SCIENCE STUDIO: A PROJECT STATUS UPDATE*

D. Liu, E. D. Matias#, D. Maxwell, D. Medrano, Canadian Light Source, University of Saskatchewan; Saskatoon, SK, Canada
M. Bauer, M. Fuller, N. S. McIntyre, University of Western Ontario, London, ON, Canada
Y. Yan, Concordia University, Montréal, QC, Canada
C. H. Armstrong, J. Haley, IBM Canada, Markham, ON, Canada

Abstract

The Canadian Light Source (CLS), the University of Western Ontario (UWO), IBM and Concordia University are in the process of building an integrated experiment management system. This system utilizes web-browsers as a thin client that can be connected to servers at the CLS over conventional Ethernet or User Configurable Light Paths. The system is based on a Service Oriented Architecture and provides access control, data acquisition, data storage and data visualization. More recently, work on the system has included implementation of user office functionality, integrated control of an EPICS based beamline and analysis codes.

INTRODUCTION

This paper presents a short overview of ScienceStudio as well as providing an outline of recent development in the ScienceStudio project.

Recent developments centre in the following areas:

- Enhancements to the remote access and online visualization capability.
- Development of a user training and user server module.
- Integration with Active Directory for user authentication.
- Development of complementary data analysis capabilities using Cell based grid computing.
- Capability to work with CDF as a canonical format for multi-variable data sets.

ARCHITECTURE

The design uses a service oriented architecture. Figure 1 presents a conceptual overview.

The architecture consists of the following parts:

- User Interface (UI) - Users interact with the system through a web browser that acts as the user agent to the UI services component. The browser communicates with the UI services using AJAX.
- The UI services manage the user's interaction with the system. The web browser clients interact with

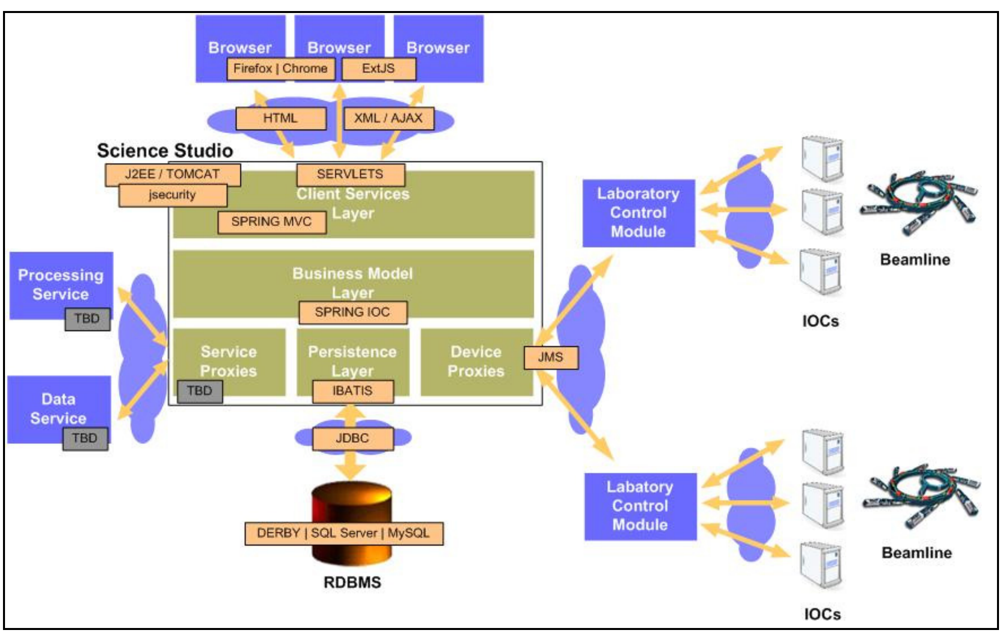

Figure 1: ScienceStudio system architecture, showing the relationship between the ScienceStudio portal, related relational databases, EPICS based control systems and other services.

*This work was funded by CANARIE as part of the Network Enabled Platforms Program, Project NEP-01.
#elder.matias@lightsource.ca

the UI services to obtain content, e.g. web pages and XML, and for servicing events from the user.

