Collaboratory for Multi-Scale Chemical Science

Status as of October, 2002 / Quarterly Report for Q3 and Q4 of FY 2002

Project Staff

Larry Rahn-SNL*, Christine Yang, John C. Hewson, Carmen Pancerella, Wendy Koegler, David Leahy, Michael Lee, Theresa Windus-PNL*, Brett Didier, James D. Myers, Karen Schuchardt, Eric Stephan, Al Wagner-ANL*, Branko Ruscic, Michael Minkoff, Lee Liming, Sandra Bittner, Gregor von Laszewski, Brian Moran, Reinhardt Pizon, S. Nijsure, K. Amin, B. Wang, William Pitz-LLNL*, David R. Montoya-LANL*, L. Xu, Y.-L. Ho, Thomas C. Allison-NIST*, William H. Green-MIT*, Jr., Michael Frenklach-UCB*

* denotes Institutional Point of Contact

Summary

During this performance period the pilot Collaboratory for Multi-Scale Chemical Sciences (CMCS) made significant progress towards the development of Version 1 of the software infrastructure - including the portal and data infrastructure as well as chemistry applications.

An evaluation of the technologies and requirements from application use cases and technology options developed in the first two quarters of the project was carried out. Decisions were made on the infrastructure and application interfaces that we would further prototype to demonstrate the central goals of the project.

The project team held a workshop to communicate these results in June. It was held in conjunction with an advisory board meeting and a BES SciDAC workshop. At the CMCS workshop, the high-level goals of the project were broken out into a work breakdown structure to enable the development of an integrated implementation of CMCS software and data. The CMCS project was presented to the BES SciDAC team and the Advisory Board concurrently at the end of the workshop.

The results of the workshop were further focused into seven demonstration use cases to be presented at SC2002. The latter months of this reporting period involved intense focus on the development of these demonstrations, their requirements, the software for the infrastructure and applications, and the collection and translation of needed data. The CMCS Infrastructure Workshop was held in August 2002 to develop a plan to implement the CMCS architecture for SC2002. Significant advantage is being taken of the Scientific Annotation Middleware project, as well as the CHEF (Univ. Mich.) project in this implementation.

The CMCS project was also represented in numerous other technical contexts. These include presentations made in poster sessions in the Turbulent Nonpremixed Flame Workshop and the 29th International Combustion Symposium, both held during July 2002 in Sapporo, Japan. Project team members also participated in the MAGIC Workshop and in the High Performance Network Planning Workshop, both held during August, 2002.

Progress

This document summarizes the work done over the third and fourth quarters of the CMCS project. A summary listing of project activities is presented, then the results of planning activities are summarized, and, finally, summaries of CMCS Working Group technical accomplishments areas are summarized.

Summary of CMCS Project Activities – April 2002 through September 2002

- Established formal collaboration agreement with University of Michigan CHEF Team (4/2002)
- Provided draft CMCS pedigree documents to Earth Science Grid System project (4/2002)
- Presented CMCS pedigree tutorial to CMCS Application Working Group (4/2002)
- Investigated digital library technology (4/2002)
- Researched scientific publishing (4/2002)
- Drafted guidelines for using Dublin Core XML in CMCS (4/2002)
- Researched mapping between DAV properties and Dublin Core properties (4/2002)
- Modified Active Tables to create XML Extended Species Dictionary (4/2002)
- Defined data properties needed for pedigree and search (4/2002)
- Discussed translation services to be provided by SAM project (5/2002)
- Established and configured project software development environment (5/2002)
- Began investigation of portlet characteristics and behaviors (5/2002)
- Developed CMCS Project Charter (5/2002)
- Drafted CMCS Project WBS (Work Breakdown Structure) (5/2002)
- Compiled current list of available WebDAV client resources (5/2002)
- Drafted general portlet reference document for developers (5/2002)
- Developed and documented security use cases for CMCS (5/2002)
- Put LLNL thermodynamic and reaction rate data on DAV server (5/2002)
- Put demo CRDB data on DAV server (6/2002)
- Began developing documentation on portal installation (6/2002)
- Discussed publish/subscribe/broadcast model for CMCS (6/2002)
- Started examining portal design issues in the context of implementation (6/2002)
- Began discussion of open source software for CMCS project at SNL (6/2002)
- Held two day CMCS Workshop II at ANL (6/2002)
- Held meeting with CMCS Advisory Board at ANL (6/2002)
- Coordinated and participated in two day BES SciDAC Workshop at ANL (6/2002)
- Documented CMCS WBS Task Info from CMCS Workshop II (6/2002)
- Drafted initial definitions of CMCS demonstrations for SC2002 (6/2002)
- Presented poster at 29th International Symposium on Combustion in Sapporo, Japan (7/2002)
- Created draft of portal objects definitions and relationships (7/2002)
- Held two day CMCS Infrastructure Workshop at PNNL (8/2002)
- Set up developers portal for testing (8/2002)
- Developed prototype for CMCS Explorer portlet (9/2002)
- Revised CMCS guidelines for defining pedigree and other metadata using DAV properties and XML for SC2002 (9/2002)

