Integrating data sources in CCP4 Cloud: A complete online platform for structure determination

Jools Wills, Andrey Lebedev, Maria Fando, Ville Uski, Charles Ballard, Ronan Keegan, and Eugene Krissinel

Scientific Computing Department, Science and Technology Facilities Council UK, Didcot, OX11 0FA, United Kingdom

jools.wills@stfc.ac.uk

Over the past decade, Macromolecular Crystallography (MX) has increasingly embraced online solutions across both experimental and computational stages of structure determination. This online workflow typically begins with shipping crystals to a synchrotron facility, followed by remote data collection, then progresses through data processing and structure solution using web-based software, and concludes with deposition in the Protein Data Bank (PDB).

The online approach offers numerous advantages, such as enhanced collaboration, freedom from geographic constraints, improved data security, and reliable long-term data management. Reflecting these trends, CCP4 has developed CCP4 Cloud [1] – a comprehensive platform supporting the entire crystallographic computation workflow, from diffraction image processing to structure refinement and preparation of the PDB deposition package.

In an ideal online workflow, researchers would send their samples to a synchrotron and, upon completion of the diffraction experiment, seamlessly receive the collected data in their CCP4 Cloud accounts for further processing and structure solution. To enable this, we have developed DataLink – a file and database storage service with a web-based API designed to retrieve raw diffraction image data from a variety of sources. These include online repositories such as SBGrid, Protein Diffraction, Zenodo, and XRDa, as well as potential direct integrations with synchrotrons and laboratory diffractometers.

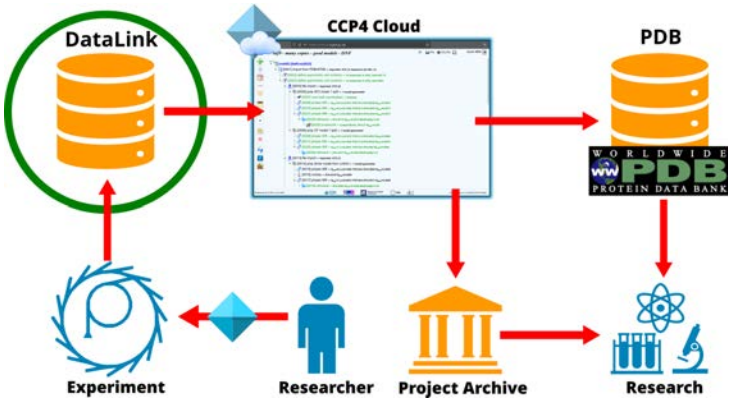

Figure 1. Diagram showing how DataLink fits within the CCP4 Cloud MX setup.

In this presentation, we showcase the implementation and key features of DataLink, with a focus on its deployment at the Diamond Light Source. DataLink is included in CCP4 Software Series 9 as part of the standard CCP4 setup. This presentation will be of interest to MX facility managers, beamline scientists, and synchrotron support teams, as well as researchers aiming to optimize data management, particularly in high-throughput experimental workflows. Additionally, we present advanced tools for local installation of CCP4 Cloud, enabling institutions such as synchrotrons, research labs, and companies to meet internal data management requirements while leveraging local computational infrastructure.

[1] Krissinel, E., Lebedev, A.A., Uski, V., Ballard, C.B., Keegan, R.M., Kovalevskiy, O., Nicholls, R.A., Pannu, N.S., Skubak, P., Berrisford, J., Fando, M., Lohkamp, B., Wojdyr, M., Simpkin, A.J., Thomas, J.M.H., Oliver, C., Vonrhein, C., Chojnowski, G., Basle, A., Purkiss, A., Isupov, M.N., McNicholas, S., Lowe, E., Trivino, J., Cowtan, K., Agirre, J., Rigden, D.J., Uson, I., Lamzin, V., Tews, I., Bricogne, G., Leslie, A.G.W. & Brown, D.G. (2022) CCP4 Cloud for structure determination and project management in macromolecular crystallography. Acta Cryst. D78: 1079-1089; doi:10.1107/S2059798322007987. URL: https://cloud.ccp4.ac.uk