

New Generation of Collaborative Science Portals Enabled by CMCS and SAM

The collaborative infrastructure developed by the Collaboratory for Multi-scale Chemical Science (CMCS) and Scientific Annotation Middleware (SAM) projects was initially deployed to support combustion research. However, the innovative approaches taken, and the early successes in the combustion community, are leading to additional uses in chemistry and reuse of the infrastructure in a variety of projects in other disciplines. In chemistry, the number of groups using the CMCS server at Sandia National Laboratories (SNL) has dramatically increased during the last year, and the National Institute of Standards and Technologies, one of the institutions developing CMCS, has decided to use the CMCS infrastructure to support its new Real Fuels Initiative, and has joined in supporting a community data initiative labeled PRIME (Process Informatics Model).

In biology, CMCS and SAM software will be used within the "Data Portal Enabling New Protein Structure Collaboration" recently funded by NIH and NSF. This 5-year, \$3M effort led by SNL and the University of California, San Francisco (NIH funded, led by Carmen M. Pancerella) and University of Maryland, Baltimore County (NSF-funded, led by Dan Fabris) builds on the core CMCS and SAM infrastructure to construct a knowledge grid for the biomedical community which will be piloted in direct collaboration with scientists leading the development of MS3D, a new method that combines intramolecular chemical crosslinking with high-resolution mass spectrometry to glean structural information about proteins and other biological macromolecules. Within Pacific Northwest National Laboratory's (PNNL) Biomolecular Systems Initiative, the "BioCollaboratory" (led by George Chin) and "Bioinformatics Resource Manager" (led by Eric Stephan) projects are extending CMCS with a rich array of bioinformatics capabilities.

The CMCS infrastructure is also being adapted for use in homeland security as part of a pilot program aimed at determining the best architectures for a nation-wide extensible data management system for the DHS. With the installation of hundreds of radiation monitors around the country to detect possible terrorist threats, Customs and Border Protection has developed an immediate need for an extensible data management system. DHS has tasked SNL, in collaboration with Los Alamos National Laboratory and PNNL, to take on the monumental task of creating the requirements for this distributed, secure, multi-user system. The CMCS infrastructure is being used in this effort as an initial testbed to demonstrate secure links, access controls, multi-user functionality, meta-data searching, data interoperability, and the ability for analysts to wrap in their own tools as they are developed.

CMCS and SAM have contributed to the NSF-funded George E. Brown Network for Earthquake Engineering and Simulation Grid (NEESgrid) as well, with an

Electronic Laboratory Notebook (ELN) component. The ELN, which was integrated with the NEESgrid security and portal infrastructure, and extended to support note-taking in multiple languages, provides NEES researchers with a flexible mechanism for documenting research and securely sharing unstructured information within distributed research teams.