- Drafted portal objects schema descriptions (9/2002)
- Discussed and defined Active Tables integration with CMCS (9/2002)
- Created PURL for CMCS (http://purl.oclc.org/NET/cmcs/) (9/2002)
- Discussed and provided input to SAM project on CMCS requirements for SC2002 (9/2002)
- Provided feedback to SAM project on translation functionality (9/2002)
- Defined keywords and chemistry properties for search (9/2002)

Project Management, Structure and Planning

The CMCS management structure has been developed and previously documented in the CMCS Management Plan, and in more detail in the last report. CMCS leadership team (Director Larry Rahn, Chief Technology Officer Jim Myers, and Chief Integration Officer Christine Yang) has worked to evolve the tasking and focus of the project team as the project has evolved. The Point of Contact (POC) team has served to coordinate site-related issues, and some tasking assignments. The work of the three CMCS working groups: application, portal, and infrastructure, have become more integrated as their work has progressed. The portal working group is now completely integrated with the infrastructure group (as the major portal technology decisions have been made), and the nature of the tasks recognized in the application-working group has changed significantly.

In preparation for the Project Workshop in June, the CMCS leadership team assembled a Project Charter that delineates the high-level goal and supporting objectives of the project. This Charter supplied the framework for a draft work break down structure (WBS) which was drafted during our June workshop. The WBS activity was a useful step in our planning, but also indicated a critical need to further focus the CMCS project team activities. To achieve this, we focused on developing integrated demonstrations of application use cases. This focus on integrated demonstrations has proven to be energizing, productive, and a natural way to prioritize the huge array of development tasks indicated in the WBS. The CMCS charter and demonstrations are summarized next in this report, follow by a more technically detailed description of progress taken from workshop results.

CMCS Charter

The complete charter is available in the public documents folder on the project website, but the goal, objectives, and scope statements are repeated below.

Goal

Enhance chemical science research by developing an adaptive informatics infrastructure and demonstrating proof-of-concept by publicly deploying an integrated set of key collaboration tools and chemistry-specific applications, data resources, and services.

Objectives

Architect and build an adaptive informatics infrastructure.

Define and build proof-of-concept capabilities that demonstrate the value of an informatics infrastructure in increasing collaboration and coordination across disciplines, and chemistry scales.

Demonstrate the power of an adaptive infrastructure as opportunities arise by supporting changes in CMCS standards as the project evolves, and by integrating with external projects without requiring adoption of CMCS schema, etc.

Develop and execute a plan for community outreach that includes demonstrations of significant new capabilities to attract participation and continued support.

Evaluate and document the success of the project in terms of community interest and support, technical capabilities, etc. and recommend a continuation path to key stakeholders.

Scope

The project is a proof of concept of the "adaptive informatics" approach. It includes the development of the infrastructure, the development and integration of specific chemistry data and applications that are representative of or key to the community, and pilot use by the community. As a pilot effort, scaling beyond what is necessary for proof of concept is out of scope although architecting to support scaling is within scope.

CMCS Demonstrations

The CMCS SC2002 demos were designed to focus on combustion applications and highlight the value of CMCS for enhancing combustion research. The concepts behind the demonstrations were developed, then a Main Demonstration, and several In Depth Demonstrations were planned. The concepts and demonstrations are briefly described below.

Concepts

CMCS improves the data access for chemical sciences researchers by:

- Providing ubiquitous access to community repositories of annotated data.
- Providing mechanisms to automate the translation of data between formats and to/from standard formats.
- Providing mechanisms to automate the generation of annotations, particularly pedigree information.
- Providing an extensible set of tools for data analysis, comparison, mining, and visualization.

CMCS increases the speed at which the community adopts new data values and techniques by providing:

- Automatic notifications, powerful search capabilities, and collaboration tools.
- A rich, fine-grained, and dynamic model for publication and evaluation of data.

CMCS directly facilitates scientific collaboration by providing project/community management tools relevant to researchers that are integrated with the public CMCS data and capabilities.

CMCS will support the transfer of knowledge across scales and disciplines in a consistent and optimized manner by providing new technologies that formalize and automate the handling of data dependencies, error bars, and sensitivities across scales.

CMCS will evolve ‘at the pace of research’ by minimizing the up-front costs and coordination required to support new data types and science applications. This will be accomplished through the development of a flexible, lightweight, open-standards-based integration architecture.

