

Collaboratory for Multi-Scale Chemical Science

Status as of April 2003 / Quarterly Report for Q2 of FY 2003

Project Staff

Larry Rahn-SNL*, Christine Yang, Carmen Pancerella, David Leahy, Michael Lee, Renata McCoy, Theresa Windus-PNL*, James D. Myers, Karen Schuchardt, Brett Didier, Eric Stephan, Carina Lansing, Al Wagner-ANL*, Branko Ruscic, Michael Minkoff, Sandra Bittner, Gregor von Laszewski, Reinhardt Pinzon, Sandeep Nijsure, Kaizar Amin, Baoshan Wang, William Pitz-LLNL*, David R. Montoya-LANL*, Lili Xu, Yen-Ling Ho, Thomas C. Allison-NIST*, William H. Green, Jr.-MIT*, Michael Frenklach-UCB*
* denotes Institutional Point of Contact

Summary

During this performance period the pilot Collaboratory for Multi-Scale Chemical Sciences (CMCS) has focused its efforts towards achievement of an improved version of the portal and infrastructure software suitable for use by early adopters. A new science use-case emphasizing multi-scale chemical science was developed to exercise the improved portal and infrastructure. This was demonstrated to the CMCS project Advisory Board on April 15. The resulting discussion helped set project priorities, prepare for our Peer Review, and also resulted in a new collaboration proposed by Wing Tsang (NIST) to be pursued in a workshop this coming September.

Other progress included continuing activities to prepare for support of the PrIME collaboration, initial discussions on an “open data policy” with appropriate DOE and Lab officials, presentation of a CMCS poster at the DOE SciDAC PI meeting in March, and discussions with Reaction Design on collaborations *via* data for Chemkin applications.

Progress

This document summarizes the work done over the second quarter of FY03 by the CMCS project. First, a summary listing of project activities is presented, then more a detailed discussion of accomplishments is presented, highlighting an improved pedigree user interface.

Summary of CMCS Project Activities – January 2003 through April 2003 (note: detailed development activities listed separately below)

- CMCS project team meeting (1/16/2003)
- Dave, Larry, Christine meet with Reaction Design team: Ellen Meeks, Graham Westmacott, & Chen-Pang Chou (2/4/2003)
- CMCS project team meeting (2/20/2003)
- Jim and Larry meet with Mary Anne to discuss FY03 funding shift and future funding of CMCS related activities (2/21/2003)
- Dave Leahy and Larry Rahn present CMCS poster at SciDAC PI meeting in Napa, CA (3/10/2003)
- CMCS project team meeting (3/20/2003)

- VNC testing with CMCS Advisory Board Members (4/7/2003)
- CMCS Advisory Board Meeting (4/15/2003)
- CMCS project team meeting (4/17/2003)
- Phone conference on establishment of an DOE Open Data Policy with Paul Gottlieb, DOE HQ, Craig Smith, SNL, Steve May, PNNL, Dickson Kehl, DOE AL, Gary Drew, DOE OAK and Larry Adcock DOE NA116 (not present, Arlene DeBlanc, DOE Science). Representing CMCS were Jim Myers, David Leahy, Larry Rahn.

Project Management, Structure and Planning

The CMCS leadership along with working group leads developed a focused set of high-level tasks for this period. These priorities are listed below:

1. Develop, implement tasking for V1 production infrastructure release
2. Implement production server infrastructure
3. Re-plan & prototype next phase of infrastructure and applications
4. Outreach to PrIme chemistry collaboration & others, as alpha users
5. Public Data: Thermochemistry, NIST Data, PrIme Data, CDAT, ...
6. Prepare for Peer Review
 - a. Beta infrastructure version & Multi-scale use-case demo
 - b. Presentations, draft publications, met with Advisory Board

The resource distribution issues raised last quarter were discussed and resolved with the sponsor. These resulted from our decision to build rather than buy portal software.

The CMCS Portal and Infrastructure

Significant advances were made this period in the development of an improved and more robust portal and infrastructure. Significant enhancements have been made to the portal interface, notification capabilities, metadata searching, user and group management, and we integrated SAM's Electronic Laboratory Notebook into the portal. The requirements were derived from testing the infrastructure prototype developed for SC02 in November, 2002. This infrastructure was developed using 7 use-cases based on multi-scale chemical science and was further demonstrated during December and January among CMCS team members and to interested colleagues.

