

Raw data and external links in the Crystallography Open Database

A. Zeleckytė¹, A. Merkys¹, A. Vaitkus¹, S. Gražulis^{1,2}

¹Vilnius University, Life Sciences Center, Institute of Biotechnology, Saulėtekio av. 7, LT-10257 Vilnius, Lithuania, ²Vilnius University, Faculty of Mathematics and Informatics, Naugarduko 24, LT-03225 Vilnius, Lithuania

grazulis@ibt.lt

The Crystallography Open Database (COD, <https://www.crystallography.net/cod>) [1] has been openly disseminating small organic, metal-organic and inorganic (mineral) structure data for more than 20 years. With increasing computational capabilities, more data are stored in the COD structure collection. Since 2009, the COD deposits scattering intensity and structure factor data in separate files. As modern data-driven science imposes stricter requirements for data provenance, we are announcing a new function – the possibility to store links to raw data (diffraction image) in the COD CIF records, making it searchable in the COD database tables. The raw data images are typically orders of magnitude larger than coordinate and even structure factor records, thus making their storage in one place a challenge [2]. The COD has therefore adopted a policy to support any combinations of distributed raw data archives. For example, data can be stored in the XRDa [3] raw data archive. Vilnius University and the Research Council of Lithuania also support Data Center for Machine Learning and Quantum Computing in Natural and Biomedical Sciences, which can in the future also host raw data for structures that will be deposited to the COD.

The implemented COD data links are not only suitable for linking COD structures (or any other CIFs) with raw data, but can establish semantic links, using documented predicates, with other databases or data sets. Other databases are invited to collaborate on building a distributed scientific data Web.

The figure shows two screenshots. The left screenshot is the COD record for entry 3000450. It features a sidebar with navigation options like 'COD Home', 'Accessing COD Data', 'Add Your Data', and 'Documentation'. The main content area includes an 'Information card for entry 3000450' with a 'Preview' section showing a 3D ball-and-stick model of a crystal structure. Below the model are links for 'Coordinates', 'Original paper (by DOI)', and 'External links'. The right screenshot is the XRDa (Xtal Raw Data Archive) interface for entry XRD-00142. It displays 'Entry information' such as 'Entry ID', 'Entry title', 'Entry type', and 'XRD entry authors'. It also provides links for 'Download', 'XRDa entry DOI', 'CCDC entry', 'Crystallography Open Database entry', and 'Citation DOI'.

Figure 1. An example of a COD record linked with the XRDa raw image set, for the entry from the T. Sasaki et al. [4].

- [1] Gražulis, S., Daškevič, A., Merkys, A., Chateigner, D., Lutterotti, L., Quirós, M., Serebryanaya, N. R., Moeck, P., Downs, R. T., & Le Bail, A. (2011). Crystallography open database (cod): An open-access collection of crystal structures and platform for world-wide collaboration. *Nucleic Acids Research*, 40(D1), D420–D427. <https://doi.org/10.1093/nar/gkr900>.
- [2] Helliwell, J. R.; Hester, J. R.; Kroon-Batenburg, L. M. J.; McMahon, B. & Storm, S. L. S. (2024) The evolution of raw data archiving and the growth of its importance in crystallography. *IUCrJ* 11(4), 464–475. International Union of Crystallography (IUCr). DOI: <https://doi.org/10.1107/s205225252400455x>.
- [3] Bekker, G.-J.; Juniar, L. & Kurisu, G. (2021) X線画像アーカイブXRDa (X-ray image archive XRDa). URL: https://pdbj.org/cms-data/workshop/20210120/WebSeminar_2021XRDa.pdf.
- [4] Sasaki, T.; Nakane, T.; Kawamoto, A.; Zhao, Y.; Fujimoto, Y.; Nishizawa, T.; Kalita, N.; Tsuzuki, S.; Ito, F.; Ramamurty, U.; Thakuria, R. & Kurisu, G. (2024) Mechanically-sensitive fluorochromism by molecular domino transformation in a Schiff base crystal. *Journal of Materials Chemistry C* 12(23), 8508–8513. Royal Society of Chemistry (RSC). DOI: <https://doi.org/10.1039/d4tc00406j>.