Demonstrations

Main Demonstration, Collaboratory for Multi-scale Chemical Sciences (CMCS)

The main demonstration will highlight all of the Demonstration Concepts points briefly in the context of thermodynamics.

In-depth demonstrations are to target additional or expanded chemistry contexts and should sketch the infrastructure and science connections to the main demo. In-depth demos will be ‘vertical’ in the sense that they can be presented alone, without supporting material. An initial, representative set of in-depth demos was developed, and is listed below.

In-Depth Demo 1. Chemical Data Access

Investigating the pedigree of a mechanism (GRIMECH3) – browsing back to input data and literature refs, technique used, etc. Doing a search across multiple data sources to discover inputs for a possible GRIMECH 4.

In-Depth Demo 2. Chemical Data Analysis Workflow

Demonstrate how ubiquitous access to data in easily translated formats can lead to an improvement in research productivity. ‘Feature’ metadata created by feature tracking analysis tools will enhance the mining and analysis of data. Large datasets can be mined and analyzed via feature metadata. Feature tracking tools associated with CMCS will be able to read data formats associated with CMCS and output data formats appropriate for annotation of those datasets. A CMCS Feature Browser will visualize feature metadata.

In-Depth Demo 3. Tools for Sharing Combustion Mechanisms

Demonstrate how CMCS facilitates sharing data and application environments across the combustion mechanics community via shared data formats, translators, and simple web-based data visualization tools.

In-Depth Demo 4. Chemical Research Project Management Portals

Through development of project management capabilities of the CMCS portal, CMCS directly facilitates scientific collaboration by providing tools relevant to researchers.

In-Depth Demo 5. Information Transfer Across Physical Scales

CMCS provides new technologies that facilitate transfer of knowledge across physical scales and disciplines. Notification services, data standards and translators will significantly reduce the barriers separating separate scientific communities.

In-Depth Demo 6A. New Collaborative Modes of Research Using Informatics Infrastructure: Active Tables

The adaptable informatics infrastructure being created by CMCS has a synergistic relationship with the development of Active Tables. Together we find the development of qualitatively new modes of collaborative research.

In-Depth Demo 6B. New Collaborative Modes of Research Using Informatics Infrastructure: GRI-Mech

The adaptable informatics infrastructure being created by CMCS facilitates the development of an international collaboration of experts in a given field. GRI-Mech Plus is a prototypical example of the development of qualitatively new modes of collaborative research.

Application Working Group

The following includes vignettes summarized from selected posters delivered by task groups reporting at the CMCS Project Workshop held during June, 2002 at ANL.

Data Flow Between Scales

The purpose of this task group has been to explore and develop applicable modes of data flow between applications and/or participants at various chemical scales. The members are Theresa Windus, Branko Ruscic, Elena Mendoza, Reinhardt Pinzon, Eric Stephan, Karen Schuchardt, John Hewson, Sandra Bittner, and Gregor von Laszewski.

The essence of multi-scale science is the transfer of needed information from one scale to the other. For example, species-specific molecular (spectroscopic) data and species-interconnecting data (e.g. kinetic equilibria, bond dissociation energies, etc) feed directly into the Active Tables (as well as others), which in turn produces thermochemical data that will be used by other applications (e.g. by programs that fit this data and supply polynomials to modeling suites). Therefore, Active Tables represent both a consumer and producer of data. This consumer/producer model continues up all the chemical scales, up to the numerical simulation stage (for the current scope of this project). discussing and documenting several issues related to CMCS capabilities. Among these issues have been the data model definitions (what data needs to be transferred across scales, what data is useful essentially only at one scale, etc.), data schemas to be used, work flow issues with the portal (these have been documented in the form of use cases and other associated documents), data storage and transfer using the DAV protocol, translation issues between legacy codes and newly developed codes that will use the CMCS data schema, search capabilities within the CMCS and with other public chemical databases, and data/model comparison.

The focus has been on the development of Active Thermochemical Tables (ATcT) and recent progress is outlined below;

- Data stores: The hierarchical organization and data sharing aspects (XML) of information for the ATcT input datastores has been blueprinted.
- ATcT Kernel Engines: (Q engine, Search engine, Network engine, Output engine, etc) development includes ~ 15 000 lines of Fortran 95 code so far. The Q engine (calculatingQ function used in JANAF and Gurvich et al.) is currently nearing the end of phase I. Search engine version 1.0 is in place and the Output engine is in progress.
- Test network: Assembled principal data needed to define the thermochemical network that includes GRI-Mech species (Excel file). Assembled small network pertinent to refinement of $\Delta H°f(OH)$ (Excel) (combined MICS and Chemical Sciences effort). For testing purposes developed Excel add-in that implements a simplified network solution strategy.
- GUI and ATcT Portal: Early stage development is a JAVA Application (can be modified into JetSpeed portlet or HTML). It creates a portal environment to ATcT, launches ATcT, passes input, displays output, launches dependent applications (e.g. Chemkin III FitDat), and graphically displays data (network, fit results).
- Graphical network visualization: This is an important component of the ATcT user interface that can be used during network analysis and is in early development using Graph Viz Engine.

