

# Scientific Annotation Middleware

Software infrastructure to support rich scientific records and the processes that produce them

Presented by

**Jens Schwidder**  
**Tara D. Gibson**

Computer Science Research Group  
Computer Science and Mathematics Division

**James D. Myers**

National Center for Supercomputing Applications

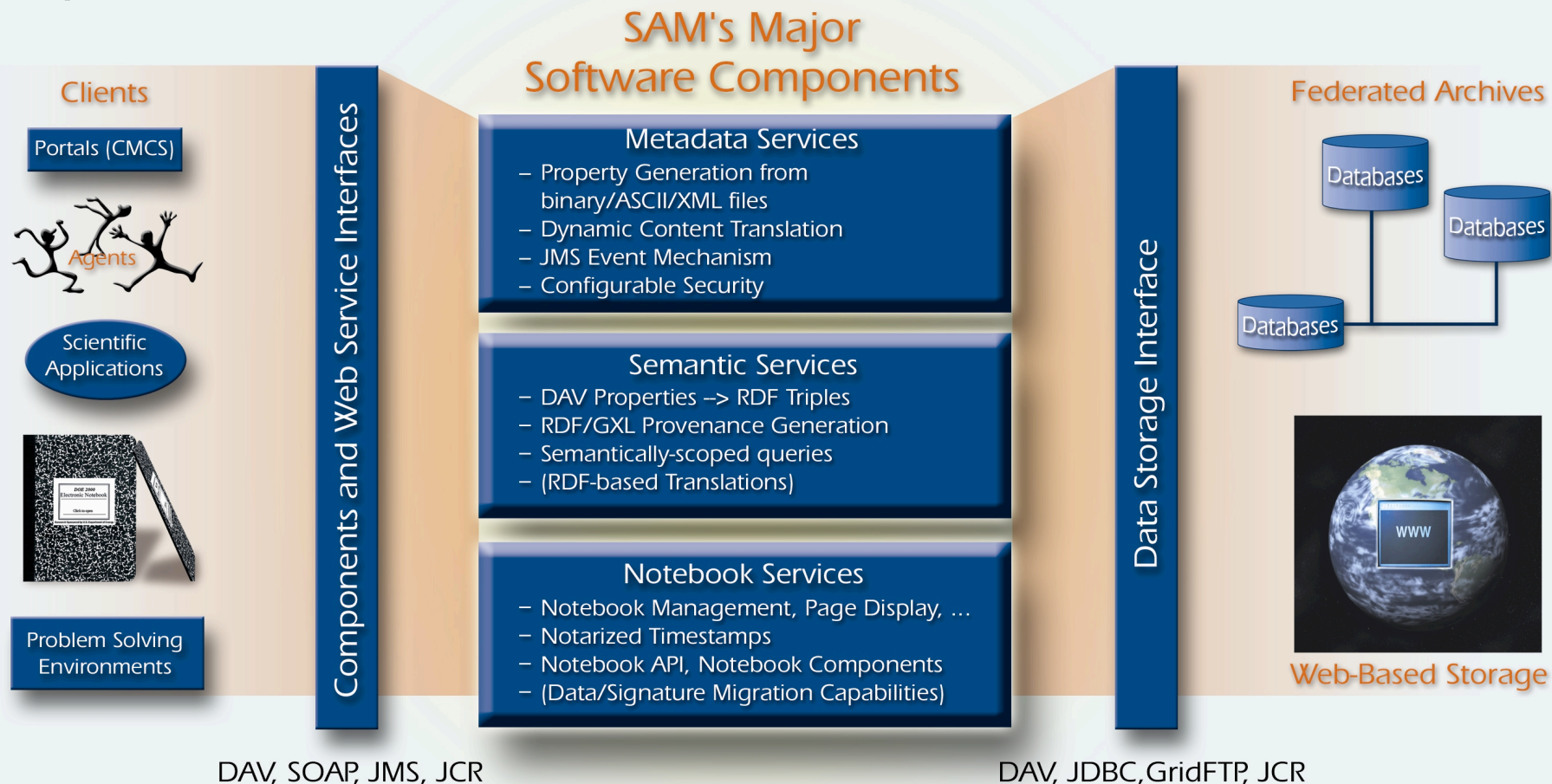


# Scientific Annotation Middleware (SAM) objectives

- Develop a lightweight, flexible middleware to support the creation and use of metadata and annotations
- Support sharing of annotations among scientific applications, portals, problem-solving environments, and electronic notebooks
- Improve the completeness, accuracy, and availability of the scientific record
- Support mapping between the annotation schemas of different scientific domains, thus enabling collaboration

