Oral presentation

Linking data sources with CCP4 Cloud: on the verge of a fully online setup 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 last decade, MX has seen a prominent shift toward online solutions in both experimental and computational parts of the whole structure determination process. Online approach may start with contracting a crystallisation facility, followed by a remote diffraction experiment at a synchrotron with subsequent data processing and structure solution using a suitable online software setup, and finishing with the PDB deposition. The online approach presents many benefits, including the overall efficiency, transparency, team working, independence of geographic location, data safety and long-term data maintenance. In line with current trends, CCP4 developed the CCP4 Cloud [1] as a robust solution for the complete cycle of online crystallographic computation, starting from processing diffraction images to structure refinement and preparation of the PDB deposition package.

In an ideal online setup, researchers could send their samples to one of available synchrotrons and then receive collected data in their CCP4 Cloud accounts for further processing and structure solution once their diffraction experiment is complete. While all bits of this picture are in place, the very link between data sources and CCP4 Cloud is currently missing, so that the data needs to be moved with manned procedures. Such an approach is inconvenient for moving large volumes of image data between home setups and remote CCP4 Cloud servers; this is why, up until now, image processing remains mostly a local operation in CCP4 Cloud.

We report on a considerable progress in developing a data link framework in CCP4 Cloud, capable of acquiring raw image data from various sources, including image repositories at PDBj/XRDa, SBGrid, Protein Diffraction and, potentially, any MX data source, such as a synchrotron or a diffractometer. The framework is presented to CCP4 Cloud users as a data acquisition task, and it can be also used through a Node JS API for pushing data from the source to the user area in CCP4 Cloud.

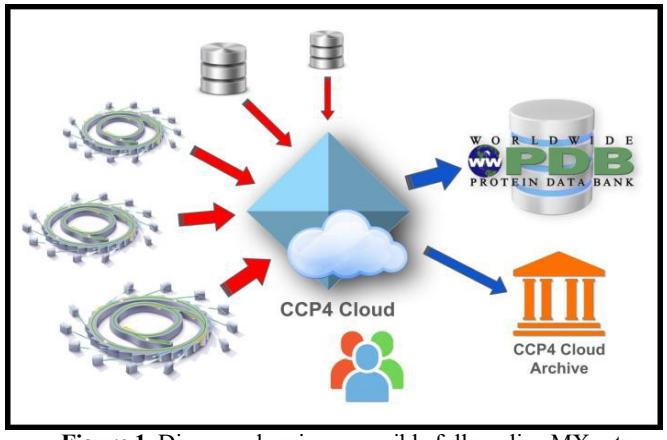

Figure 1. Diagram showing a possible fully online MX setup comprising data repositories, data producing facilities, and archives. Data links, reported in this communication, are shown with red arrows

In this presentation, we will discuss design and operational concepts of the data link framework and give further technical details. The framework is released with CCP4 Software series 9 as part of a standard CCP4 setup. The presentation will be of interest for MX facility managers, beamline scientists and synchrotron support teams, as well as for researchers wishing to optimise their data management practices, especially in large-series experimentation.

[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