

A Collaborative Informatics Infrastructure for Multi-scale Chemical Science

Larry A. Rahn,¹ Thomas C. Allison,² Brett Didier,³ Michael Frenklach,⁴ William H. Green, Jr.,⁵ David Montoya,⁶ William Pitz,⁷ Al Wagner⁸; Points of Contact

¹Sandia National Laboratories, ²National Institute of Standards and Technology,

³Pacific Northwest National Laboratory, ⁴University of California, Berkeley,

⁵Massachusetts Institute of Technology, ⁶Los Alamos National Laboratory,

⁷Lawrence Livermore National Laboratory, ⁸Argonne National Laboratory

Summary

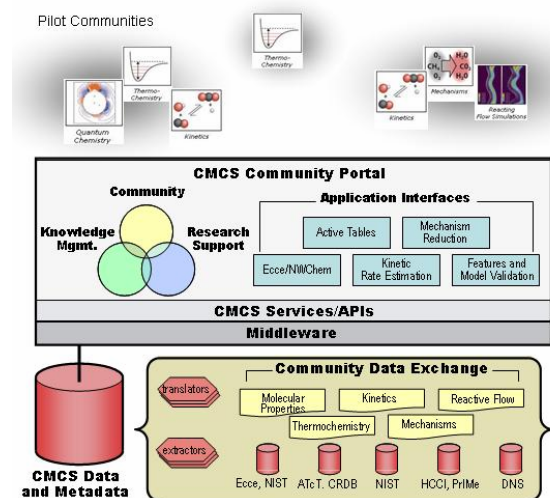
The Collaboratory for Multi-scale Chemical Science (CMCS) is developing a powerful informatics-based approach to synthesizing multi-scale information in support of collaborative systems-based research and is applying this approach to combustion research. An open source multi-scale informatics toolkit is being developed that addresses a number of issues core to the emerging concept of knowledge grids. A number of scientific pilot groups are actively collaborating with CMCS to develop and use capabilities to enable collaborative combustion science, new community-based paradigms for research, and improved access to validated data.

Chemical science, like many DOE science missions, addresses complex multi-scale phenomena. One of the major bottlenecks common to multi-scale research enterprises today is in the passing of information from one level, or scale, to the next in a consistent, validated, and timely manner. The rap-

idly increasing pace at which we can produce data and computational results elevates such data/information bottlenecks to rate-limiting processes that can impact the output of a whole multi-scale research community.

The Collaboratory for Multi-scale Chemical Science (CMCS) project is working to accelerate multi-scale science by prototyping a 'knowledge grid' that incorporates advances in semantic web, informatics, collaboratory, and grid communities. The vision is to enable the transition to a 'systems-science' approach, and to reduce the delay from discovery to DOE mission impact.

CMCS has emphasized the development of data-centric tools supporting knowledge management within an overall the grid-capable portal framework depicted in the figure. CMCS builds upon components from the CompreHensive collaborativE Framework (CHEF), Scientific Annotation Middleware (SAM), Commodity Grid



A schematic view of the CMCS multi-tier approach integrating chemistry data, applications, and informatics tools to enable 'systeme' approach to chemical science research.

(CoG) toolkit and other projects. A number of portlets and underlying services provide a full set of scientific content management operations including support for automated metadata extraction, data translation, cross-scale data provenance management and graphical display, meta-data based search, and multimedia data annotation. These capabilities are supplemented by a suite of collaboration tools and several innovative chemical informatics applications.

CMCS APIs are being used to integrate key combustion community data and application resources and to provide sophisticated portal-based interactive views of combustion data including, for example, molecular structures and XY-graphs, from a variety of widely-used file formats. Overall, the CMCS production server currently provides access to data spanning 5 chemistry disciplines and 10 orders of magnitude in length scale. 25 metadata extractors, 40 translations/views, and several web services simplify data movement and analysis.

A number of national and international scientific expert groups have been attracted by CMCS' vision and unique capabilities and are working as pilot users and collaborators. A 13 member IUPAC (International Union of Pure and Applied Chemistry) Task Group is working to critically evaluate existing data and recommend improved thermochemical values of important radicals. This task group is using the CMCS infrastructure to coordinate and will be using the Active Thermochemical Tables (ATcT) chemical network analysis tool to statistically combine large amounts of data managed in CMCS to produce new reference values for radical thermochemistry. The PrIME (Process Informatics Model) group is a research team of about 40 international scientists that is forming for a similar purpose — to assemble curated kinetics data and to develop optimal reaction

models. A High Quality Electronic Structure pilot group consisting of BES SciDAC researchers has formed to develop community standard benchmarks for assessing the accuracy of computational methods for predicting molecular properties for large and open shell systems. A multi-university consortium, formed to address the challenges in the development of Homogenous Charge Compression Ignition Engines, is working with CMCS to deal with complex chemical kinetic mechanisms and their translation and/or reduction for modeling applications. The NIST Real Fuels Initiative is working with PrIME and the combustion modeling community through CMCS to address problems in combustion chemistry and to make validated data available to industry. A DNS Simulations of Turbulent Combustion group, involving two more BES SciDAC projects working with the Scientific Data Management SciDAC center, is advancing a combustion feature analysis capability for large reacting flow simulations data sets.

Together, these interacting pilot groups represent a significant fraction of leading combustion research efforts and they are poised to have a revolutionary impact on the field. CMCS continues to extend capabilities, and is working to expand provenance to include sensitivity and error information and to develop community peer review mechanisms. The team will continue to address issues related to knowledge grid research while developing an operational, scalable, and sustainable community research resource.

For further information on this subject contact:

Dr. Larry A. Rahn
Sandia National Laboratories
PO Box 969, MS9051
Livermore, CA 94551

Phone: 605-555-1234
E-mail: rahn@sandia.gov
Website: www.cmcs.org