



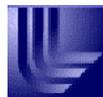
CMCS Search

Brett Didier

brett.didier@pnl.gov

Pacific Northwest National Laboratory

June 2002





CMCS Search

- Goal: 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.
- Team Members:
 - › Brett Didier (lead)
 - › Eric Stephan
 - › Carmen Pancerella
 - › Karen Schuchardt
- Application Contributors:
 - › Tom Allison
 - › Dave Leahy
 - › Branko Ruscic



Why is Search Important?

- 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

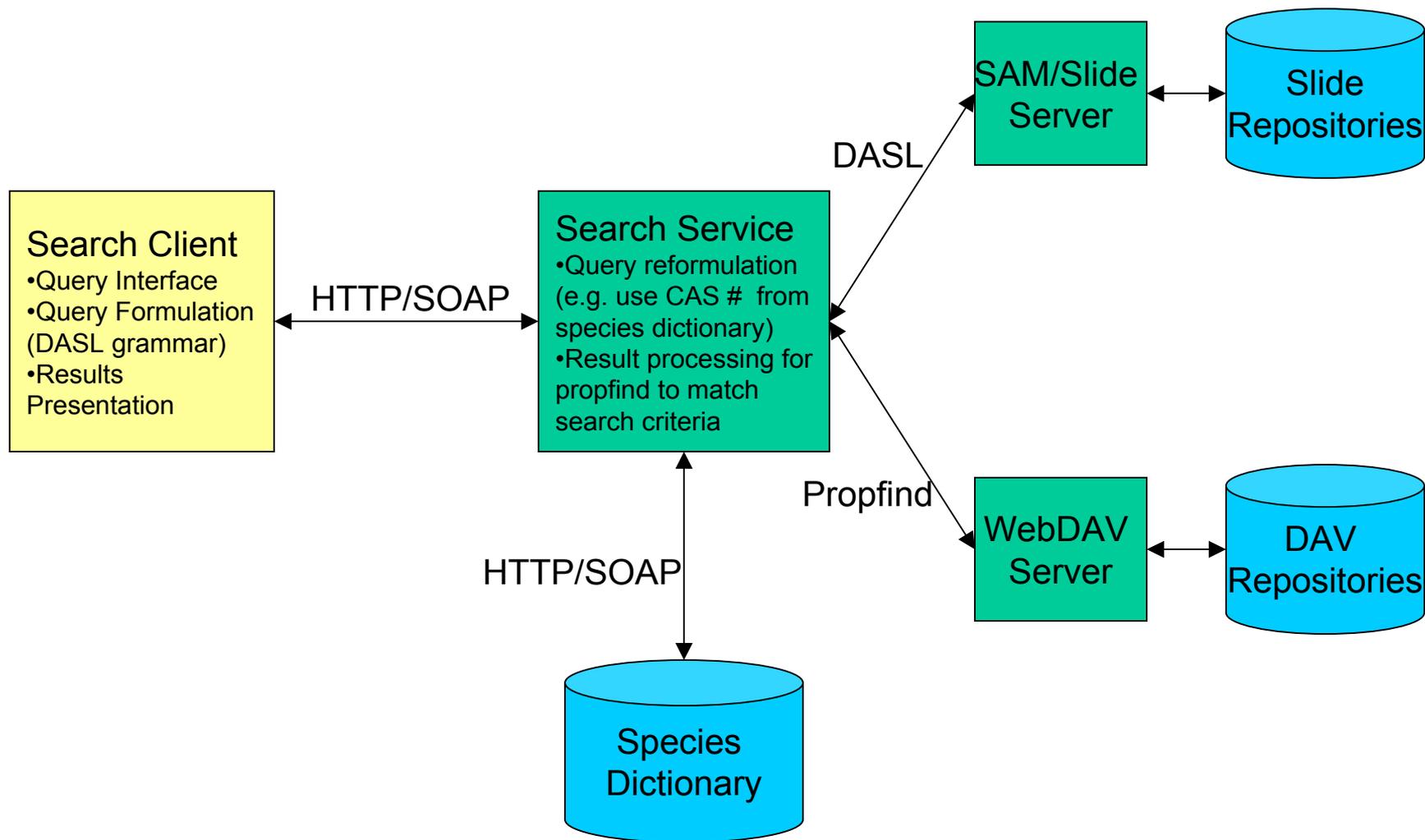


Requirements

- Provide search interface for “most-common” search criteria
 - › Chemical system name
 - › Chemical system formula
 - › Author
 - › Data type (e.g. Thermo, Kinetics, etc)
- Allow user entry of less common search criteria
- Support searching of resource metadata and resource content
- Provide Web-style presentation of search results
- Support a single schema (CMCS) for query construction and result display
- Allow searching of reference links found in metadata (pedigree searching)



