



New Paradigms for Collaborative Multi-scale Science

Imagine a cross-disciplinary, geographically distributed project team that wants to model combustion under conditions found in a new type of internal combustion engine: homogeneous charge compression ignition (HCCI). To properly treat ignition kinetics for methane in the HCCI engine, new chemical species must be included. Reactions involving these new species with accompanying rate constants must also be added to the detailed reaction mechanism. One of the species that is expected to significantly affect the predictions for the HCCI ignition event is methyl peroxy radical (CH₃OO). But the thermo-chemical properties of this transient molecule are not well known, presenting a problem to a team member at Sandia that who seeks to model chemical reactions involving methyl peroxy.

How does the researcher go about finding, validating and utilizing the new information that is needed? The conventional approach involves literature searches, contacting colleagues, and personal knowledge of unpublished research results. At the same time, the rapid rise of the Internet and web application technologies have not been fully exploited by the scientific community.

Collaboratory for Multi-scale Chemical Science (CMCS)

The goal of the Collaboratory for Multi-scale Chemical Science (CMCS) is to enhance chemical science research by breaking down the barriers to rapid sharing of validated information and by opening new paradigms for multi-scale science. To accomplish this, CMCS is developing an a data-centric adaptive informatics infrastructure and is demonstrating proof-of-concept by publicly deploying an integrated set of key collaboration tools and chemistry-specific applications, data resources, and services. This becomes a problem-solving environment for the scientist that also facilitates data sharing and collaboration.

CMCS is piloting new infrastructure and data-sharing concepts among a multi-disciplinary team of chemical scientists working to advance combustion science. The research of the chemical scientists is focused on different physical scales in the combustion problem, and the scientists are geographically distributed. The physical scales range from the electronic structure of atoms and molecules to direct simulations of turbulent combustion phenomena that occur in engines or industrial processes. The proof-of-concept capabilities that CMCS is building are also applicable to many other research areas.

Web Portal

A portal serves as the web interface for the adaptive informatics infrastructure being developed by the CMCS team. The data infrastructure takes advantage of a variety of many 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. Figure 1 shows how CMCS users interact primarily with the top layer, the CMCS portal and chemistry applications. The applications can appear within the portal or provide their own user interfaces that interact directly with the underlying metadata, data, and other CMCS services.

Data pedigree is at the heart of the CMCS project and is a key technology enabling new approaches to science. It allows researchers to categorize and trace scientific data across disciplines and scales and to identify the ultimate origin of scientific data. We have developed a DAV-aware pedigree browser (see lower screen in Fig. 2), which can easily display pedigree data (as well as annotations) and allows users to search, browse, and retrieve a data set's pedigree. DAV, or WebDAV (Web-based Distributed Authoring and Versioning), is a set of extensions to

