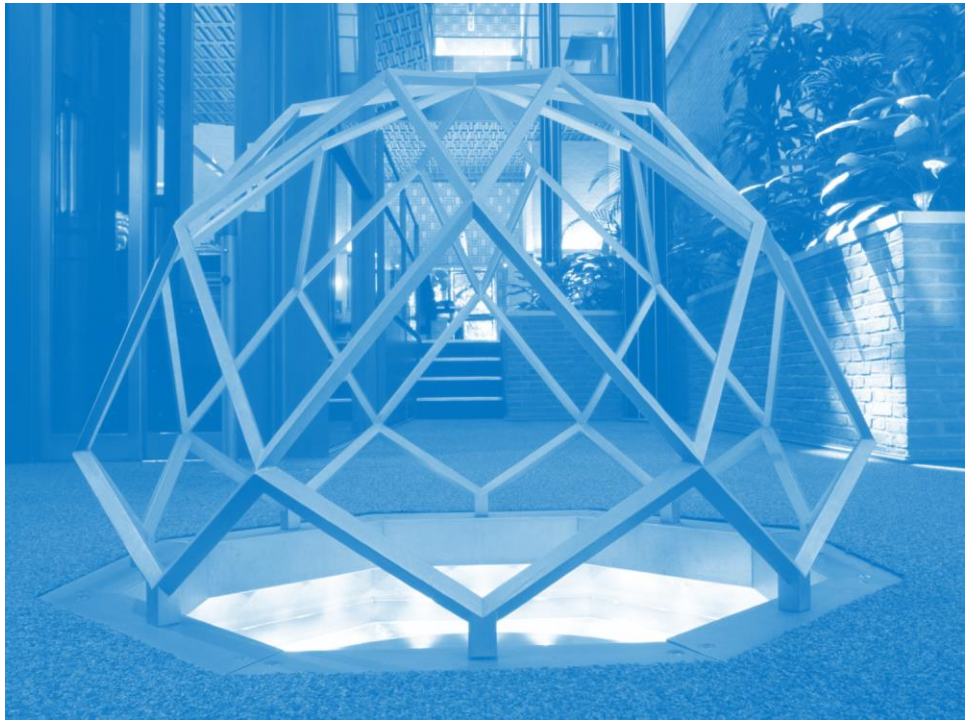
Raw Diffraction Data

- CCDC does not currently store raw diffraction data. However, a DOI for the raw data can be linked to the entry.
- 35 structures in CSD have raw data DOI.





Thank you for listening



Acknowledgements

Stephanie Bowers, **Jason Price** (Australian Synchrotron), **Mike Hoyland**, **Brian McMahon**, **Peter Strickland** (IUCr), **Vasily Bunakov**, **Brian Matthews** (STFC), **Suzanna Ward**, **Seth Wiggin** (CCDC), **Simon Coles**, **John Helliwell**, **Mark Warren**, and **Amy Sarjeant**

CCDC