How Diamond Light Source has achieved a data archive of 10 years of experiments and how we are preparing for the next 10 years

Karl Erik Levik[1], Silvia Da Graca Ramos[1], Neil Smith[1], Christopher Prosser[2], Gordon Brown[2], Andrew Richards[1], and Alun Ashton[1*]

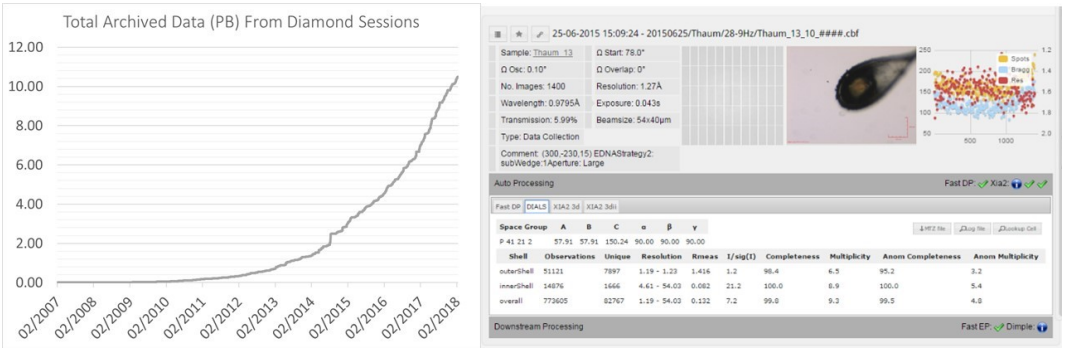
[1]*Diamond Light Source, Chilton, Didcot, Oxon, OX11 0DE, U.K.*

[2]*Scientific Computing Department, STFC Rutherford Appleton Laboratory, Harwell, Oxford, U.K.*

[*]*alun.ashton@diamond.ac.uk*

Managing the balance between the long-term benefits and requirements of capturing and cataloguing up to 25TB of experimental data a day along with meaningful information about the experiment against the immediate needs of answering immediate scientific investigation is a challenge faced by every user facility. Diamond Light Source became a user facility in 2007 and had aspirations for self-describing data files using the NeXus (http://www.nexusformat.org) or CBF format; every discipline having access to information management systems using ISPyB/SynchWeb (http://diamondlightsource.github.io/SynchWeb/); and a secure experiment data catalogue and full data preservation and provenance using ICAT (http://icatproject.org) and StorageD.

Now in our 11^{th} year of operations many of these goals have now been achieved. But with an ever- changing landscape of evolving experiments, increasing data rates and data access, improvements in beamline throughput, equipment evolution and most recently the cultural requirement for Findable, Accessible, Interoperable and Reusable (FAIR) data in every domain along with the availability of commercial cloud services the landscape is quickly evolving and new opportunities lie ahead.



Left: total archived data against time; Right: screenshot example of a rich interface to experimental data in SynchWeb.