# The Xport Collaboratory for High-Brilliance X-ray Crystallography

Donald McMullen, Randall Bramley, John C. Huffman, John Bollinger, Edward Dambik, Kia Huffman, Indiana University;
Gregor von Laszewski, Argonne National Laboratory

Xport, a DoE NGI funded project at Indiana University, Argonne National Laboratory, and the Lawrence Berkeley National Laboratory targets revolutionary improvements in telepresence for major scientific instrumentation systems. Our goal is to exploit a combination of advanced networking, middleware services, and remote instrumentation technologies to achieve interactive "better-than-being-there" capabilities for remote experiment planning, instrument operation, data acquisition, reduction and analysis. These capabilities are being deployed and evaluated at several X-ray crystallography facilities including the Advanced Light Source and the Indiana University Molecular Structure Center.

The primary aim is to develop a set of components for high brilliance X-ray crystallography that will simplify the design and execution of macromolecular crystal structure determinations at two DOE facilities, the Advanced Light Source and the Advanced Photon Source, and one University facility, the Molecular Structure Center at Indiana University.

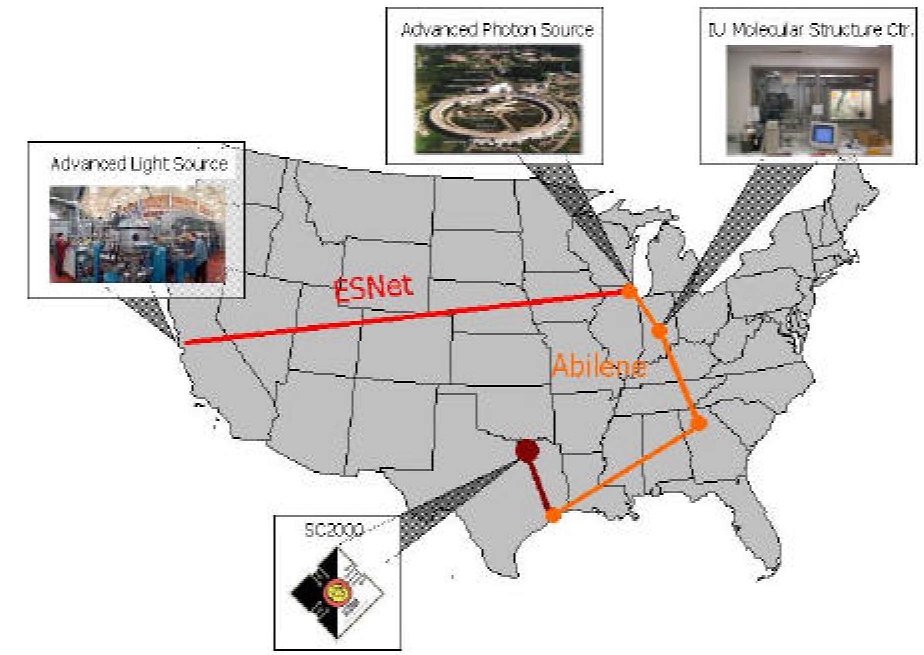
Building on work currently underway in the Globus group and the DoE 2000 Common Component Architecture (CCA) Forum, the Xport project addresses several issues related to NGI network-based instrumentation including high speed data collection, reduction, storage and visualization, and real-time instrument control for the acquisition of macromolecular X-ray crystallographic data from high brilliance X-ray sources.

