

## Sparking a Revolution in Data Sharing<sup>1</sup>

Why isn't sharing data with remote collaborators as easy as saving it to a local disk? For users of the Collaboratory for Multi-scale Chemical Science (CMCS), it is. With CMCS, researchers can also see which files were used as inputs to produce derived data sets, automatically translate files to other formats, search for data based on metadata (e.g. by author, experiment name, or chemical species involved), pull up interactive views of data in a web browser, and annotate data with additional comments and links to other data, (see figure) all while working with groups of colleagues that may be distributed around the globe.

These capabilities, which lower the barriers to large-scale collaboration are sparking a revolution in the way chemical information

is assembled, validated, and conveyed to users. Pilot CMCS user groups are sharing everything from basis sets for quantum chemistry calculations to thermodynamic and kinetics reference data, to optimized mechanisms for modeling combustion. And they are sharing this information across sub-disciplines, with the details of how it was derived and the tools to allow anyone to review or improve it! "Traditionally, development of chemical kinetic mechanisms is done in isolation by individual researchers. It has been too difficult to share all of the inputs and the reasoning that went into its development, or to compare it with other mechanisms. CMCS is changing that." says Dr. Bill Pitz, a chemical scientist at Lawrence Livermore National Laboratory.

CMCS' advanced capabilities build on services provided by CHEF collaborative portal toolkit and the Scientific Annotation Middleware (SAM). SAM provides an extremely flexible repository for scientific data and associated metadata with layered capabilities for managing metadata, semantic relationships between data sets, and notebook-style annotation and research documentation. SAM enables queries and display of metadata and data produced by independent applications. SAM also allows CMCS to automate the tedium of documenting data and interpreting data in custom formats. The net result: CMCS has been able to provide users with a rich suite of collaboration capabilities and chemistry-specific tools, with single sign-on access across the portal and for data access within users' desktop applications, and to rapidly integrate new tools and support new data types as users' needs evolve and grow.

For the growing CMCS user base, the ability to concentrate on science and ignore the logistics of collaboration, while avoiding the up-front costs of data standardization, is more than a convenience, it's an enabler of new types of coordinated research efforts.

<sup>1</sup> Point of contact is Larry Rahn [rahn@sandia.gov]. The CMCS and SAM projects are funded by the U. S. Department of Energy Office of Science, Mathematical, Information, and Computational Science Division of the Office of Advanced Scientific Computing Research.