THE GRID ADVENTURES: SDSC'S STORAGE RESOURCE BROKER AND WEB SERVICES IN DIGITAL LIBRARY APPLICATIONS

Arcot Rajasekar, Reagan Moore, Bertram Ludäscher, Ilya Zaslavsky

San Diego Supercomputer Center
University of California San Diego

{sekar|moore|ludaesch|zaslavsk}@sdsc.edu

The data handling infrastructure being developed at the San Diego Supercomputer Center, includes a range of approaches and technologies for managing data, information and knowledge, specifically:

(1) self-instantiating and self-validating persistent archives;

(2) data handling system providing ubiquitous access to data resources stored in a variety of systems, epitomized in the development of SDSC Storage Resource Broker (SRB);

(3) data collection management services, with Extensible Metadata Catalog working in conjunction with SRB;

(4) digital library services for information navigation and discovery;

(5) information integration across multiple heterogeneous sources and web services, based on SDSC XML-based mediation technologies, and

(6) model-based knowledge management, providing conceptual-level mediation and ontology services.

This paper reviews the key components of the SDSC data handling infrastructure for scientific data, highlighting our experience with the SRB in a variety of application domains; XML-based information integration (mediation) technologies employing grid Web services, and knowledge management techniques.

1. What is Data Grid

The term "Data Grid" refers to a network of distributed data resources linked using a logical name space, to enable global uniform mechanisms for data access and query, collection-building and manipulation, data preservation and presentation. Within a grid, the distributed storage resources may include archival systems, caches, databases and their collections, along with software enabling information integration and computation across heterogeneous resources. Grid-based federation of data collections has been increasingly popular in various scientific disciplines, with several emerging grid prototypes which include:

- The Grid Physics Network [4],
- The Particle Physics Data Grid [17],
- Biomedical Informatics Research Network [3],
- The Knowledge Network for Biocomplexity [9],

and several others, with application areas ranging from physics and astronomy to ecology and geography.

While most of the data grid prototypes are still under construction, they exhibit a set of common features. The distinct "data grid" organization of data storage and management infrastructure is typically characterized by:

- Location transparency: users have the ability to seamlessly access data resources on any node on the grid, connecting from any other node;
- User transparency: users have a single point of authentication/authorization to access data holdings from distributed sites, based on user access rights, and authorize shared access to data holdings across sites, while maintaining strict levels of privacy and security; auditing mechanisms may be also available;
- Resolving heterogeneity of resources: ability to handle a variety of disparate data and computational resources: computer platforms, file systems, databases, collections, data types and formats (most recently – datasets with different semantics), as well as computational services;
- Persistent archiving: ability to coordinate lossless migration of data resources to newer computing platforms, storage systems, data formats;
- Data replication: ability to seamlessly create data replicas and maintain their consistency, to ensure quality of service, including fault tolerance, disaster recovery and load balancing;
- Collection management: ability to create, maintain and navigate a hierarchical virtual data organization, according to user-defined context-dependent structures;
- Inter-collection integration and scalability: ability to integrate large (in terms of number of files, number of datasets, and file sizes) data collections and associated metadata (system metadata, domain-specific metadata, and user-defined metadata);
- Application-level metadata handling: facilities to ingest, extract and maintain application-level metadata, and use it to support navigation and attribute-based query across datasets;
- Efficient data staging: ability to automatically cache frequently used data in distributed caches, or move data to archival systems, to provide optimal data access;
- Ontology management: ability to store and manipulate conceptual data models describing concept hierarchy, relationships and integrity constraints in each dataset, and reconcile domain-specific differences across datasets for navigation, query and collection integration;
- Multiplicity of access mechanisms: ability to be invoked from a variety of computing environments and portals, with multiple interfaces and language bindings, including web-based access using URLs and WSDL/SOAP-based services.

The data handling infrastructure being developed at the San Diego Supercomputer Center, includes a series of technologies and software that handle all aspects of data grids. The core of this infrastructure is the SDSC Storage Resource Broker (SRB), as well as specialized tools for managing persistent archives, inter-collection query mediation, and ontology services. This paper reviews the SDSC approach to data grids, focusing on the core software and recent experiences. The next section describes the SDSC Storage Resource Broker. It is followed by an overview of SDSC MIX (Mediation of Information using XML) technology, specifically focusing on ontology management extensions of MIX, for knowledge-based mediation across semantically different data sets within a data grid. In section four, we discuss SDSC's approach to persistent archives, as an application of data grid approaches integrating data, information and knowledge management aspects.

2. SDSC Storage Resource Broker

The SDSC Storage Resource Broker [2, 18, 19, 20] is a client-server middleware for organizing information from multiple heterogeneous systems into logical collections. It supports collection-building, preserving, managing, querying and accessing data in a distributed data grid environment. Working in unison with a Metadata Catalog (MCAT) [14], it provides a scalable facility for publishing and sharing scientific data and metadata, and information discovery in networks of heterogeneous data resources. The general organization of SRB, and common supported APIs, are shown in Figure 1. Following the main principles of data grid construction outlined above, SRB has the following key features:

- Support for a federation of SRB servers, where each server manages a set of storage resources, and can seamlessly access data from other servers if required by user requests (possibly, more than one SRB server can be configured to manage a particularly large resource with stringent fault-tolerance requirements);
- User-definable hierarchical organization of data in a logical name space, to express application- and context-dependent logical structure independent of physical data storage, and provide an intuitive navigation environment;
- Automatic maintenance of data consistency in data replicas created to ensure fault tolerance and disaster recovery;
- Single-point SignOn/authentication for user access to distributed resources, eliminating the need of user authentication on each individual resource;
- Single-point authorization, where a user can conveniently grant access to distributed data resources under her control, to collaborators, with no loss in privacy or security;