***Goals for the Xport project are:***

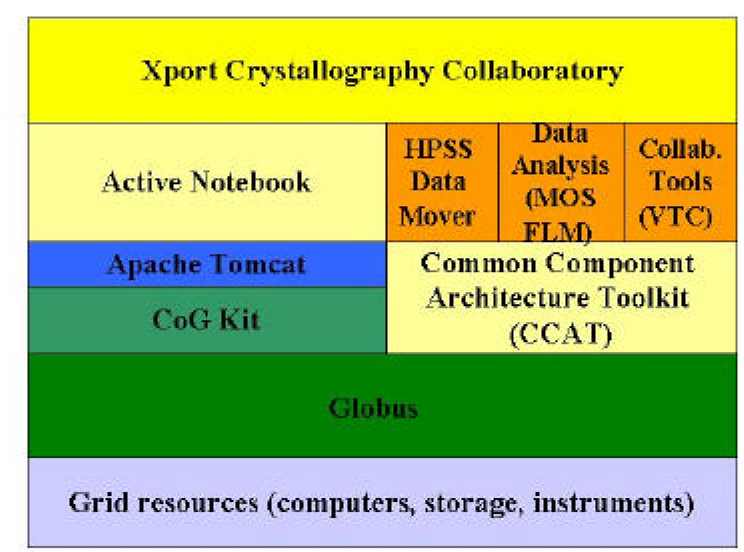
- establishment of an NGI (primarily ESNet and Abilene) hosted testbed for macromolecular crystallography;
- development of a component-based application environment for instrument control and experimental data collection reduction and storage system based on the Common Component Architecture Forum's specifications and deliverables;
- a high-speed data caching and transfer middleware to support direct connection of the ALS and APS instrumentation to grid mass storage and computing systems;
- distributed real-time data reduction components which can be run during data collection to determine quality of the data being collected and to provide feedback for instrument tuning;
- extensions to the current Component Architecture Toolkit (CCAT) to enable quality of service (QoS) and real-time network performance analysis for a production application, initially through the GARA API on the Emerge QoS testbed;
- extensions to the CCAT to allow interactive performance monitoring and analysis of computation and communication in individual components;

***Testbed and data transport requirements***

The APS, the Molecular Structure Center at Indiana University, and the ALS were selected as the initial testbed sites. Data transport requirements for current generation CCD detectors (9-32MB/image every 2 sec.) were estimated at 128Mbps sustained for ~12 h. per experiment. In addition to CCD image transport we expect each experiment to require two or more channels of streaming video and audio at 384Kbps each by which researchers at remote sites can monitor the progress of the experiment and consult with local expertise at the beamline.



