The Great Light Sources of Europe: Studying Approaches to Synchrotron Data Management and Analysis

Lawrence Berkeley National Laboratory

The 10-day tour of Europe was not your typical itinerary — Garching, Karlsruhe, Villigen, Hamburg and Oxford. In January. But David Brown and Craig Tull of the Computational Research Division and Alex Hexemer of the Advanced Light Source weren't touring to see the sights — they more interested in seeing the lights — powerful scientific instruments known as light sources that use intense X-rays to study materials down to the macromolecular scale.

Scientific user facilities like Lawrence Berkeley National Laboratory's Advanced Light Source (ALS) are becoming increasingly powerful tools for scientific discovery with the development of higher-resolution imaging devices, which not only capture higher resolution images, but at a much faster rate. Thus, beamline scientists are faced with unprecedented amounts of data, but not always with the hardware or software needed to effectively manage, analyze and share that data.

To get a better picture of how this situation is being addressed elsewhere in the world, Brown, Tull and Hexemer visited some of Europe's leading light source, as well as other facilities. Stops included the Heinz Maier-Leibnitz Neutron Source and the Leibniz Computing Center at the Technical University of Munich in Garching; the Steinbuch Computing Center at the Karlsruhe Institute of Technology; the Swiss Light Source at the Paul Scherrer Institute in Villigen, Switzerland; the PETRA III light source at the German Electron-Synchrotron (DESY) in Hamburg; and the Diamond Light Source near Oxford, England.

"Among our goals were understanding what kind of hardware and software these facilities are using, how they manage their data and workflows and to what extent the facilities worked closely with high performance computing centers," Brown said. "We also wanted to understand the funding models used by other facilities to support their IT infrastructure with the idea of possibly leveraging the most successful models here in the U.S."

In general, Brown said, the European facilities have well-funded IT infrastructures that include computing hardware, networking and user software development and support.

"Many of our facilities here in the U.S. have some catching up to do in other to provide IT support commensurate with the emerging big data challenges," Brown said." On the other hand, in the area of new mathematics algorithms and software development, we appear to be ahead of the Europeans."

At Berkeley Lab, for instance, mathematician James Sethian leads a project called CAMERA (The Center for Applied Mathematics in Energy Research Applications) to design and apply mathematical solutions to data and imaging problems at Berkeley Lab scientific user facilities supported by the Department of Energy's Office of Basic Energy Sciences.

Tull said he found the discussions with the Europeans "extraordinarily interesting" and that the series of visits "validated many of our views, but also led me to rethink some of our views. There was a lot of commonality on the problems, and some commonality on the solutions," Tull said.

During the visits, Tull gave presentations on SPOT Suite, a Laboratory Directed Research and Development project he is leading between CRD, the ALS, NERSC, ESnet and the Materials Sciences Division. SPOT Suite is a collection of software providing a data portal, data management and processing, a database and workflow management. The system automatically transfers data from ALS beamlines to NERSC where the data is processed in real time, with the results automatically transferred back to the scientist working at the beamline.

Brown noted that while the European sites have their own computing infrastructure, they did not see the kind of working relationships with supercomputing centers that Tull is developing.

Hexemer said that many of the people they met with were also interested in HipGISAXS, the high-performance software developed by CRD's Slim Chourou, Abhinav Sarje and Sherry Li (PI), along with Hexemer. HipGISAXS is a high performance, massively parallel analysis code to support GISAXS (Grazing-Incidence Small-Angle X-ray Scattering), an experimental measurement technique characterizing materials properties at the nanoscale.

"Most of the software we heard about did not have the same capabilities as HipGISAXS," Hexemer said.

Brown said their hosts also expressed interest in better sharing of information and in holding joint workshops to understand and develop solutions to the big data challenges shared by the U.S. and European facilities.

“The Big Data challenge is just coming on for the light sources, due mainly to advanced detectors, and everyone is struggling with it,” Brown said.

About Berkeley Lab Computing Sciences

The Lawrence Berkeley National Laboratory (Berkeley Lab) Computing Sciences organization provides the computing and networking resources and expertise critical to advancing the Department of Energy's research missions: developing new energy sources, improving energy efficiency, developing new materials and increasing our understanding of ourselves, our world and our universe. ESnet [1], the Energy Sciences Network, provides the high-bandwidth, reliable connections that link scientists at 40 DOE research sites to each other and to experimental facilities and supercomputing centers around the country. The National Energy Research Scientific Computing Center (NERSC [2]) powers the discoveries of 5,500 scientists at national laboratories and universities, including those at Berkeley Lab's Computational Research [3] Division (CRD). CRD conducts research and development in mathematical modeling and simulation, algorithm design, data storage, management and analysis, computer system architecture and high-performance software implementation.

Source URL (retrieved on *11/25/2015 - 10:18pm*):
http://www.scientificcomputing.com/news/2014/02/great-light-sources-europe-studying-approaches-synchrotron-data-management-and-analysis

Links:
[1] http://es.net
[2] http://www.nersc.gov
[3] http://crd.lbl.gov