- Facility to aggregate data into physical blocks ("containers"), for optimal data placement, archiving, and caching in distributed caches;
- Support for large number (millions) of datasets, as well as large files (tens of GigaBytes in size), comprising collections measured in TeraBytes of storage;
- Brokering access to multiple resources including file systems (UNIX, NTFS, MacOSX), archival systems (HPSS, UniTree, ADSM, DMF), databases (Oracle, DB2, Sybase, SQLServer).
- Accessibility via command-line access, APIs for C, C++, Java, Python, VBasic, GUI access from Windows and UNIX, as well as web-based access using URLs and Web services.
- Cross-platform implementation: IBM AIX, Sun, SGI, Linux, Cray T3E and C90, Windows NT/2000/Me, Max OSX, etc.

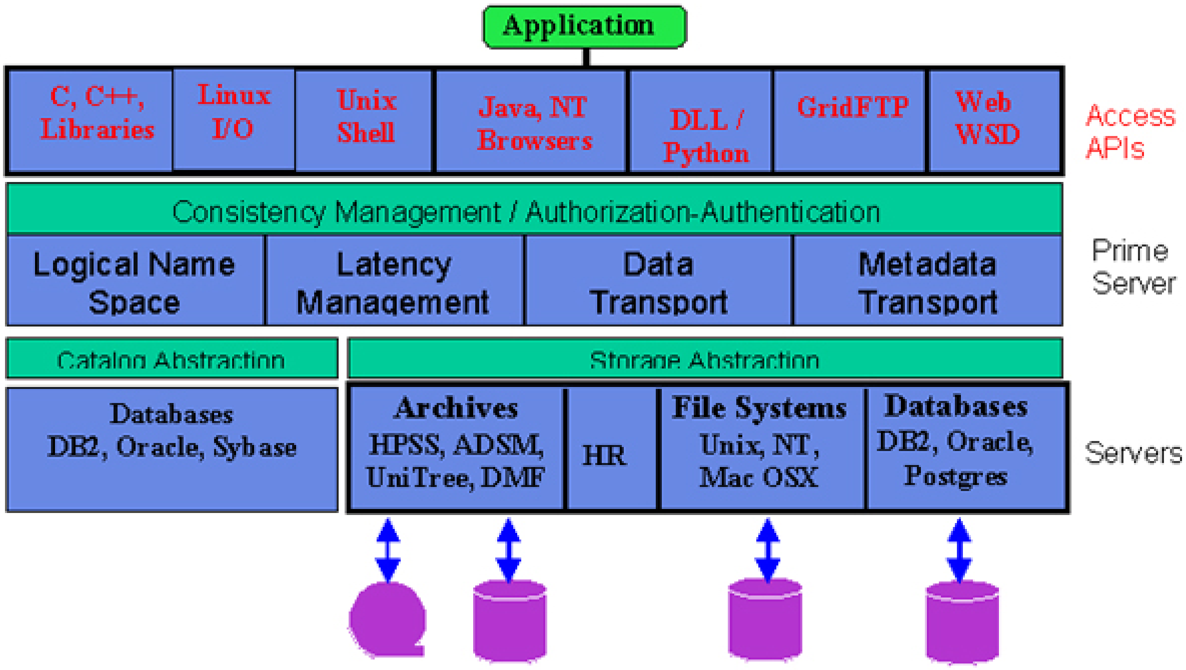

Figure 1. Common APIs of SDSC Storage Resource Broker and Metadata Catalog

Currently, SRB brokers more than 150 resources in a variety of projects, in disciplines ranging from physics and astronomy to biomedical informatics and ecology, providing access to more than 6 million files and over 38 TeraBytes of data. Here is an incomplete list of data grid projects where SRB is being used (see *http://www.npaci.edu/dice/srb/Projects/main.html* for more information about these and other projects):

1. Astronomy:
2MASS: the 2Micron All Sky Survey
Hayden Planetarium project
NVO: National Virtual Observatory
CACR (Center for Advanced Computing Research, CalTech) Computing Resource
DPOSS Collection (the Palomar Digital Sky Survey)
2. Earth-systems and Environmental Sciences:
ROADNet: Real-time Observatories, Applications, and Data management Network
HyperLTER (Long-Term Ecological Research) Project
LDAS: Land Data Assimilation System
CEED: Caveat Emptor Ecological Data Repository
Bionome: Biology Network of Modeling Efforts
3. Medical Sciences:
Visible Embryo Project
4. Molecular Sciences:
SSRL: Synchrotron Data Repository
AfCS: Alliance for Cellular Signaling
5. NeuroSciences:
NPACI Brain Data Archiving Project
BIRN: Biomedical Informatics Research Network
The Telescience Portal at NCMIR (National Center for Microscopy and Imaging Research)
6. Physics and Chemistry:
PPDG: Particle Physics Data Grid
GriPhyN (Grid Physics Network) Project
GAMESS (General Atomic Molecular Electronic Structure Systems) Portal
BaBar (B and B-bar experiment) Project
7. Digital Libraries and Archives:
NSDL: National Science Digital Library
National Archives and Records Administration (NARA) and Library of Congress
University of Michigan Digital Library Archive
CDL: California Digital Library
SIO Digital Libraries
ADEPT: Alexandria Digital Earth Prototype
8. Education:
Transana (Transcription and Analysis of Video Data)
Digital Insight

Projects that use the SRB-MCAT for its data grid capabilities include:

- NASA Information Power Grid
- NPACI Grid Portal Project
- UK eScience Grid at CLRS and UK Grid Starter Kit
- DOE ASCI Data Visualization Corridor
- DOE SciDAC - Portal Web Services
- Visible Embryo Project
  etc.