The new pedigree user interface, with an improved textual presentation of pedigree/metadata is shown below in Figure 1. A new feature has been developed, the display of an actual pedigree graph, showing relationships among resources (inputs, outputs, references, and translations). In Figure 2 below, the relationships for showing the pedigree of a computational data file entitled "AtomizationEnergy-CH3OOH.xml". This file was produced to enable new thermochemistry for this molecule, and is part of our multi-scale use-case. Note that the pedigree contains five levels including multiple computations.

There was also extensive testing to identify performance bottlenecks with various configurations of the SAM/Slide server and hardware and database configurations. Also, further steps were taken with the National Institute of Standards and Technology to define schema mappings for integration of their databases into our SAM repository.

Finally, a list of modifications to the infrastructure are listed at the end of this report for more detailed information.

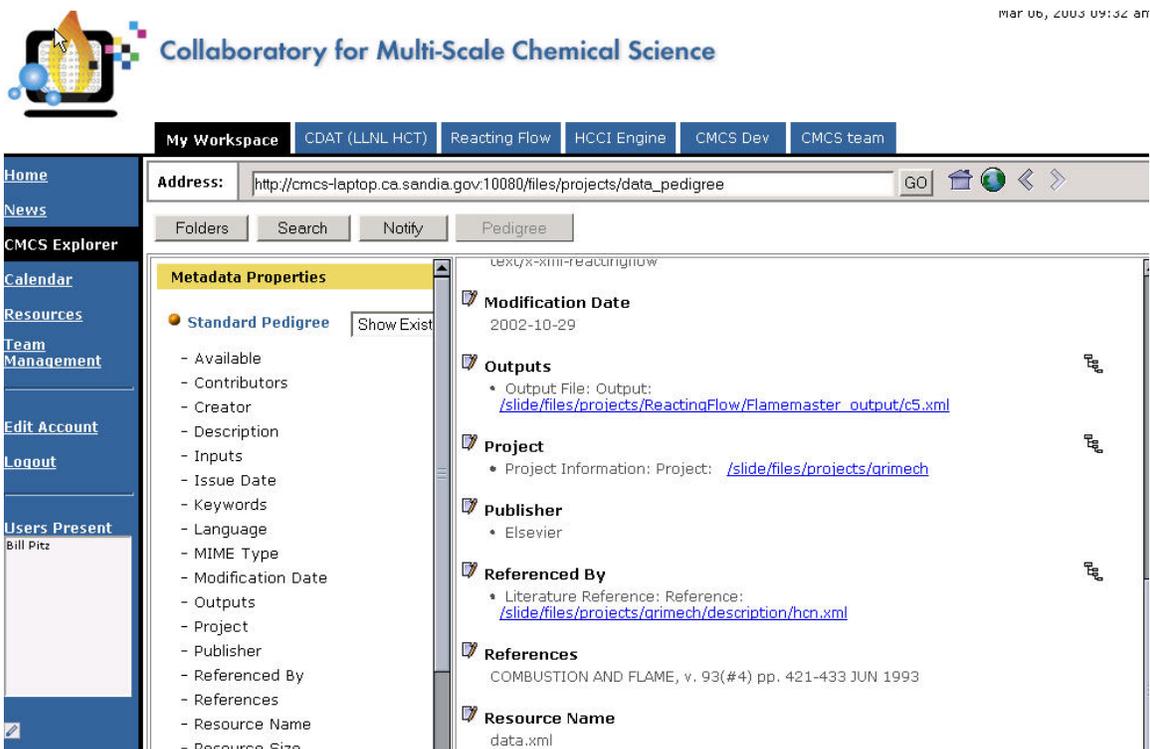


Figure 1. Pedigree user interface, improved textual presentation of pedigree/metadata.