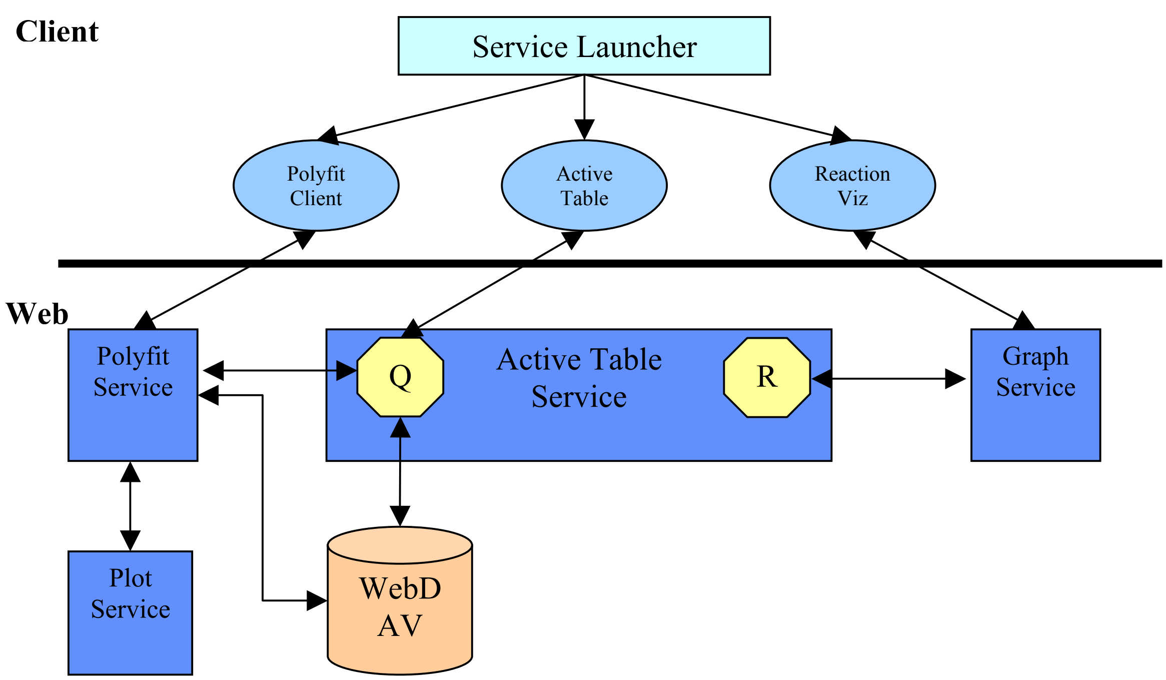

Figure 1. Architecture for Active Table Service.

Data Modeling and Translation

Collaborating researchers each have access to his/her own preferred set of analysis tools. Difficulties arise because researchers typically expend a great deal of effort in making data and models compatible with their unique tools. The focus of this task is to define how CMCS will provide an environment to ease translation by taking advantage of certain technologies described in this poster. We must understand data models in order to translate between formats and develop a means of translation that is much easier than current practice. Team members are John Hewson, Bill Pitz, and David Leahy.

The eXtensible Markup Language (XML) is a widely supported standard for data description. XML provides a framework for generating a sufficient descriptive format for models and data. Core elements allow ease of exchange for basic information. XML descriptions are readily extensible allowing change as necessary. Extensions build on core elements. XML elements can be given definitive meaning by associating them with a definition following XML/RDF (Resource Description Framework). Pedigree associated with models will be described primarily with existing frameworks: Dublin Core (DC) and RDF. These indicate relationships between models/data and their creation. XSLT (eXtensible Stylesheet Language for Translations) provides a means of translating data from XML to arbitrary formats.

Accomplishments:

- Draft XML format describing Species and Reactions and Reacting Systems designed.
- Draft schema and object models developed for Reacting System descriptions.
- Translators available from Chemkin, FlameMaster and LLNL data formats into XML formats. Next step is the translation out of XML format into native file formats.
- Entire LLNL Chemistry database has been converted to XML (Thermodynamic parameters for 4,000 species and rate constants for 16,000 reactions)
- Utility program written to produce property data for DAV (David Leahy)
- Representative data put into DAV.

Feature Tracking in Numerical Simulation Datasets

Feature tracking is a data mining approach with the motivation to extract further scientific understanding from valuable DNS data sets. This task is led by Wendy Koegler. There are significant challenges, including a multitude of different data formats, large data sets that are difficult to move, and requirements for diverse types of search criteria (e.g. Which species?, What code?, Input summaries, Results summaries, etc.)