The 2MASS Digital Sky Project [1] is an example of an SRB-based data grid project in which the ingestion stage is complete. 2MASS stands for "2-Micron All Sky Survey", which represents a catalog of stars with images of stars taken with two highly automated 1.3 m telescopes, at the 2-micron level. The survey contained over 5 million images of point sources and other objects, initially stored off-line on tape in a raw format generated by the telescopes. The total size of the raw image files exceeded 10 TeraBytes of data. To provide astronomers with near-line access to the raw images of stars, the SRB group at SDSC developed procedures for migrating the data files from off-line tape storage to HPSS hierarchical storage system, under control of SRB, and making them Web-accessible. The ingestion phase, sketched in Figure 2, took 1.5 years. During this time data were streamed from CalTech (where the data files were

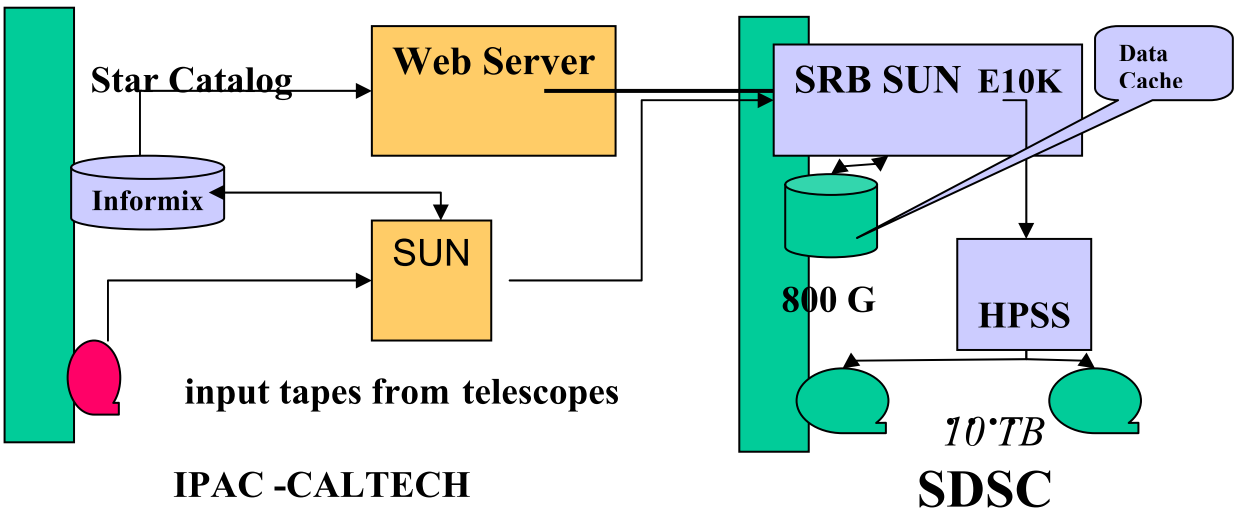

Figure 2. 2MASS Digital Sky Project. The project involved IPAC at Caltech and SDSC. Tapes were read at IPAC and data transferred over CalRen2 to SDSC where it was spatially sorted into containers in the data cache and finally archived into HPSS.

repeatedly read from tapes into staging space for streaming) to cache at SDSC, using the CalRen2 network, and subsequently moved to HPSS. Because of multiple space-related bottlenecks in the ingestion process at various times (such as lack of staging space at CalTech and cache space at SDSC for data streams of this size), the SRB-controlled ingestion procedures had to be fault-tolerant, with automatic restart and recovery for most types of errors. Also, raw images were stored on tapes in the order they were generated by the telescopes (i.e. in temporal order), while astronomers typically request images of a particular region of the sky (i.e. spatially). The temporal-to-spatial sorting involved lot of data movement from cache to archive at SDSC thus increasing ingestion time.

Another interesting 2MASS ingestion problem was related to handling of large number of relatively small (2 Mbytes) files by HPSS. HPSS does not manage small files efficiently, allocating excessive space to each ingested file; it would be overwhelmed by 5 million individual files. To resolve this problem, we used the SRB's "container" facility. A container is a physical aggregate of several files, which still maintain their individual metadata records in MCAT. Container-based manipulation, with automatic chaining of containers, are seamlessly handled by SRB during the ingestion process. The 2MASS data collection is currently accessed at a rate of over 1000 hits per day.

A new way to access SRB functionality is via MySRB [19], a Web interface for accessing SRB-brokered data and metadata. Enabling users to organize and share scientific data in a secure and platform-independent fashion, in a Web browser environment, is the specific goal of MySRB development. MySRB supports three primary functionalities: (1) collection and file management (including operations for data creation, ingestion, registration, replication, movement and deletion), (2) metadata handling (management of standardized (e.g. derived from the Dublin Core) and user-defined metadata), and (3) browsing files in the collection and metadata searching and querying, with subsequent display of files and metadata. Providing user-friendly access to core SRB functionality over the Web, MySRB completes the SRB-based data grid infrastructure, providing facilities for building distributed data collections, digital libraries, and persistent archives.

