

# Common Instrument Middleware Architecture: Extensions for the Australian e-Research Environment

Ian M. Atkinson,<sup>\*a</sup> Douglas du Boulay,<sup>b</sup> Clinton Chee,<sup>b</sup> Kenneth Chiu,<sup>c</sup> Kia L. Huffman,<sup>d</sup>  
Donald F. McMullen,<sup>d</sup> Romain Quilici,<sup>b</sup> **Peter Turner**,<sup>\*b</sup> and Mathew Wyatt<sup>a</sup>

<sup>a</sup>School of Information Technology, James Cook University, Townsville, Qld, Australia.

<sup>b</sup>School of Chemistry, The University of Sydney, Sydney, NSW, Australia.

<sup>c</sup>Computer Science, State University of New York (SUNY) at Binghamton, NY, USA.

<sup>d</sup>Pervasive Technologies Lab, Indiana University, Bloomington, IN, USA.

## Abstract

The Common Instrument Middleware Architecture (CIMA) model is being adapted and extended for a remote instrument access system being developed as an Australian eScience project. Enhancements include the use of SRB federated Grid storage infrastructure, and the introduction of the Kepler workflow system to automate data management and provide facile extraction and generation of instrument and experimental metadata. The SRB infrastructure will in effect underpin a Grid enabled network of X-ray diffraction instruments. The CIMA model is being further extended to include instrument control, and is being embedded as a component in a feature rich GridSphere portal based system for remote access.

## 1. Introduction

### 1.1 The Australian Context

Australia is a large island continent with most of its population primarily residing in a relatively small number of coastal cities. Given the size of the country, it is easy to appreciate that the provision of remote access services to resources, such as scientific instruments and their data, offers significant efficiency and cost benefits. Further gains may be provided by harnessing distributed resource technologies such as grid storage and compute resources, work-flow tools and web services. In order to explore this potential, we're developing a pilot network of X-ray diffraction instruments equipped with Grid enabled services for remote access.

The development network encompasses instruments at James Cook University, Monash University, the University of Queensland and the University of Sydney (Fig. 1). That is, geographically the sites span the east coast of Australia across a distance of over 3000 km.

The selection of X-ray diffraction instruments for a remote access development programme, emerges naturally from the observation that crystallography offers well defined work-flows and data structures, and

utilises relatively common (if not standardised) instrument types. X-ray diffraction instruments are sufficiently expensive items that many institutions are not in a position to own and operate such equipment.

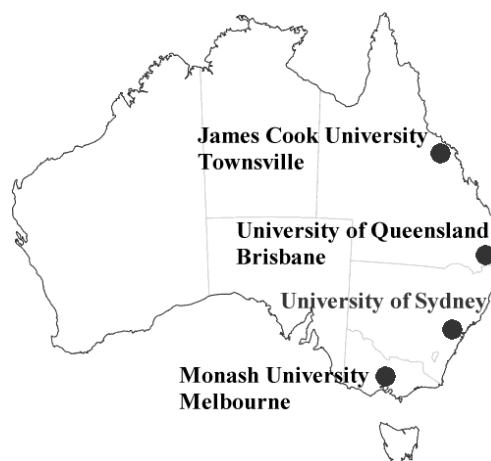


Figure 1. Remote Access Network Sites

### 1.2 Technology

The Common Instrument Middleware Architecture (CIMA)<sup>1</sup> project is developing a web services based API that embodies a generic or abstract representation of an instrument in terms of its type, identity, data and metadata output streams. The CIMA model provides a

middleware layer between the instrument and the network within which to standardize the representation of the instrument, to mediate access by downstream components, and to host extensions to the instrument's functionality. This middleware layer need not be co-located with the instrument, and could be elsewhere on the network. The CIMA model is intended to promote re-usability and hence reduce coding effort in changing instrument environments or eco-systems.

## 2. Adaptations and Extensions to CIMA

In close collaboration with the CIMA project team, we're adapting and extending the capabilities of CIMA in developing a rich GridSphere portal environment for remote instrument and data access.

### 2.1 Workflow and Data Management

Several significant changes to the CIMA model have been made with a view to increased flexibility and capability. The changes are depicted in Fig. 2 and are summarised as follows:

- NFS and MySQL data manager replaced with an SRB<sup>2</sup> and MCAT backend
- Use of Personal Grid Library (PGL)<sup>4</sup> for user friendly SRB data manipulation
- Moved from C++ and Perl to Kepler<sup>3</sup> to facilitate the handling of data management
- Ability to customise data storage via a visual workflow system
- Ability to create metadata schema definitions for experimental metadata
- Highly extensible, storage repository not restricted to SRB

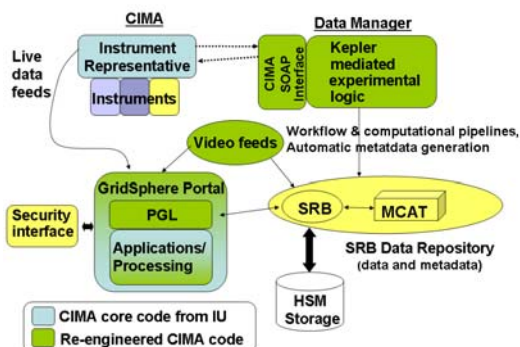


Figure 2. Extensions to the CIMA System

A significant goal of the project is to realise the possibility of instruments serving as first class members of the Grid. As Fig.3 suggests, Storage

Resource Broker technology is being utilised to link the participating instrument sites.