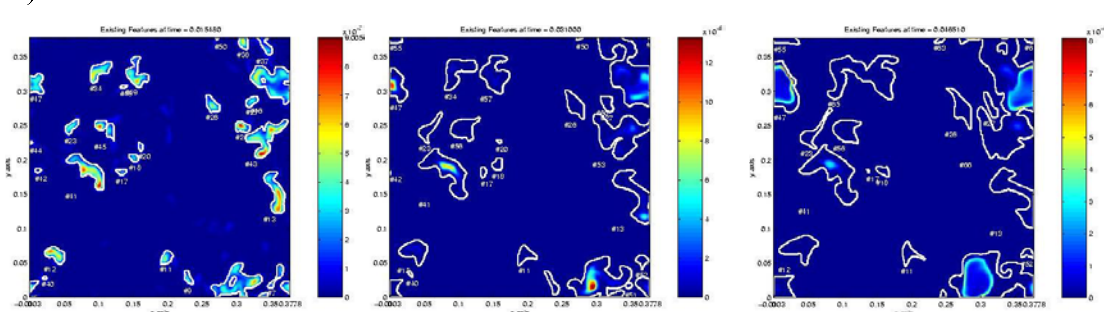

Figure 2. Features (outlined with white lines) of the H_2O molecule concentration tracked versus time in a DNS of hydrogen-air autoignition.

The capabilities presented by CMCS will enable many people to mine data created by a few, enable searches for data using metadata (pedigree, feature properties, etc.), and will enable data analysis & comparisons to be made easily.

A prototype CCA-based feature tracking capability is now in place and has been used to analyze a hydrogen-air autoignition (data courtesy of J. Chen, T. Echekki) problem by tracking concentration isocontours of the H_2O molecule (see figure).

We have also developed a way to visualize the time-dependent fate (showing mergers, death, birth, etc.) that is helpful in analyzing the parameters set for the identification of features.

The goals for future work include:

- Finish development of Feature Tracking software
- Input & Reference DNS data in CMCS repository
- Annotate DNS data with feature metadata
- Configure 'Search' to support feature metadata
- Integrate a 'Feature Viewer' for feature browsing

GRI-Mech - Data Collection

A CMCS WebDAV collection of GRI-Mech data was assembled by M. Frenklach and D. Leahy. An example of the xml description file for HCN is shown

- Initiated translation of GRI-Mech reaction data into CMCS-DAV XML format. The data is organized into 5 groups; chemSpecies, chemElements, thermo, description, and references. We have finished translation of about 40 % of the complete GRI-Mech data set.
- Developed an HTML-JavaScript Wweb application to demonstrate archival, retrieval, and processing GRI-Mech thermodynamic data stored on CMCS-DAV site in an XML format.
- Organized an international collaboratory group on modeling combustion reactions (PrIMe - Process Informatics Model)
- Organized a special session of the PrIMe collaboratory at the International Symposium on Combustion, held in Sapporo, Japan, on July 21-26.

Portal and Infrastructure Working Groups

Portal Working Group

CMCS project's most central requirement involves data/information access and publishing through the portal. This working group is generating requirements of user needs for data access, management, summary, and notification. Individual scientists will be developing applications that are being invoked

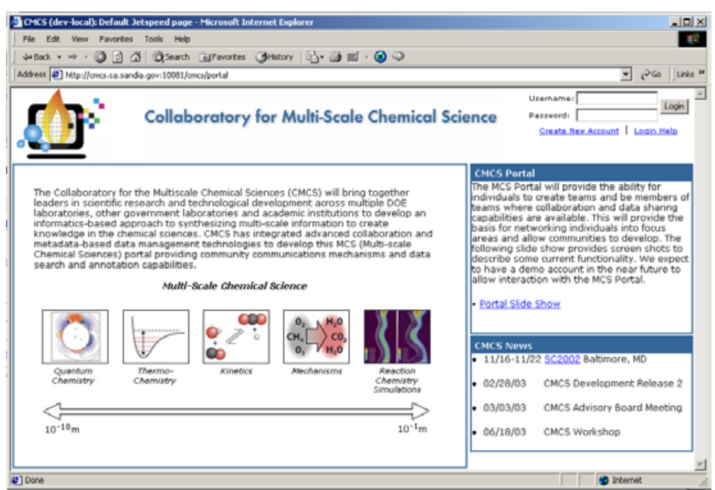

Figure 3. Prototype of CMCS Main.

outside the portal and would interact with a DAV server directly; the portal will be a reporting and management window into different aspects of data. The portal will also be playing the role of file manager / document manager window to DAV servers.