3. Mediation of Information using XML (MIX), and conceptual-level mediation with domain maps

Information mediation technology is another core element of the data handling infrastructure being developed at SDSC. To enable querying across multiple heterogeneous sources and web services in a data grid, we follow the information mediation approaches outlined in [22] and explored in such systems as TSIMMIS [16], DISCO [21], Information Manifold [10], etc. The Mediation of Information using XML project at SDSC explores mediation architecture where sources register as XML sources, and communicate by passing queries expressed in a declarative XML query language (initially: XML Matching and

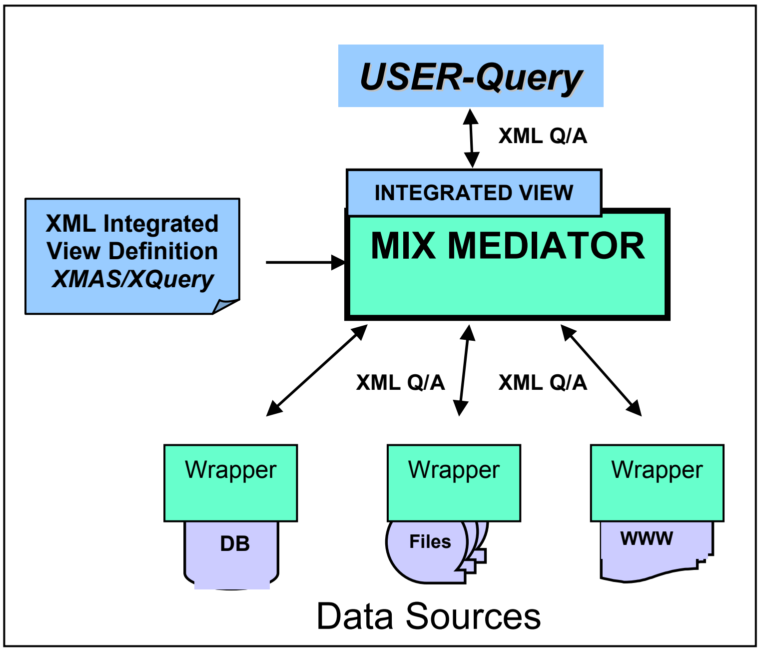

Figure 3. Standard wrapper-mediator architecture within SDSC MIX.

Structuring Language, or XMAS; today: XQuery) and XML-formatted results. Homogeneity in a mediator system is thus based on a set of source wrappers which convert XML queries against each legacy source into the language of the source, and transform query results into XML. Complex user queries are processed by mediator middleware, or a group of linked mediators, responsible for breaking up the initial user query into query fragments for each data source (based on source capability descriptions), orchestrating query execution, and merging results produced by individual sources, into composite presentations. The mediator provides virtual integrated views over distributed heterogeneous information sources, expressed in XMAS. The MIX system supports "lazy evaluation" of queries, i.e. on-demand query evaluation is driven by the client's navigation into the virtual XML view reflecting the query result [13]. The XML integrated views are not materialized; instead they are computed at runtime. The 3-level architecture of the MIX wrapper-mediator system is presented in Figure 3, various aspects of MIX and application experiences are described in [5, 6, 8, 24], including research focused on mediating across heterogeneous sources of spatial data.

In the recent experiments with MIX, we extended MIX architecture to handle spatial and survey information exposed as WSDL/SOAP web services. As an illustration, consider the following sample query:

```
"Find school districts in San Diego where com-
puter ownership rates among residents are over
80%"
```

In this query, “computer ownership” is not an attribute of school districts. Thus, the query has to be executed over two data sources with associated Web services:

- a survey data analysis server (in our example, it is the Sociology Workbench service at *http://www.edcenter.sdsu.edu/swb/*), which provides access and query mechanisms, as well as Web-based analytical interface, for the Digital Divide in San Diego survey [25], as well as a number of other surveys. Two questions asked in the course of that survey were "Do you have a computer or laptop in your house?", and "Where (what municipality) do you live in?" The survey variables can be queried with a set of servlets exposed via WSDL/SOAP.
- An Oracle Spatial source with boundaries of school districts and cities in San Diego county, and a set of servlets exposed via WSDL/SOAP and implementing a variety of requests against Oracle Spatial

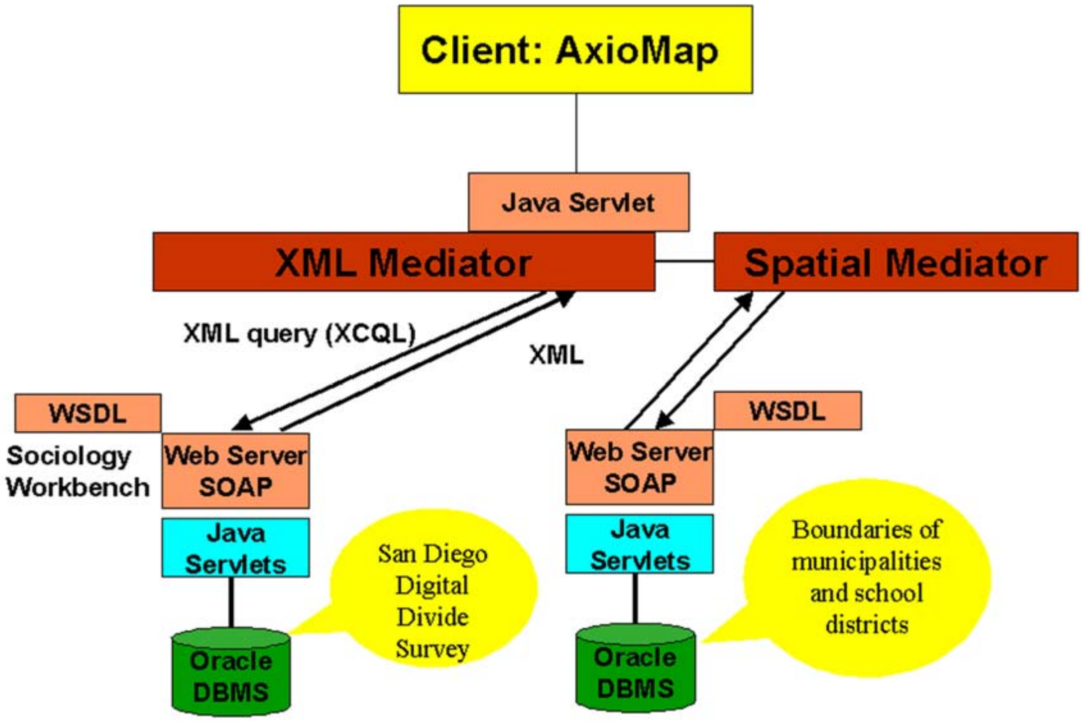

Figure 4. Mediator architecture for answering complex queries against heterogeneous sources of spatial and survey data.