At present the Data Manager for all sites is located at JCU with a production scale SRB storage facility. It is intended that each site will run its own Data Manager and SRB storage facility, which is to be federated across the sites into a single shared data space. The security and access rights issues associated with this store are presently being considered.

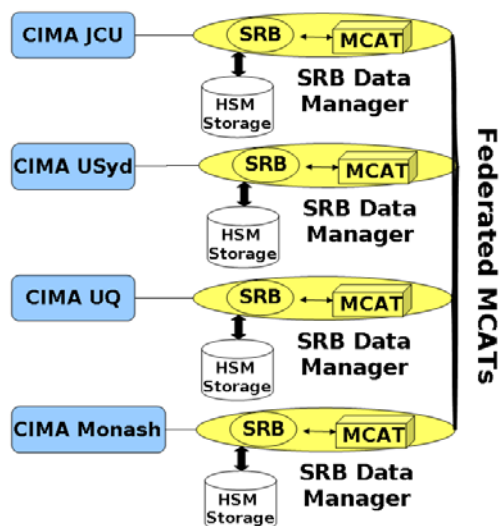


Figure 3. SRB Based Network Storage

The Personal Grid Library has been developed at James Cook University to provide easy web browsing and access to standalone or federated SRB instances. The PGL also supports annotation and the convenient display of object metadata. Annotation or metadata based searches can be undertaken, and stored images may be viewed.

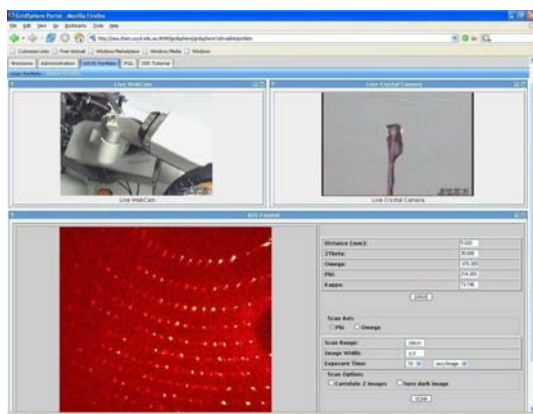
As can be seen in Figure 2, Kepler plays a key mediating role in the work flow and data management. For use cases outside of X-Ray crystallography, the data management requirements will most likely be very different. Data management customisation for new applications would normally impose a heavy coding over-head. However, by exploiting Kepler workflows the development effort can be dramatically reduced. Using Kepler a customised data manager workflow can be configured in days, rather than weeks or months. Workflows in Kepler can be exported into XML, which can then be deployed to other instances of Kepler at different sites.

### 2.2 Remote Instrument Control

The remote desktop approach to remote instrument access, such as typified by the use of

VNC<sup>5</sup> (and its many variants), CITRIX<sup>6</sup>, Tarantella<sup>7</sup> and NX<sup>8</sup>, has the significant advantages of ease of set-up and familiarity. While convenient, these approaches are not ideal and can afford remote instrument users with excessive control over expensive and potentially dangerous instruments, and have a relatively poor security model. A significant disincentive in building custom-built remote access systems, is that there is a high coding overhead that may reproduce functionality already provided by an instrument manufacturer. A significant advantage of the custom built interface approach however, is that the actions of the remote instrument user can be tightly controlled, while at the same time services outside the desktop environment can be provided to offer a richer operating environment.

Accordingly a Gridsphere portal is being built that includes a portlet providing a purpose built interface to a diffractometer instrument that will tightly define the actions of a remote user, and so minimise accidental damage and injury. The portlet allows the user to enter data collection specifications and allows the execution of those instructions. Live views of the instrument and crystal sample are provided, together with the most recent diffraction image (see screen shot in Fig. 4.)



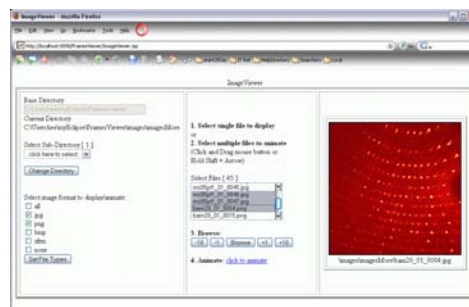
**Figure 4.** Instrument Control Portlet

A second portlet pane provides instrument status information, such as generator status, detector temperature and laboratory temperature.

A further portlet allows the user to browse the images collected to date (See Fig. 5), and includes a high speed ‘flick through’ capability. The browsers capability is to be further extended to include annotation and analysis.