In the previous reporting period the portal evaluation task group identified Apache Jetspeed as base for the Portal User Environment. Accomplishments in the past two quarters include;

- Established a CVS control and build environment on CMCS computing resources
- Initial level of product capability understanding and trial applications -Base level of product tutorials
- Established collaboration with University of Michigan CHEF project.
- Initial portlet development utilizing webDAV data access with velocity portlet
- Reported progress and plans in a poster at the June, 2002 CMCS Project Workshop
- Attended the CHEF tool review and training, and decided to implement their Jetspeed implementation and collaboration suite under the CMCS interface and add the CMCS Explorer product created for access to DAV data objects.
- A Jetspeed portal prototype has been developed for the CMCS project main website. This portal presents a login interface, project information, and demonstrates the implementation of portlet technology. See Figure 3.

A dynamic and configurable portal framework is required for users. Users must be able to establish teams and configure their space to fit the work goals and the approach that they want to use. Users will also have access to data, research and expertise in focused

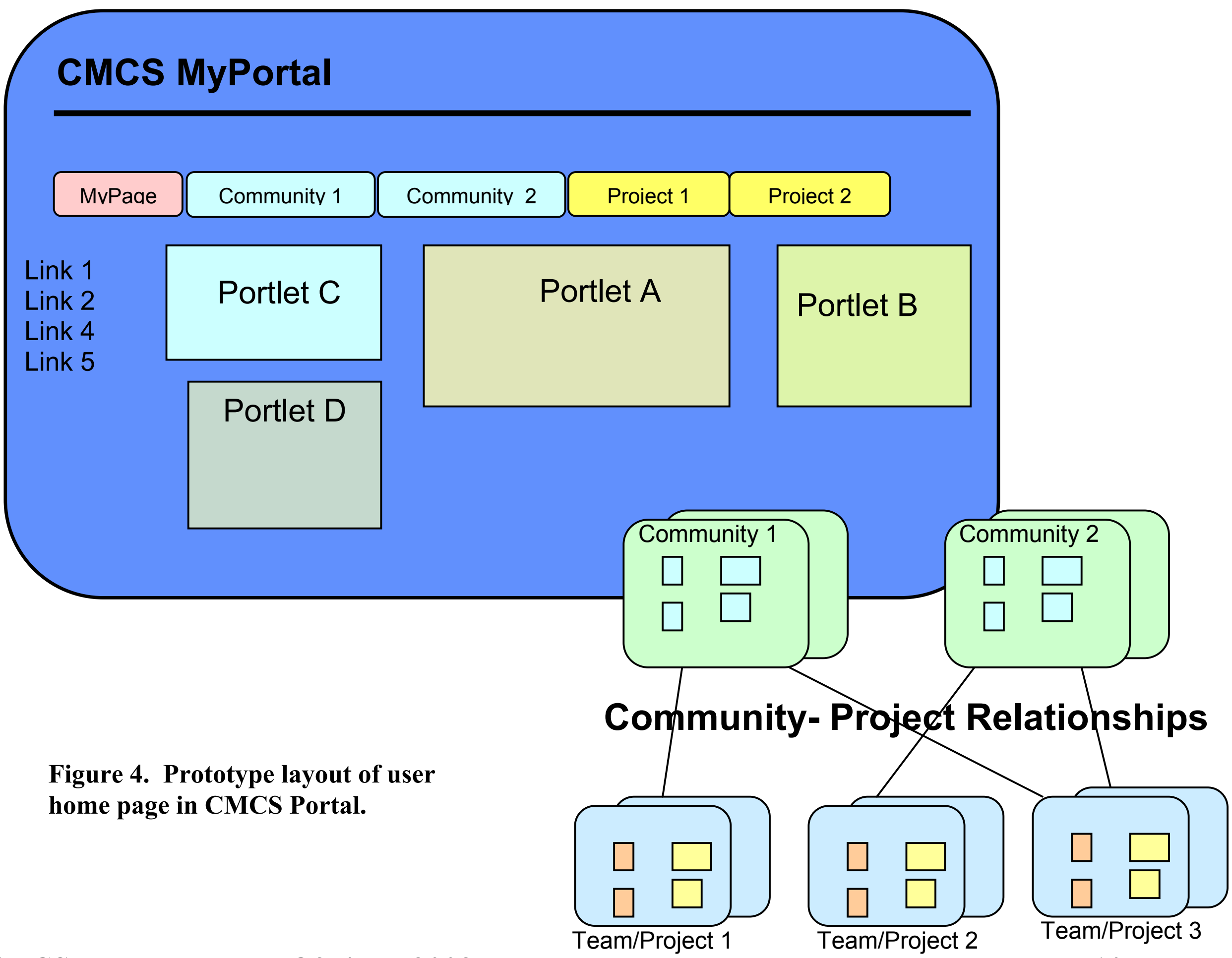

Figure 4. Prototype layout of user home page in CMCS Portal.