The general framework for answering this query is shown in Figure 4. The mediator receives the query from a Web mapping client called AxioMap [23], and generates a query evaluation plan using capability descriptions of the two sources. In the first step, an XML query is sent to the Sociology Workbench server, requesting a cross-tabulation between “computer ownership” and “municipality” variables. An XML response from this query is then routed to the spatial mediator which issues a sequence of queries against the spatial database of municipal and school district boundaries stored in Oracle Spatial. In particular, Oracle Spatial initializes the appropriate tables, performs attribute selection on the “municipalities” table, followed by spatial selection of school districts that intersect with the selected municipalities. At the last step, the coordinates of the selected school districts are written out as an XML string conforming to AxioMap DTD, and sent to AxioMap for rendering.

While system-level and structural level integration issues are addressed by the SRB technology and the mediation approaches respectively, and syntactic heterogeneity of data formats is resolved by XML-wrapping of information sources and using XML for data interchange, these approaches are not sufficient when information integration spans several domains with different and often conflicting semantics. Data from multiple domains may be complex, have “hidden semantics” that cannot be explicated on the schema level, and often seem unrelated or indirectly related. Using such indirect semantic links for information integration requires formalization of domain expert’s knowledge in the form of conceptual models, and collections of rules and integrity constraints defined over such models. To address this problem, a novel approach called model-based mediation has been proposed [7, 11], where domain semantics is represented in processable form as domain maps (a kind of formal ontologies) and process maps. The architecture of a model-based mediation system is depicted in Figure 5. In addition to syntactic source wrappers of the traditional mediation which translate the raw data into a common XML data format, source wrappers in a model-based mediation system export conceptual models (CMs), which include classes, relationships, and integrity constraints. The mediator employs conceptual models and domain maps to define complex semantic relationships on the formalized domain knowledge, and use these relationships for evaluating user queries. Specifically, the mediator maintains a Generic Conceptual Model (GCM) which represents a meta-model for CMs and is used to link CMs exported by individual sources. Mediator also includes a rule processor to process relationships specified in a declarative rule language (F-logic).

This technology allowed researchers to approach information integration scenarios in such fields as neuroscience, ecology and geosciences, where integrating data with unspecified semantics and vaguely defined relationships is part of common research practice.