# Middleware approach

- Various client and storage interfaces provide transparent integration of legacy applications as well as new applications using new, more powerful semantics



# Characteristics

- Middleware design capable of integrating into multiple service architectures
- “Schema-less” store that accepts arbitrary content and metadata
- Dynamic metadata/data translations to support evolving standards and lightweight integration
- Layered design to allow basic and advanced clients and interactions between them

# Features

- Meta–data translation/extraction
- Semantic services
- Distributed Authoring and Versioning (DAV)
- Notebook services and user interfaces
- Event notification using Java Messaging Service (JMS)
- Prototype implementation of Java Content Repository (JCR)(JSR 170)  
-based SAM layer that allows adding SAM capabilities to JCRs



# Benefits of the SAM system

- Rich, accessible, integrated scientific records
- Support for system-science cyber environments and collaboration across disciplines
- Increased automation of metadata capture and data/metadata translation
- Integrated electronic notebook, semantic relationship (e.g., provenance) tracking, and third-party annotation services
- Open source, standards-based scientific content management services
- Flexible authentication and authorization support

# SAM-based electronic notebooks

- Take advantage of advanced SAM features, such as data translation
- Provide hierarchical chapters/pages/notes
- Provide add/view/search notes
- Provide multiple client interfaces
  - Internationalized Electronic Laboratory Notebook (ELN) client
  - HTML-based Web interfaces
- Enable applications to provide notebook functionality using SAM notebook API/components
- Can serve as record with electronic signatures
- Allow scientists to share notes in distributed teams
- Allow email notifications

# Community interactions



Collaboratory for Multiscale Chemical Science (CMCS)—using SAM to support a portal-based community knowledge grid

SAM-based internationalized grid-capable notebook



Mid-America Earthquake Center MAEviz—“Consequence-based Risk Management Cyberenvironment”—using SAM to support shared data and provenance