(either specific project or community) areas of the portal and have the mechanism to search within and outside the portal. Services will be established within the CMCS that the Portal will access to integrate tools and data for the needs of the user. The vision is not a static portal, but one where the whole environment is managed, updated and grown by the user base rather than by administrators.

A proposed Portal User Environment is shown in Fig. 4. Tabs (MyPage, Community 1, etc.) provide easy access other Portals representing groups and/or projects in which this user participates. Portlets can be selected that may include tools, apps, data access, news, collaboration, etc. In this example, the user has a page configured for their own use (MyPage) that includes portals for two Communities and two Project teams. Also depicted is the multiple relationships possible between user communities and project teams.

Security Task Group

The CMCS project must consider security issues in its design from the very beginning. Too often a complicated and ambitious project can find out late in the development cycle that neglected security concerns wind up requiring huge resource allocations or significant de-featuring of the project. Task group members are David Leahy, Sandra Bittner, Gregor von Laszewski, Karen Schuchardt.

CMCS must be easy to use in order to be successful with the larger goal of attracting researchers and information science developers. Analysis of the security challenges associated with possible avenues of development have helped to define the scope and the timeline of the CMCS project as a whole.

Our goal is a security implementation that facilitates research by assuring the users that their concerns are addressed, while minimizing the costs of security development and administration. We have documented this vision, outlined below, a typical use case, and specific use cases for a Active Tables, and Notification.

Security Landscape and Vision:

- Data-Centric Project: While the CMCS has many facets, at its core it is a data-centric content management and publishing system. Thus, we have initially focused our security effort on securing a WebDAV server.
- Secured Workspaces: CMCS will supply chemical sciences research projects with secured workspaces and secured communications. Initially, widely-adopted commodity protocols such as HTTP and SSL will be used for prototyping.
- Data Publishing with Non-Repudiation: In the first prototype, published data will be globally readable. It will be digitally signed and time-stamped, and safely archived. This comprises a high standard for permanent data storage and publishing.
- Integration of Remote Computing Resources: CMCS will follow a Web Services model, where remote sites that access the CMCS data store will typically be responsible for securing their own assets; CMCS will secure their data.
- Notification: CMCS will provide notifications on a subscription basis to users and applications that want to know about updates. Resulting security implications (especially privacy concerns) will be addressed.

Future Plans for CMCS Security

Implementation-Independent Security Model (IISM): We plan to abstract the authentication and authorization process from the specific technologies that need to be secured (initially, the WebDAV data repository, and the CMCS portal). This abstraction will facilitate the use of additional authentication and authorization schemes for an existing infrastructure without having to update or modify multiple web services.

Migration Towards Single-Sign-On: It is desirable for all web services associated with the CMCS to enjoy a single sign-on. We will continue to assess the costs and benefits of this feature as our web technologies mature.

Support Additional Authentication/Authorization Schemes: It is desirable to extend the initial Implementation-Independent Security Model to allow configurations that support and/or require Public Key Infrastructure (PKI) and sophisticated Grid-based security. Such additional schemes will extend the reach of CMCS to communities that have additional security requirements.

CMCS Data/Meta-data Management

Goal of this task is to provide data/metadata storage and management services to CMCS applications and infrastructure developers and end users. The team members are Eric Stephan (lead), Brett Didier, Elena Mendoza, Carmen Pancerella, Theresa Windus with contributions from Tom Allison and David Leahy.

Metadata is important because it semantically describes data objects in the language of the scientist. Metadata makes binary or text data objects quickly searchable, can be used to show data object pedigree, it facilitates exchange (manual and automated) of data