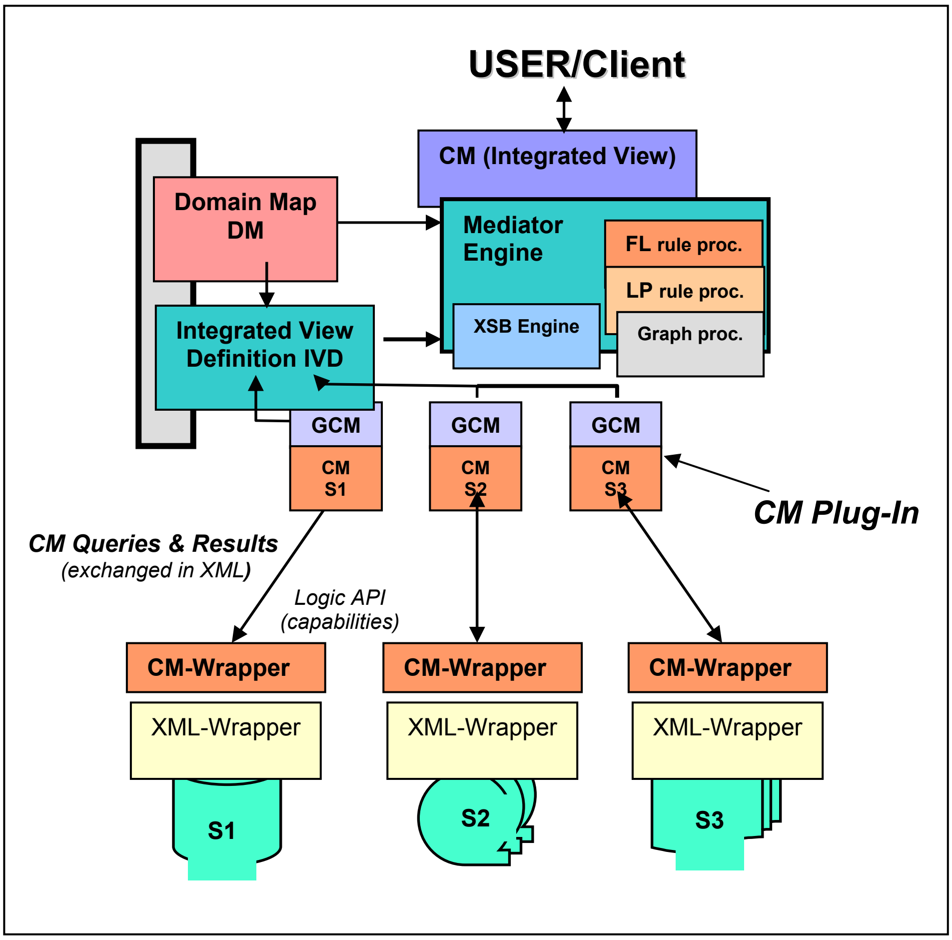

Figure 5. Model-based mediator architecture

4. SDSC Persistent Archives technology, an application of SDSC data grid

The main challenge in preserving digital information for long periods of time, is to maintain the ability to discover, access and display digital objects in the constantly evolving computing environment, across heterogeneous data storage systems, changing data formats and access mechanisms. Archiving digital objects "persistently" requires software infrastructure based on the integration of archival storage technology and information models, with domain-specific preservation models. This infrastructure challenge is similar to the common challenges of creating a data grid. The SDSC approach to persistent archives, therefore, focuses on the development of infrastructure independent representations for the information content of collections, interoperability mechanisms to support migration of the collection onto new software and hardware systems, and use of a standard tagging language to annotate the information content in an infrastructure-independent way. The persistent collection process is shown in Figure 6, as a sequence of steps necessary to preserve digital objects.

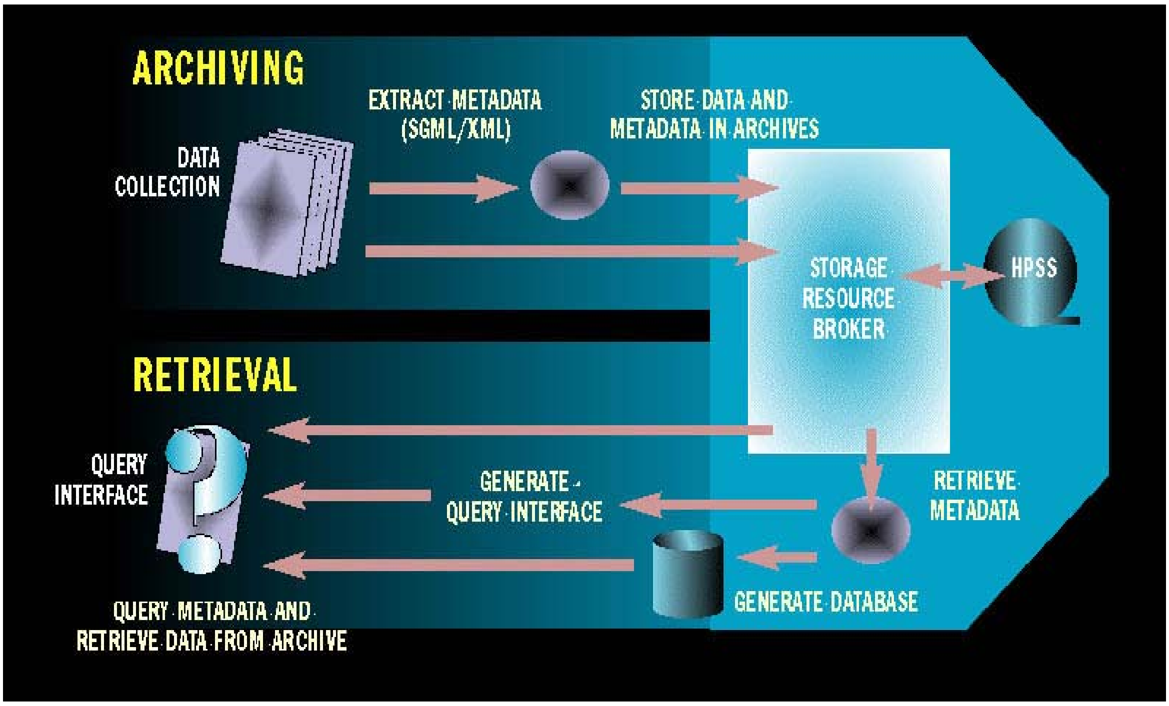

Figure 6. Archiving and retrieval of digital objects in a SDSC persistent archive

The concepts behind the SDSC persistent archive include:

- Infrastructure independence: all components of a persistent archive, including digital entity format, and organization of storage and information repository, should be accessed through abstraction mechanisms that support a common set of operations for the manipulation of data and information;
- Scalability: the archival processes associated with the preservation of digital entities can be automated through the use of data grid technology;
- "Migration" rather than "emulation" archiving strategy: rather than migrating display applications to new operating systems preserving the look and feel of the old technology, digital entities in a persistent archive are periodically migrated to new standard encoding formats, to enable more sophisticated navigation and query mechanisms, as well as annotation, analysis and consistency checking;
- Information and knowledge layers: every digital entity contains information (attributes used to assign semantics to the data) and knowledge (structural relationships and semantic integrity constraints described by a data model), so that a digital entity can be interpreted and displayed correctly;
- Abstracting digital entities and repositories: managing abstractions of digital entities and repositories (including information and knowledge repositories), rather then explicit physical representa-

tions of storage resources, simplifies technology management in a persistent archive;

- Self-instantiation: through archiving processes used to arrange, describe and preserve digital entities, and applying them to digital object at initialization, enable re-creation and re-validation of information content of a persistent archive.

As illustrated in Figure 7, knowledge-based persistent archives require software infrastructure to support interoperability between different implementations of ingestion, management, and access infrastructure components. Between the “Ingest platforms” column and the “Management” column, standards are needed to define consistent tagging mechanisms for knowledge (XML Topic Maps, RDFS, etc.), for information (XML DTDs and schemas), and for data organization (logical folders and physical containers). Between the “Management” column and the “Access services” column, standard query languages are needed for knowledge-based access, attribute-based access, and feature-based access. Between the “knowledge” and “information” environments, a standard representation is needed to map from concepts to attributes, such as topic maps or model-based access systems. Between “information” and “data storage” environments, a data handling system is needed to map from attributes to storage locations, such as the SDSC Storage Resource Broker described above.