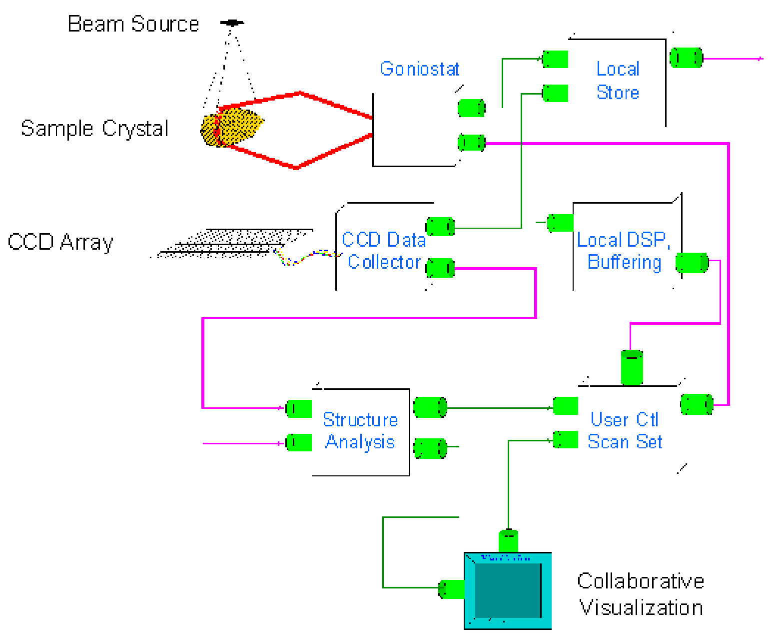

**SC2000 Xport collaboratory testbed**




**Xport software architecture overview**

### *Xport Software architecture*

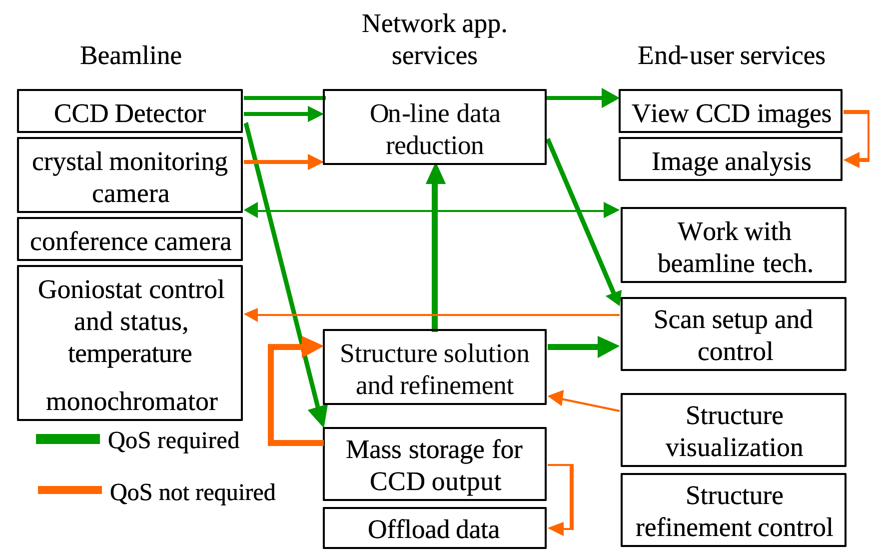
The overall software architecture for Xport is shown in the following diagram:




Each block in the diagram represents a CCAT component written in C++ or Java. CCD frame data flows from the detector to a cache at the beamline, then to a network archive (currently an HPSS server at IU or ANL) where (possibly asynchronous) data reduction and visualization applications that may be distributed across a computational grid can retrieve individual frames and intermediate data sets. Each component is a CCAT instance running on a machine appropriate for its function.

Since CCAT uses Globus extensively components running on NT systems represent a special challenge. Currently these components include data acquisition, local cache management and instrument control for Bruker goniostats and CCD detectors. The intermediate solution for using CCAT on NT machines is to establish a Globus-based proxy that instantiates NT component via ssh (secure shell). An alternative approach being explored is a more general mechanism for instantiating CCAT components through servlets without using the Globus resource manager. These changes are tracking developments within the CCA Forum, and leading the Forum in some areas.

In particular, we are experimenting with using the shared object access protocol (SOAP) as a run-time system to replace the relatively heavy-weight Globus mechanisms for services such as instantiating components, connecting them, sending control signals among them, and more generally for data transfers which do not require high-speed, broad bandwidth communications.




### *Network Quality of Service*

A key element of the architecture for this project is the use of a "network storage" paradigm to eliminate file system dependencies in data reduction and visualization component codes. To prevent local cache overflow and stalling of the experiment some level of QoS is needed in the link between instrument and network storage. At present we are characterizing the GARA bandwidth reservation API from Argonne using the Emerge testbed.

QoS requirements between functional blocks is indicated in the diagram above. Paths that need some level of QoS (bandwidth guarantees primarily) are indicated in green. Components for data transfer to an HPSS server and for video are being modified to support bandwidth reservation.

As an adjunct to this project a Cisco 7500 series router has been placed in the Abilene network operations center to extend Emerge from ESNet through Abilene to Indiana University.

### *Data acquisition components*

Data acquisition consists of two phases: instrument setup and moving CCD images to the storage server as they are generated. We do not replace existing code for acquisition directly from the CCD, rather the data acquisition components will monitor a directory into which existing code for reading the detector places images. This approach allows Xport components to be added to an existing experiment incrementally and without disrupting current data acquisition practices. At the present setup and acquisition components are being developed for Bruker goniostat/detector systems and these are expected to port easily to other vendors' detectors.

### *Data analysis and management components*

Macromolecular biologists uses several packages for analyzing diffraction patterns. Our approach to the "community code" problem is to provide transport and storage formats for CCD images and intermediate reflection data that can be easily converted to an appropriate format for a given analysis suite. We are demonstrating a CCAT component wrapper for the MOSFLM analysis and display code and expect to provide similar component versions of both the CCP4 and D*TREK crystallography analysis packages. A component for visualizing the "reciprocal lattice" (a 3D representation of the X-ray diffraction pattern) is also being developed for a range of display devices from desktop workstations to immersive CAVE™ systems.