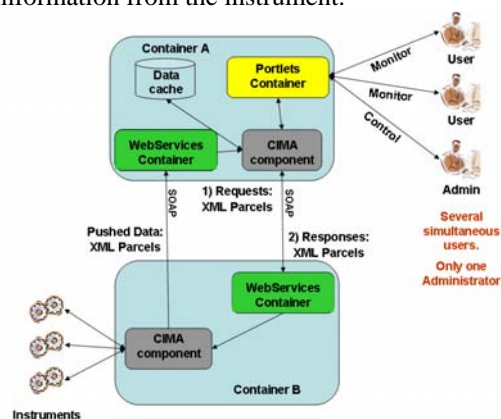
Thus far CIMA has been developed solely for instrument and sensor monitoring. We are now extending the role of CIMA to include

support for instrument control, and this involves including CIMA services as components in a modular portal architecture (see Fig. 6).



**Figure 5.** Portlet Diffraction Image Browser

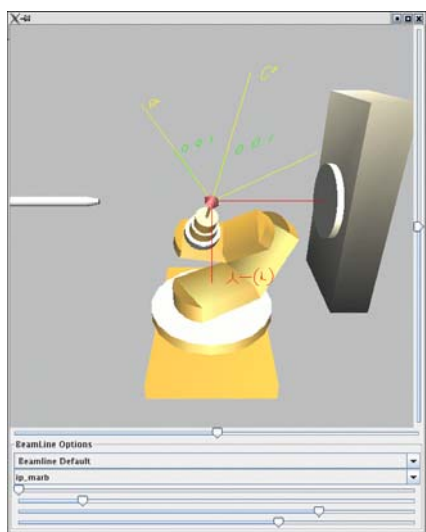
A modular SOA model using Web Services has been adopted to provide maximum flexibility, with Web Services providing language and location independence. Container communication is based on SOAP messages containing XML parcels. A data cache minimises the need to seek (get) updated status information from the instrument.



**Figure 6.** Instrument Control Architecture

The remote steering portlet is also being enhanced with a virtual depiction of the instrument, allowing a data collection simulation and assessment before execution. The instrument simulation utilises and extends the DS diffractometry simulation program.<sup>9</sup> The DS software has been ported from HP based C and OpenGL to Linux, Java, JOGL and OpenGL for this work, and can now be incorporated into the portlet system. Currently we’re adapting the OpenGL graphics to X3D<sup>10</sup> for further flexibility and functionality. The DS implementation can reflect the current state of the diffractometer, and can simulate the axis and detector movements of a full data collection. Further work is being undertaken to more precisely represent the instrument being simulated. The simulator has training benefits

and can help minimise the risk of damage to the instrument. The simulator also provides a back-up for a web camera failure and reduces the impact of the dark lab problem.



**Figure 7.** DS<sup>9</sup> Based Instrument Simulation

A significant challenge remaining is the incorporation of a single sign on security system, and currently we're exploring the use of Shibboleth.<sup>11</sup>

## References

- (1) (a) Huffman, K.L., Gupta, N., Devadithya, T., Ma, Y., Chiu, K., Huffman, J.C., Bramley, R., McMullen, D.F. Instrument Monitoring, Data Sharing and Archiving Using Common Instrument Middleware Architecture (CIMA): A new approach to modernize and standardize the data collection procedures in an X-ray Crystallography Laboratory. *Journal of Chemical Information and Modeling*, accepted February 27 2006. (b) CIMA: <http://www.instrumentmiddleware.org>
- (2) Storage Resource Broker: <http://www.sdsc.edu/srb>
- (3) Kepler workflow project: <http://www.kepler-project.org>
- (4) Portable Grid Library: <http://plone.jcu.edu.au/hpc/hpc-software/personal-grid-library/releases/0.3>
- (5) Virtual Network Computing; <http://www.realvnc.com/> (also TightVNC, RealVNC, UltraVNC, and TridiaVNC)
- (6) CITRIX: <http://www.citrix.com>
- (7) Tarantella; now renamed Sun Secure Global Desktop: <http://www.sun.com/software/products/sgd>
- (8) NoMachine (NX);

- <http://www.nomachine.com/>
- (9) *J. Appl. Cryst.* (1995). 28, 225-227. DS - a 3D graphics simulation program for single-crystal diffractometer. Zheng, M. Yao and I. Tanaka.
  - (10) X3D: [www.web3d.org](http://www.web3d.org)
  - (11) SHIBBOLETH: [shibboleth.internet2.org](http://shibboleth.internet2.org)

**Acknowledgements:** The support of the Australian Department of Education, Science and Training (DEST), for the DART project ([dart.edu.au](http://dart.edu.au)), the Australian Research Council (ARC) eResearch Program, the ARC Molecular and Materials Structure Network ([mmsn.net.au](http://mmsn.net.au)) and GrangeNet is gratefully acknowledged. Dr Min Yao is thanked for generously providing the DS source code, and the assistance of Dr Darren Spruce is also gratefully acknowledged.