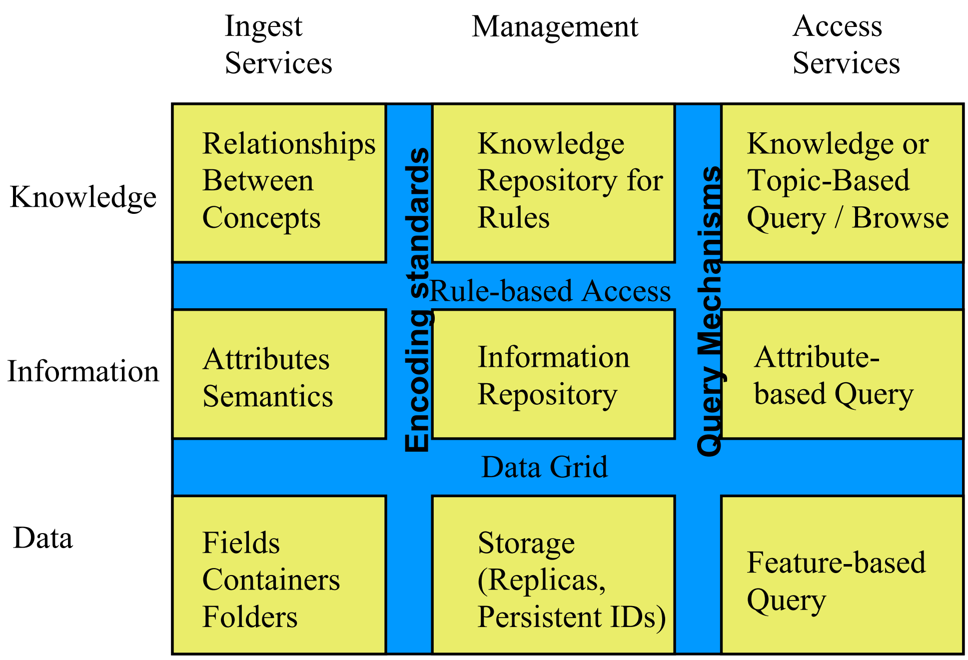

Figure 7. Knowledge-based persistent archive

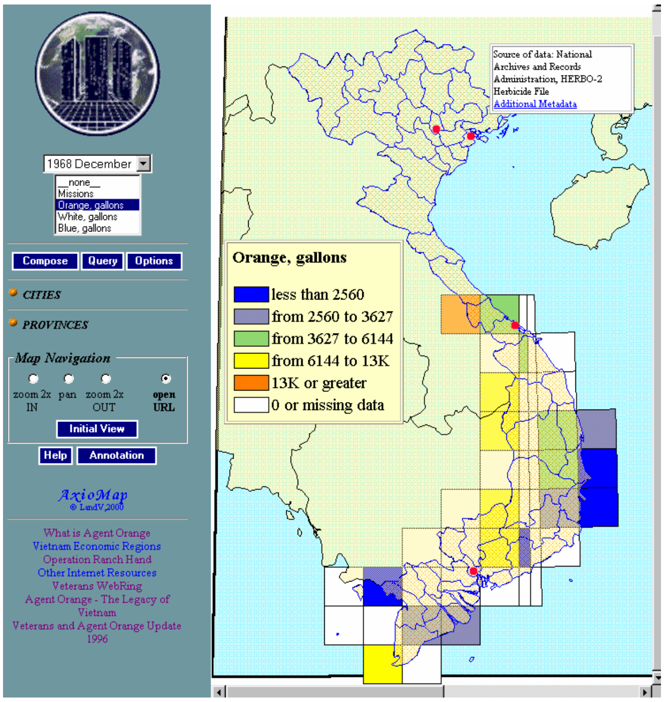

Figure 8. A snapshot of an XML map viewer used for archived NARA Herbicides collection. The viewer [22] is capable of displaying and navigating XML-tagged spatial data, using XML-based grammars for 2D vector rendering (VML: Vector Markup Language, and SVG: Scalable Vector Graphics).

Various elements of SDSC grid-based technology for persistent archiving have been demonstrated in several applications, including archiving NARA (National Archives and Records Administration) e-mail collection (1 million records; 2.5 Gb of data; 6 required, 13 optional and over 1000 user-defined fields – see [15]), the NARA HERBICIDES collection (the use of various defoliation agents: orange, white, blue, pink, purple, during the Vietnam war, in a program code named *Operation Ranch Hand* – see Fig. 8), NARA collection of U.S. Senate bills, amendments and orders [12], and others.

Conclusion

Data Grids are becoming increasingly important for sharing large scientific data collections, archiving, disseminating and querying them across system, structural and semantic boundaries, as solving difficult data interoperability problems moves to the forefront of research agenda in many research communities. This paper reviewed several core data grid technologies being developed at the San Diego Supercomputer Center. Modern grid software enables information integration at several levels: from system level where multiple data storage resources are presented in a common logical space and managed via persistent identifiers; to structural and syntactic level, where queries against heterogeneous information sources are mediated by enforcing XML encoding of data and system messages, and reconciling different data schemas with mediated views over XML sources; to semantic level, where domain knowledge is exported to mediators in the form of conceptual models, and navigated with domain maps. The software projects reviewed here, including the Storage Resource Broker (SRB) and SRB-based persistent archiving tools, Mediation of Information using XML (MIX), and model-based mediation, are at different development stages. While SRB is already deployed in a large number of research projects, the mediation software, especially for knowledge-based mediation, is still largely a series of research prototypes. However, moving the research prototypes to stable production systems is the task of not so distant future.