- The Application Service provides the mechanism to create, retrieve and modify metadata relating to experiments such as projects, sessions schedules, samples, etc. The metadata assists in organization and automation, improving the experiment process.
- The database holds all experiment metadata and some data. All metadata and data for a given project or experimental session is provided to users in the form of a "zip archive" available directly from the User Interface. With metadata in XML, and experiment data in Common Data Format (CDF) [5] (and/or the original detector format).
- The underlying beamline control system is implemented using EPICS based Input/Output Controllers (IOCs) and with remote access capability [1]. A beamline abstraction layer has been developed to decouple ScienceStudio and the beamline controls for maintenance and security purposes [2].
- Scanning is performed using the standard CLS scanning library commonly used on the beamlines.

USER OFFICE

Training Module

Developed by Concordia and CLS the User Training Module provides the capability to track the status of users, notify users when their training is about to expire, and grade training tests. A separate web interface is also provided to permit the upload of new training material and configure tests. The system is fully integrated into active directory.

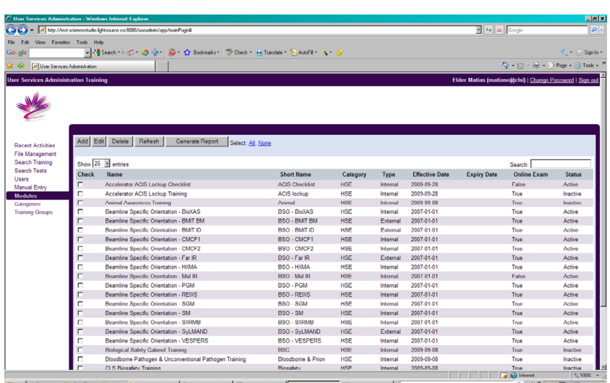

Figure 2: Sample screen for training module

Survey Module

The survey module has been developed by Concordia and is currently being integrated into ScienceStudio by CLS. This module provides an online method for users to complete a survey after performing an experiment providing feedback to CLS on the quality of service they receive.

Scheduling Module

The scheduling module is currently being developed by Concordia University. This module will provide the ability to automate the scheduling of user time using a heuristic algorithm [3].

Proposal Submission Module

This module is currently being developed by CLS and will provide the ability to manage user proposals.

IMPLEMENTATION ON VESPERS

The VEry Sensitive Elemental and Structural Probe Employing Radiation from a Synchrotron (VESPERS) beamline is an X-ray microscope beamline at the CLS. The VESPERS beamline provides X-ray Fluorescence (XRF), Laue X-ray Diffraction (XRD), XANES and XAFS information from the same microscopic spot [4].

The ScienceStudio software (Figure 3) allows remote collection and saving of individual XRF spectra from VESPERS, and has an integrated web-based interface permitting spectra to be collected and rapidly downloaded. The concept of a tree-based browser (shown on the left) has been introduced to permit users to effectively navigate between data from different experiments.

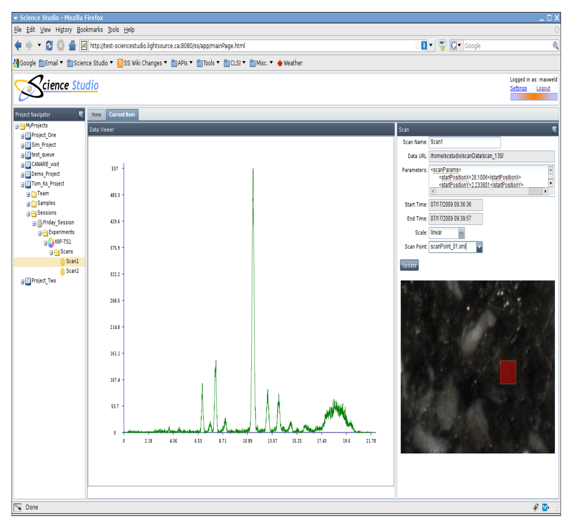

Figure 3: First outside user spectra taken using Science Studio.

REMOTE ACCESS

Within ScienceStudio there have been many improvements in the areas of experiment control and data acquisition. Two dimensional area scans are supported using the standard CLS data acquisition system. This reduces the complexity of VESPERS software systems by using the same scan configuration as the existing VESPERS beamline data acquisition system. A 2D scan region is defined by the step size, and the coordinates of start and end points. These coordinates can be either entered manually in web forms, or interactively selected from a live microscope image of the sample on an AJAX page. The sample image, along with the scan region, is saved during data acquisition and provides an important visual reference for the user when they review their data. Streaming videos from several cameras have been integrated into the web interface to provide the user with views of the experimental station from various aspects.