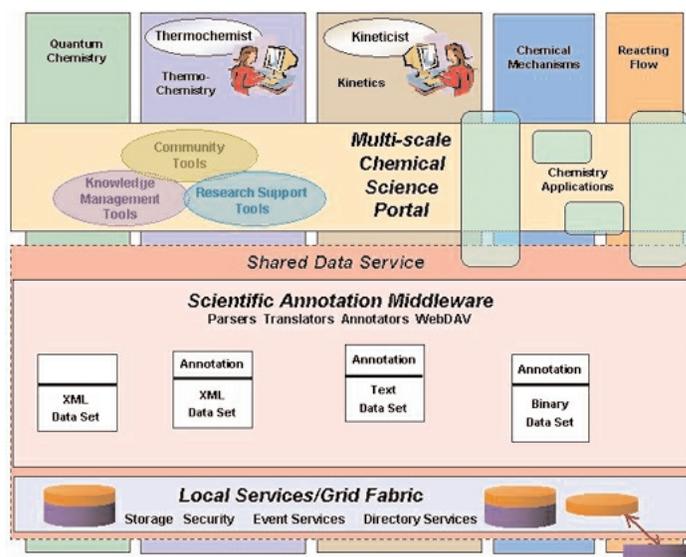


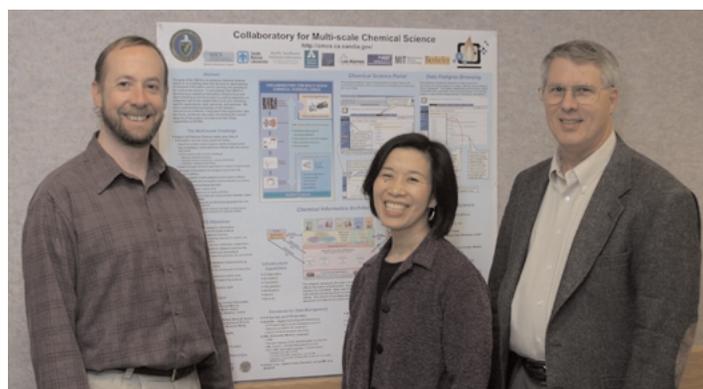
Figure 1. A diagram representing the major conceptual elements of the CMCS informatics infrastructure.



the HTTP protocol that allows users to collaboratively edit and manage files on remote web servers. Pedigree data and metadata are associated with CMCS resources by placing this information into the DAV properties of the data file in the CMCS data store. Pedigree data may also be an active link to a different, but associated, data resource.

How CMCS Works

In this new environment, our Sandian is part of a multi-scale team sharing a project space in the CMCS portal. She poses the problem to a quantum chemist at PNNL who then initiates a computation of the bond energies of the methyl peroxy molecule. Soon she is automatically notified that the results are available in the common data store, complete with metadata and links to the information about the methods that produced it. Now she can collaborate with another team member at ANL specializing in developing thermo-chemical properties from such information that is consistent with other known relevant information. She has access to his Active Thermo-chemical Tables (ATcT) application through a portlet in the project team workspace in the CMCS portal (see upper screen in Fig. 2). The displayed ATcT portlet shows the network used to produce an optimized table of thermo-chemistry data for the methyl peroxy radical. The output of this application is also deposited to the data store. The XML (eXtensible Markup Language) format allows automatic translation into the format needed to fit the data and derive the param-



David Leahy (from left), Christine Yang, and Larry Rahn (not pictured, Carmen Pancerella)

David Leahy is a Senior Member of the Technical Staff in the Reacting Flow Research Department (8351). His background includes basic research in chemical reaction dynamics as well as development of Internet software and services.

Christine Yang is a Principal Member of the Technical Staff in the Science and Engineering Problem Solving Environments Department (8964). She is one of the two technical leads for CMCS.

Larry Rahn is a Senior Scientist in the Combustion and Physical Sciences Center (8300). He directs the CMCS project and his background includes management and basic research in laser-based diagnostics and reacting flow science.

Carmen Pancerella is a Principal Member of the Technical Staff in the Science and Engineering Problem Solving Environments Department (8964). She leads the pedigree task group for CMCS.

ters needed for the reaction modeling. In fact, this routine operation is automatically accomplished by processing routines that are started upon notice that this type of data has been deposited. This new information is available long before the publications of the science that produced it, speeding the HCCI research team towards its goal.

CMCS Project Team

The Distributed Information Systems Center (8900) and the Combustion and Physical Science Center (8300) have partnered on two collaboratory projects funded by the Office of Mathematical, Information, and Computational Sciences (MICS). Sandia is the lead on the current DOE National Collaboratories CMCS pilot project, which is a partnership of eight organizations—SNL, PNNL, ANL, LLNL, LANL, NIST, MIT, and UC Berkeley.* Sandia project team members include Larry Rahn, Christine Yang, Carmen Pancerella, David Leahy, Wendy Koegler, Michael Lee, Renata McCoy and Edward Walsh.

* (SNL: Sandia National Laboratories; PNNL: Pacific Northwest National Laboratory; ANL: Argonne National Laboratory; LLNL: Lawrence Livermore National Laboratory; LANL: Los Alamos National Laboratory; NIST: National Institute of Standards and Technology; MIT: Massachusetts Institute of Technology)

The screenshot shows the ATcT application interface within the CMCS portal. At the top, there is a search bar for 'Chemical species' with the example 'CH4'. Below this is a pedigree diagram showing the reaction network for the methyl peroxy radical (CH₃OO). The diagram starts with CH₃ and O₂ at the bottom, which react to form CH₃O and O. These then react to form CH₃OO. The diagram is annotated with '0' and '1' in red circles, indicating specific reaction steps or states. Below the diagram is a metadata table with fields for Creation Date, Publisher, Resource Type, Keywords, Creator, Title, and Abstract.

Figure 2. ATcT accessed from the CMCS portal.