References

[1] 2MASS, 2001. http://www.ipac.caltech.edu/2mass/.

[2] Baru, C., R, Moore, A. Rajasekar, M. Wan, 1998. "The SDSC Storage Resource Broker," Proc. CASCON'98 Conference, Nov.30-Dec.3, 1998, Toronto, Canada.

[3] BIRN, 2001. "Biomedical Informatics Research Network" (http://www.nbirn.net)

[4] GriPhyN, 2000. "The Grid Physics Network", (http://www.griphyn.org/proj-desc1.0.html).

[5] A. Gupta, R. Marciano, I. Zaslavsky, C. Baru, 1999. "Integrating GIS and Imagery through XML-Based Information Mediation". In P. Agouris and A. Stefanidis (Eds.) *Integrated Spatial Databases: Digital Images and GIS*, Lecture Notes in Computer Science, Vol. 1737, pp. 211-234.

[6] A. Gupta, I. Zaslavsky, R. Marciano, 2000. "Generating Query Evaluation Plans within a Spatial Mediation Framework". *Proceedings of the 9th International Symposium on Spatial Data Handling*, Beijing, China, August 2000, pp. 8a18-8a31.

[7] A. Gupta, B. Ludäscher, M. E. Martone, 2001. "An Extensible Model-Based Mediator System with Domain Maps", *demonstration track, 17th Intl. Conference on Data Engineering (ICDE)*, Heidelberg, Germany, IEEE Computer Society, April 2001.

[8] A. Gupta, A. Memon, J. Tran, R. Bharadwaja, I. Zaslavsky, 2002 "Information Mediation Across Heterogeneous Government Spatial Data Sources". dg.o'2002 Proceedings, Los Angeles, May 2002.

[9] KNB, 1999. "The Knowledge Network for Biocomplexity", (http://knb.ecoinformatics.org/).

[10] A. Levy, A. Rajaraman and J. Ordille, 1996. "Querying Heterogeneous Information Sources Using Sources Descriptions". Proceedings of VLDB, 1996: 251-262.

[11] B. Ludäscher, A. Gupta, M. E. Martone, 2001, "Model-Based Mediation with Domain Maps", *17th Intl. Conference on Data Engineering (ICDE),* Heidelberg, Germany, IEEE Computer Society, April 2001.

[12] B. Ludäscher, R. Marciano, R. Moore, 2001. “Preservation of Digital Data with Self-Validating, Self-Instantiating Knowledge-Based Archives”, *ACM SIGMOD Record*, 30(3), 54-63.

[13] B. Ludäscher, Y. Papakonstantinou, P. Velikhov, Navigation-Driven Evaluation of Virtual Mediated Views, Intl. Conference on Extending Database Technology (EDBT), LNCS 1777, Springer, March 2000.

[14] MCAT, 2000. “MCAT: Metadata Catalog”, SDSC (http://www.npaci.edu/dice/srb/mcat.html).

[15] R. Moore, C. Baru, A. Rajasekar, B. Ludäscher, R. Marciano, M. Wan, W. Schroeder, A. Gupta, 2000. “Collection-Based Persistent Digital Archives”, D-Lib Magazine, 6(3 & 4), March [Part 1] and April [Part 2], 2000.

[16] Y. Papakonstantinou, S. Abiteboul, and H. Garcia-Molina, 1996. “Object Fusion in Mediator Systems”. In Intl. Conf. on Very Large Data Bases (VLDB), 1996.

[17] PPDG, (1999) “The Particle Physics Data Grid”, (http://www.ppdg.net/, http://www.cacr.caltech.edu/ppdg/).

[18] A. Rajasekar and M. Wan, 2002. “SRB & SRBRack - Components of a Virtual Data Grid Architecture”, Advanced Simulation Technologies Conference (ASTC02) San Diego, April 15-17, 2002.

[19] A. Rajasekar, M. Wan, and R. Moore, 2002. “MySRB & SRB - Components of a Data Grid,” The 11th International Symposium on High Performance Distributed Computing (HPDC-11) Edinburgh, Scotland, July 24-26, 2002.

[20] SRB, 2001. “Storage Resource Broker, Version 1.1.8”, SDSC (http://www.npaci.edu/dice/srb).

[21] A. Tomasic, L. Raschid, P. Valduriez, 1998. “Scaling Access to Heterogeneous Data Sources with DISCO” IEEE Transactions on Knowledge and Data Engineering, Vol. 10, No. 5, 1998: 808-823.

[22] G. Wiederhold, 1992. “Mediators in the Architecture of Future Information Systems”. IEEE Computer, 25(3):38-49.

[23] I. Zaslavsky, 2000. “A New Technology for Interactive Online Mapping with Vector Markup and XML”, *Cartographic Perspectives*, # 37, 65-77.

[24] I. Zaslavsky, A. Gupta, B. Ludäscher, S. Tambawala, 2001. “Query Evaluation and Presentation Planning within a Spatial Mediator: Extending XML-based Mediation to Heterogeneous Sources of GIS and Imagery Data.” DEXA Workshop 2001, pp. 853-855.

[25] I. Zaslavsky, P. Sensenig, D. Forkenbrock, 2001. “Mapping a Future for Digital Connections: A Study of the Digital Divide in San Diego”. San Diego Regional Technology Alliance. (http://www.sdrta.org/sdrta/aboutsdrta/RTA_Report_0201.pdf)