Architecture





Technologies

- Search service
 - › HTTP based access
 - › Performs pre-search processing to narrow scope of search and any post-search processing
- Grammar
 - › DASL basic grammar
 - › XML-Query (if necessary)
- Species Dictionary (translate to “unique” species identifier, e.g. CAS #)
- SAM/Slide/DAV Servers and Data Repositories
- DASL/Propfind searches



Portal Searching Prototype



Jakarta Jetspeed Portal: Default Jetspeed page - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites History Mail Print Edit Discuss Real.com

Address http://bullitt.emsl.pnl.gov:8080/cmcs/portal/user/turbine/page/default.psm1/js_peid/P-edf05869a0-10038?action=controls.Close Go Links >>

Favorites

- Channels
- From ICQ
- Imported bookmarks
- Links
- Media
- Software Updates
- MSN
- My Documents
- Radio Station Guide
- RealPlayer Home Page
- Web Events
- Welcome to RealMedia!

Welcome **Tommy Turbine**
Customize: [HTML](#)
[WML](#)
[Edit accounts](#)
[turbine](#)
[Logout](#)

Collaboratory for Multi-Scale Chemical Science

[Home](#) [Karen](#)

- About CMCS
- Portal
- Project Team
- Publications
- Contacts

CMCS Search Portlet

CMCS Search v0.1

URL:

CAS #:

System Name:

Formula:

Author:

CMCS JMS Message Publisher

Enter the information in the fields below to publish a message.

Topic:

Content:

buildNormalContext: 0 ms

[Support and Additional Information](#)

Query Interface

URL - target for search

Initial search criteria:

- CAS #
- Chemical species name
- Chemical formula
- Author

Add query input fields for:

- Data type (e.g. thermo)
- Reactant or product
- Type of comparison (e.g. equality, contains)



Portal Searching Prototype



The screenshot shows a Microsoft Internet Explorer browser window displaying the 'Collaboratory for Multi-Scale Chemical Science' website. The address bar shows the URL: <http://bullitt.emsl.pnl.gov:8080/cmcs/portal/template/Home/template/Home>. The page features a navigation menu on the left with links for 'About CMCS', 'Portal', 'Project Team', 'Publications', and 'Contacts'. The main content area displays 'CMCS Search Portlet' and 'Simulated Search Results'. A table lists search results with columns for Title, Date, Size, and Author. Each row includes a title link, a date, a size, an author, and three small icons for document, notification, and pedigree browsing. At the bottom, there are 'Next', 'Prev', and 'Search' buttons, and a status bar indicating 'Search 1750 ms'.

Title	Date	Size	Author			
config.xml	today	1774807659	Ed Itor			
prop_hct_thermo.xml	today	672504547	Ed Itor			
prop_hct_ref.xml	today	1489521409	Ed Itor			
thermo	today	926055364	Ed Itor			
prop_hct_rates.xml	today	1278299951	Ed Itor			
hct_rates.zip	today	1843337675	Ed Itor			
hct2xml3.java	today	1889196109	Ed Itor			
prop.xml	today	4848516	Ed Itor			
hct_thermo.zip	today	187594680	Ed Itor			
references.xml	today	1217870495	Ed Itor			
rates	today	1980310354	Ed Itor			
out.cdat_194	today	1403150484	Ed Itor			
org	today	71864057	Ed Itor			

Search Results Display

Show title, date, size, and author of resources matching query criteria.

Resource can be retrieved by link on title.

Buttons allow viewing of properties, notification subscription and pedigree browsing.



Search Tasks

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