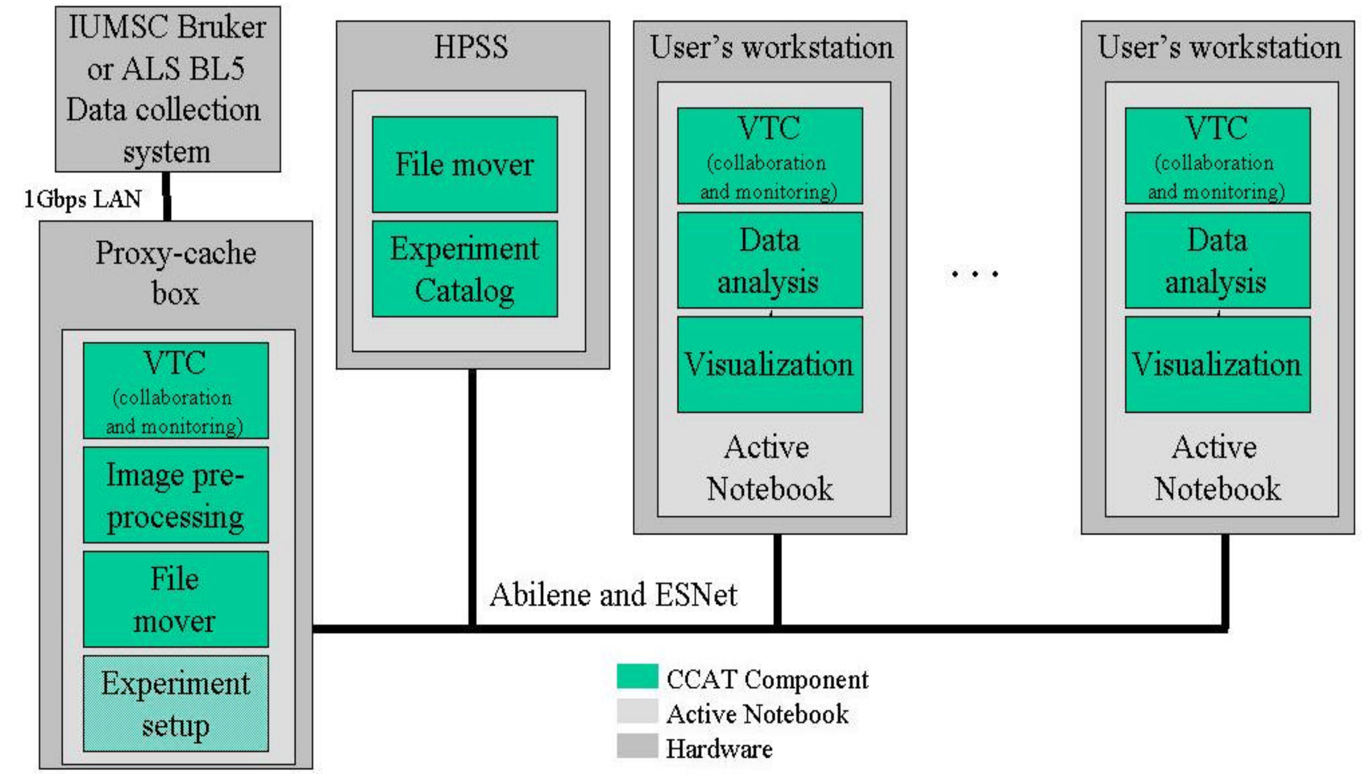
*Current Capabilities*

The diagram to the right illustrates the current capabilities and operating modes of the Xport software. A remote crystallographic experiment is conducted as follows:

- Crystal samples are sent to the beamline or MSC facility.
- At the beginning of a run a researcher remote to the X-ray source initiates a video teleconferencing session with beamline staff.
- Parameters for crystal movement and beam characteristics are determined by the remote researcher and the experiment is initiated either remotely or locally.
- The remote researcher determines where the data will be stored during and after the run and what image processing will be done at what stages during the experiment. Typically a few images are collected at different angles to determine the quality of the crystal.
- VTC software is used to monitor the crystal mounting process and the alignment of the crystal in the X-ray beam.
- While data are being collected cache software on the beamline proxy computer moves CCD images to a mass store device. Collaborators at other sites can monitor the progress of the experiment and can set up independent data reduction and visualization pipelines for data which are available at the moment.
- All software configuration and data management are provided by the CCAT Active Notebook which provides a framework for distributed software execution and a record of experimental parameters, data reduction software and parameters, and location of the data.

**Data collection system at the ALS Beamline 5. Goniostat and beamline optics are in the hutch to the right of the chair.**

**IUMSC Bruker Goniostat and 2Kx2K CCD detector**




**Xport CCAT component layout on the grid**