

## CMCS Highlight: The CMCS Chemical Informatics Portal

<http://cmcs.ca.sandia.gov/>

January, 2003

### *Chemical Informatics Portal Demonstrated*

At the SC2002 Conference, the CMCS team gave the first public demonstrations of the Multi-scale Chemical Sciences portal and infrastructure. The portal serves as the web interface for the adaptable informatics infrastructure being developed by the CMCS team and piloted within the chemical science community. The data infrastructure takes advantage of a variety of standards and open-source information technologies to provide an unprecedented ability to share data, data pedigree, and project information within groups and across communities. The portal, which can easily be enhanced and customized through the inclusion of new 'portlets', includes real-time collaboration capabilities, search and notification tools, and a pedigree browser. To support the chemistry community, the CMCS team has integrated a variety of powerful chemistry applications, data viewers, and data translators. .

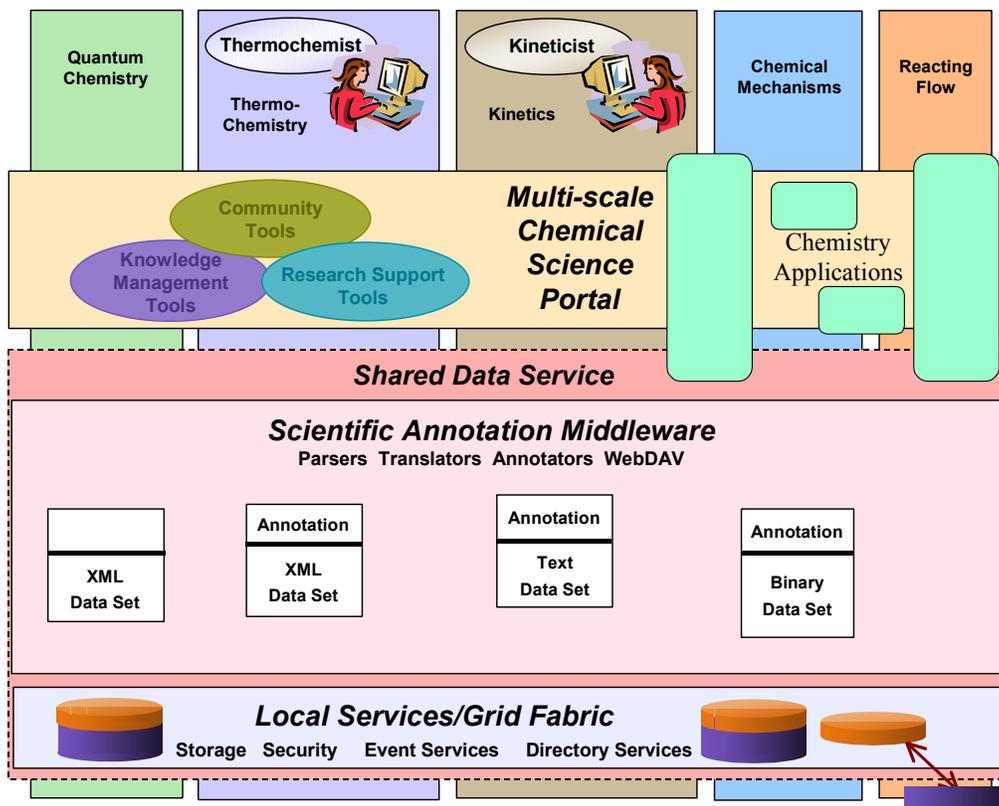
The demonstrations at SC2002 followed the work of a hypothetical kineticist who uses the portal to discover a recent electronic structure calculation for a key species in a combustion model he is in the process of developing. The kineticist is able re-optimize the overall thermochemistry used in his model using a portlet to access a remote Active Thermochemical Tables (ATcT) web service. The portlet and infrastructure automate the production of several derived and translated data sets that are required to analyze the new input and to format it for use in subsequent applications. The kineticist's collaborators stay informed via access to the kineticist's notes in an electronic notebook and the automatically generated pedigree information on his new results. Compared with the current process, the kineticist spends less time in discovering data, translating and transforming it, gathering applications, and documenting his work.

### *CMCS Infrastructure Overview*

The diagram in Figure 1 provides an overview of the intended CMCS pilot community and the high-level CMCS architecture. The diagram shows how a CMCS user interacts primarily with the top layer, the CMCS portal and chemistry applications. The applications can appear within the portal or provide their own user interfaces and interact directly with the underlying metadata/data and other CMCS services. The portal provides an array of functionality to support group and community processes, with an emphasis on simplifying the discovery and use of data. The shared data service, shown as the second layer, provides configurable capabilities for automating the generation of metadata, translating data between standard formats, and federating multiple data stores. At the lowest layer, the portal can take advantage of existing distributed services for security, event management, and data storage.

### *Technical Specifications*

The CMCS portal and infrastructure have been developed using Java and web standards. The portal itself is built upon an extension of the open-source Jetspeed portal environment developed by the CHEF (CompreHensive collaborativE Framework) project at the University of Michigan. Jetspeed is in turn based on the Java Servlet standard and



**Figure 1. A diagram representing the major conceptual elements of the CMCS Informatics Infrastructure.**

a variety of other open-source tools including Tomcat, Turbine, and Velocity. CHEF extends the basic portal capabilities of Jetspeed to include a variety of synchronous and asynchronous collaboration tools based on standard collaborative service programming interfaces.

A primary feature of the CMCS portal is the CMCS Explorer portlet. CMCS Explorer was developed to provide a rich interface to the CMCS data/metadata service that provides hierarchical (file-system-like) and pedigree-based browsing, data upload and download, metadata-based searching, data viewing, and access to dynamically generated translations of data. Another capability available within CMCS Explorer is the ability to register interest in specific types of data in order to receive email notification when such data is uploaded to CMCS.

The underlying data/metadata service uses the standard webDAV protocol and is based on software developed within the Scientific Annotation Middleware (SAM) project. SAM also produces the Java Messaging Service (JMS) events used by the CMCS Notification Email Daemon (NED) to implement the notification service. At present, the open-source MySQL database is used as the low-level data and metadata store.

### *CMCS Pilot Use*

The CMCS team is currently working with potential pilot groups to define the general and chemistry specific functionality required to support them effectively. Towards this end, several quantum chemistry, thermochemistry, and kinetics data stores are being integrated into the CMCS data service. Applications such as the Extensible Computational Chemistry Environment (Ecce) and ATcT have been extended to use the CMCS infrastructure to store data and to generate CMCS pedigree information. Third-party programming interfaces and user tutorials are also being developed. These activities are expected to culminate in a Version 1 CMCS portal that will be in active pilot use beginning this spring.

#### **Point of contact for this highlight:**

Jim Myers

Pacific Northwest National Laboratory

[Jim.myers@pnl.gov](mailto:Jim.myers@pnl.gov)

610-355-0994