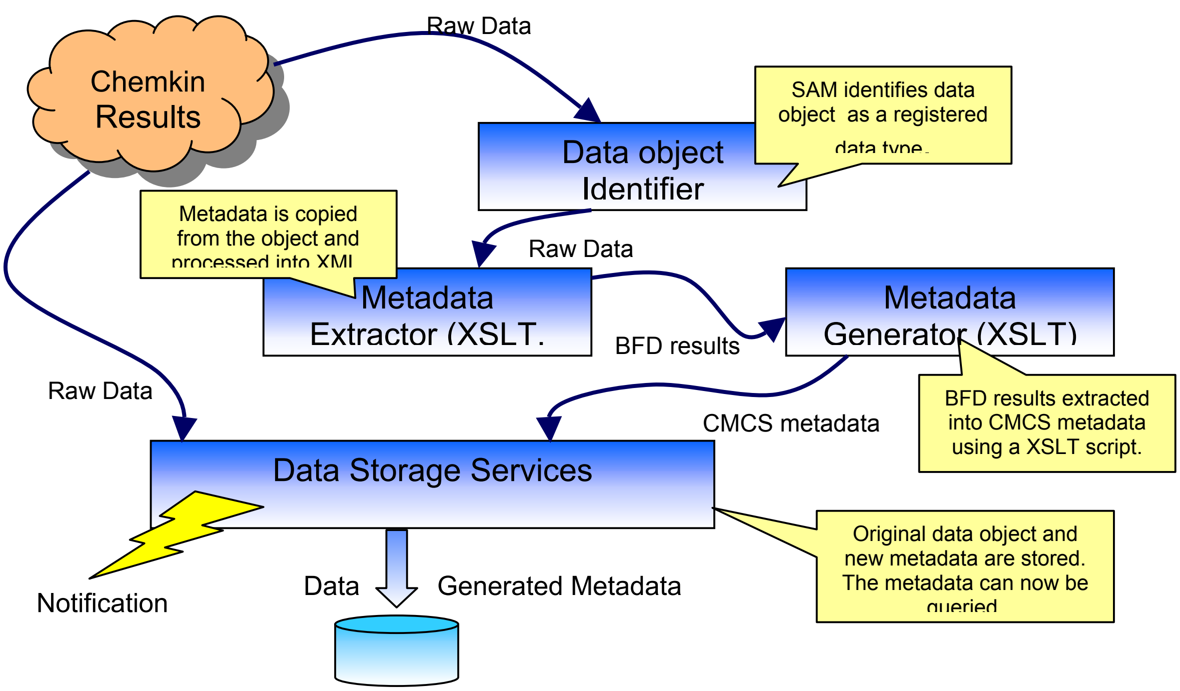

Figure 5. SAM metadata generation for a typical Chemkin application result.

across scientific boundaries, it makes data discoverable, and it enables automatic translation and notification capabilities.

We will use SAM metadata generation and WebDAV and SAM servers have also been deployed. An example of the path for metadata generation takes place is shown for Chemkin in Fig. 5.

There are a number of data/metadata management tasks in progress. These include:

- Identify and develop client DAV interfaces.
- Assist prototype tasks
- Deploy Metadata Generation Services
- Connect to NIST Repository.
- Work with Search and Pedigree.
- Develop CMCS Translation.
- Create Portal Interface to SAM.
- Configure Slide 2 Server.

CMCS Search

The goal of the CMCS Search task is to provide searching interfaces, tools and services required by CMCS end users and developers to support the finding of data resources (e.g. data sets, documents, references, etc.) and the integration of searching capabilities within the CMCS infrastructure and scientific applications. The team members are Brett Didier (lead), Eric Stephan, Carmen Pancerella, and Karen Schuchardt with application contributions from Tom Allison, Dave Leahy, and Branko Ruscic.

Objectives of Search Task:

- Reduce time spent by scientists finding research results relevant to their research interests
- Provide search execution across multiple data sources (federated search)
- Provide integrated capabilities with browsing and analysis tools to quickly assess which data should be used
- Provide input to the chemical sciences community to improve the description of scientific data so data can be more easily discovered and accurately used by other researchers
- Enable pedigree searching of data resources to more accurately identify relevant data

The following search tasks are complete

- Requirements gathering with application scientists
- Technology investigation of DASL
- Prototype initial interfaces for query specification and results display
- Prototype standalone species dictionary application

These search tasks are in progress or planned:

- Document version 1 of search requirements

- Document search architecture
- Complete prototype search servlet (using propfind)
- Provide input to CMCS schema definition efforts
- Search Tasks
- Provide unified interface to species dictionaries
- Develop a species dictionary service (providing management capabilities)
- Develop a CMCS search service and portlet that utilize the species dictionary
- Integrate with portal pedigree browsing, notification, and data analysis capabilities

CMCS Architecture for SC2002 implementation

A CMCS Infrastructure Workshop was held at PNNL August 14-15, 2002. The focus was the development of an implementation of the CMCS architecture and software for SC2002. Topics covered at the workshop included;

- Review demo scripts and WBS.
- Review of overall system diagram of all the pieces.
- Review of Current Status and discuss the current architecture and needed changes.
- Current functionality
- Rough Design Diagram
- List of known design, technology, and other issues

Software for CMCS components were checked into CVS so the team could look at them as needed. These components included:

- Current state of SAM
- DAV Portlet Interface
- Slide/DASL/Search review
- Metadata/Search/Pedigree
- File Uploading
- Portal Object Discussion
- DSI
- Notification
- Access Control

An overview and general description of CHEF tools was presented with details and Q and A for CHEF folks (Terry W. from UMich). Server configurations at SNL, PNNL, ANL, NIST, SC2002 booth were documented and the required data translators and metadata extractor were specified. At comprehensive task list was generated for follow up to the workshop.