Automated experiment records, user annotations, and customized instrument logs

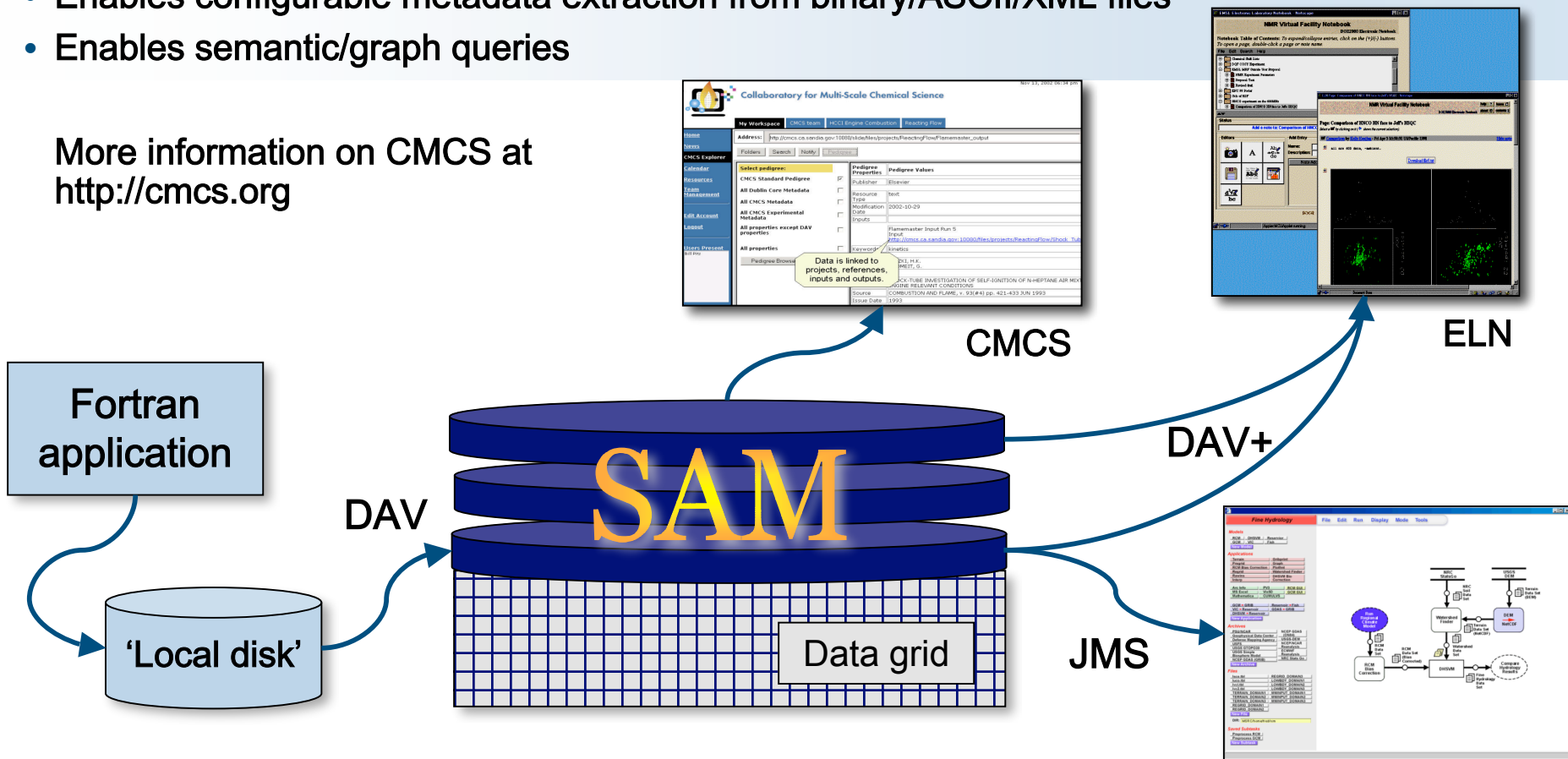


- Data Format Description Language (DFDL) standardization within the Global Grid Forum
- JCR (JSR 170) standardization within the Java Community Process
- Battelle records managers
- DOE2000 electronic notebook (Enote and ELN) communities
- PNNL Computational Science and Mathematics Division
- Semantic data grid to store, generate, and query provenance information

# CMCS use of SAM

- Powers CMCS knowledge management
- Provides a node plus metadata/relationship view of underlying data sources
- Supports put/get/search/access control of arbitrary data/metadata
- Enables configurable metadata extraction from binary/ASCII/XML files
- Enables semantic/graph queries

More information on CMCS at <http://cmcs.org>



# SAM 2.1.4 release

- DFDL, Web service, and XSLT-based metadata extraction and data translation capabilities
- Improved semantic search capabilities using an extension of DAV searching and location and Lucene indexing
- JDBC databases, file systems as data/metadata stores
- Simple Web-based SAM and notebook administration
- Internationalized ELN client (accepts UNICODE for Chinese/Japanese character sets)
- Optional fully Web-based version of the ELN client
- JAAS-based single-sign-on capabilities
- Notarization server and proxy implementation
- Command-line client and client API library
- Jakarta Slide 2.1 code base
- Requirements: Java 1.4 (or higher) and Tomcat 5.x



# Spallation Neutron Source (SNS) Notebooks

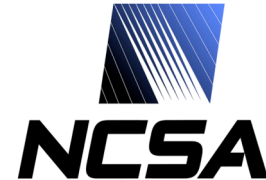
- The electronic notebook software for the SNS is being developed based on the research done in the SAM project.
- Support for different types of notebooks:
  - Instrument notebooks
    - Record events and annotations regarding an instrument.
    - Structure fixed; entries can't be edited, but can be annotated.
  - Proposal notebooks
    - Contain research annotations for a proposal and its experiments.
    - Structure and editing policies under control of proposal PI.
- Layered access control for SNS users and groups:
  - Personal notebooks.
  - Shared proposal and instrument notebooks.
- Web-based user interfaces using AJAX.
- Support for Wiki-formatting to support easy input of structured text.
- JCR-based storage system.

# More information about SAM

- Project information at <http://www.scidac.org/SAM>
- SAM source code hosted at <http://sourceforge.net/projects/sam>
- BSD/Apache-style open source license



Pacific Northwest  
National Laboratory



# Contact

## Jens Schwidder

Computer Science and Mathematics Division  
Computer Science Research Group  
(865) 576-7928  
[schwidderj@ornl.gov](mailto:schwidderj@ornl.gov)