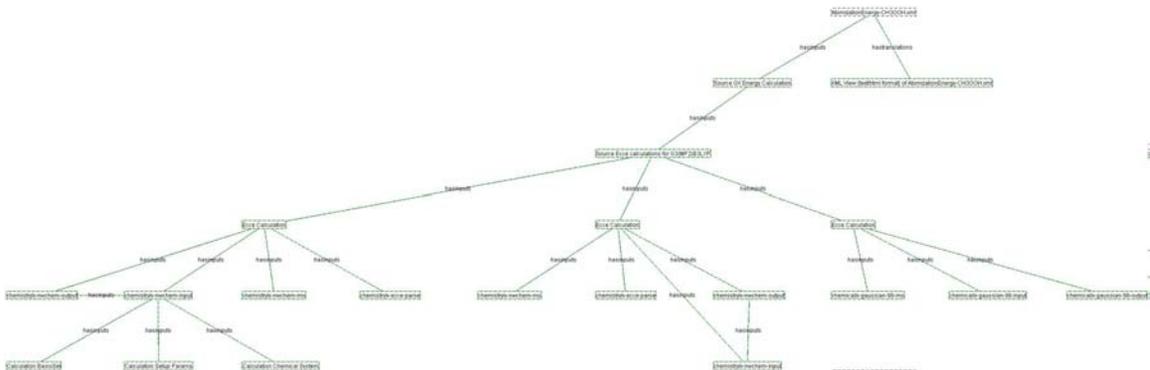


Figure 2. Actual pedigree graph, showing relationships among resources (inputs, outputs, references, and translations).

Specific infrastructure and portal development items accomplished this period include:

- Synchronized the CMCS portal and SAM repository account creation and management functions.
- Modified Chef's user and group interfaces and management capabilities to better meet our requirements for editing account information and group membership and customizing individual workspaces.
- Replaced Chef user and group persistent storage mechanism to external services to allow access to applications and services outside the portal environment (in progress).
- Provided feedback to the Chef team regarding desired features and requested bug fixes.
- Configured access controls for the CMCS SAM repository, modified the portal software to set, modify and delete these access controls based on data management activities and account and group membership interactions.
- Investigated use of SSL connections between the portal server and the SAM repository, currently waiting for a stable version of Apache's http-commons library that is compatible with other third-party portal libraries.
- Filter all urls to the data repository through the portal to avoid reauthentication for each user access.
- Added handling of special characters occurring in url's and data so they could be properly stored and retrieved from the SAM repository and entered/populated within the portal interfaces.
- Improved the file navigation capabilities and basic file operations available within the CMCS Explorer portlet.
- Integrated the JMol molecule viewer into portal.
- Added ability to add and edit metadata values in CMCS Explorer portlet.
- Improved the pedigree user interface interaction and the textual display of pedigree data.
- Developed a prototype pedigree graphing capability, which shows a 2D graph of pedigree relationships (in progress).
- Redesigned the user interfaces for notification requests, data requests and expertise profile.
- Changed the criteria used in notification requests to match the search criteria, to effectively provide a persistent search mechanism via notifications.
- Provided a notification digest capability.
- Improved the message content of notifications to be more descriptive.
- Provided an update capability for portal resources to improve the granularity of some notifications.
- Augmented notification interface to support team-based subscription definitions (in progress)
- Improved the data storage interface (DSI) layer of the CMCS software to support access control methods, standardized error and exception handling, and better encapsulate HTTP/DAV nuances.

- Performed extensive evaluation of the SAM server to better understand performance issues.
- Provided access to the NIST species dictionary through the portal interface.
- Defined preliminary schema mappings to enable the integration of the NIST kinetics database into the CMCS SAM repository.
- Improved the CMCS metadata extraction capabilities to recognize additional chemical properties tagged in data sets.
- Configured production and development CMCS portal and SAM servers.
- Developed a standardized process for deploying a CMCS SAM server.
- Performed integration of SAM's Electronic Laboratory Notebook (ELN) into the portal. Developed ELN launch portlet and incorporated SAM-based Electronic Laboratory Notebook (ELN): ELN can be launched from CMCS portal, can reference CMCS data in notes, shares CMCS data viewers with the Explorer, and produces pedigree information viewable via the Pedigree Browser.

- Integrated the CMCS Explorer portlet and the Active Tables portlet to improve file-based operations between the portal and the Active Tables server.
- Performed evaluation of portal capabilities with a leader of the PriME group in preparation for our initial release for use by this group.