An important part of the remote access system is to regulate and control access to the VESPERS beamline. An administrative control screen has been added to the web interface. This allows authorized personnel to control access by starting and stopping beamline sessions. Each session has an associated team of users, and each user has a role of either Experimenter or Observer. These users are only allowed to join a session once that session has been started. When a user has joined a beamline session, they are allowed to take control of the beamline only if they have the Experimenter role. Only one user is allowed to have control of the beamline at any time.

The next significant function to be added to the remote access system is XRD. This is currently under testing at the CLS and will provide the user with an interface to configure the CCD detector and to acquire XRD images.

DATA ANALYSIS

An important goal of ScienceStudio is the tight integration of data analysis both on conventional and grid computing resources.

The experimental results once stored in ScienceStudio can be readily downloaded and analyzed further using the 'Peakaboo' software that leads the user through the analysis of XRF spectra and the creation of elemental distribution 'maps'. Figure 5 and 6 illustrate the data analysis results using the Peakaboo software. The University of Western Ontario developed Peakaboo as part of the ScienceStudio project. Provision is provided in the web portal to easily transfer data from the current and past experiments into Peakaboo.

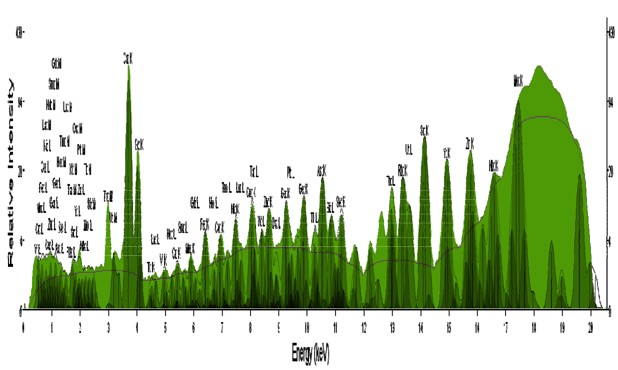

Figure 4: Analysis of first spectra data set using Picaboo.

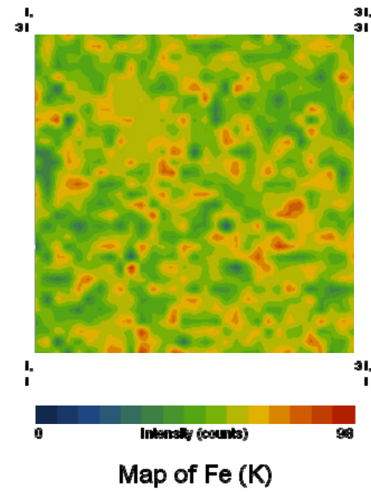

Figure 5: After collection of 1600 individual spectra from a rastered area of 100x100 microns, the entire data set is downloaded from the CLS to SharcNet at UWO in 45 seconds. The Peakaboo software is used to create maps for each of 35 elements (Fe is shown).

Work is ongoing in the development of a strain analysis code using a Cell Processor based cluster. Our ultimate goal would be to integrate this analysis code with ScienceStudio using stream computing.

DATA FILE FORMAT

To facilitate the integration of data analysis code it was important to select a common data file format. The Common Data Format (CDF) [5] was originally developed by NASA to provide a common format for scientific data analysis. The CLS selected CDF as the canonical format for experimental data within ScienceStudio. Experiment data from EPICS is parsed and translated into CDF. The data files can then be extracted from ScienceStudio in CDF and imported into other data analysis packages such as Peakaboo.

REFERENCES

[1] E. Matias, D. Medrano, D. Chabot, D. Maxwell, N. S. McIntyre, M. Fuller, C. Armstrong, J. Haley, "Remote Beamline Access at the CLS," SRN Vol. 21, Nov. 5. pp. 30-33.

[2] D. Chabot, "Beamline Control Module," EPICS Collaboration Meeting. Padova Italy. Oct. 2008.

[3] Z. Aahid, Z. Wang, C. Wang, D. Ni, Y. Xu, Y. Yan, "An Integrated Programming Model and Heuristic Algorithm for Automated Scheduling in Synchrotron Facilities," SMC'09, San Antonio, TX. Oct. 2009.

[4] R. Feng, A. Gerson, G. E. Ice, R. Reininger, B. Yates, and N. S. McIntyre, "VESPERS-A Beamline for Combined XRF and XRD Measurements," AIP Conference Proceedings, 879, pp. 872-874 (2007).

[5] NASA, "CDF User's Guide Version 3.3," Space Physics Data Facility, NASA/Goddard Space Flight Center, Greenbelt, Maryland. 2008.