Collection-Based Persistent Archives

Reagan Moore
Chaitanya Baru
Bertram Ludascher
Richard Marciano
Arcot Rajasekar
Wayne Schroeder
Mike Wan

moore@sdsc.edu
http://www.npaci.edu/DICE
San Diego Supercomputer Center

Persistent Archives Challenges

- **Information input scaling**
  - Amount of information is growing exponentially
  - Implies the amount ingested in a year is comparable to the total archived information
- **Storage infrastructure scaling issues**
  - Implies archiving the new data can be the major task
  - Migrating the old data to the new technology may be a smaller task
- **Need scalable infrastructure**

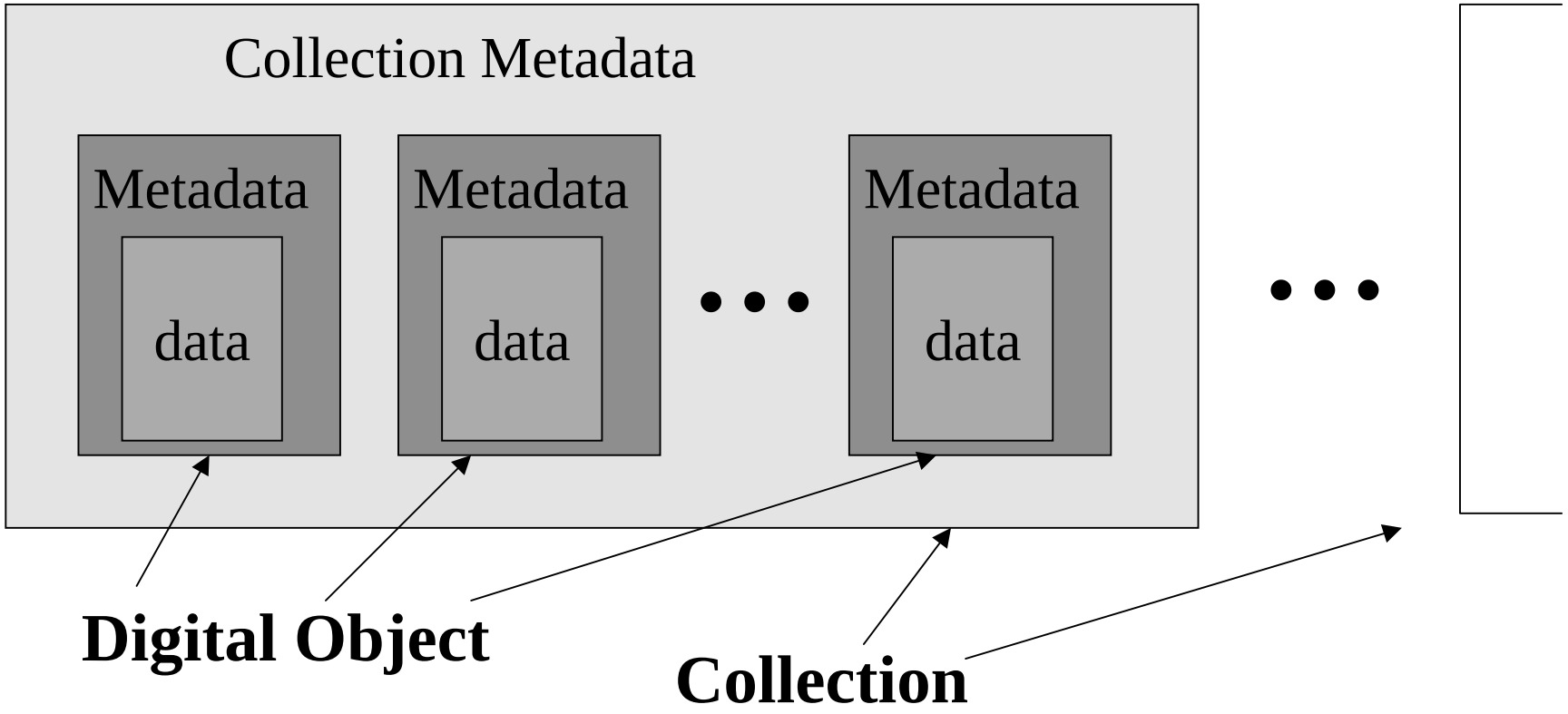
Persistent Archives Challenges

- **Maintain ability to retrieve information from a collection over 400 years**
  - The technology used to instantiate the collection changes every 3 years
  - The technology used for data presentation changes every 4 years
  - Technology used to archive the collection changes every 5 years
- **Need an Infrastructure Independent description of the collection**
  - Ability to rebuild the collection from the archived information

Approach - Context Based Objects

- **Data has value when it is given a context**
  - When archiving a digital object, must also archive the context of the object
  - Requires meta-data for defining the context
- **Use collections to define the context**
  - When archiving a collection, must also archive the information needed to reassemble the collection.
  - Requires meta-data to define the schema and the collection attributes

Collection Management



Persistent Archive Approach

- **Provide "infrastructure independent" definition of data collection**
  - Implies ability to dynamically reassemble data collection on new technology
- **Provide migration strategy for upgrading any infrastructure component of the persistent archive**
  - Implies ability to incorporate new media, new archival storage software, new collection management databases, new presentation displays

Persistent Archive Capabilities

Presentation	**Style Sheet**
Information Discovery	**Schema**
Database	**Metadata**
Archive	**Digital objects**
Infrastructure Dependent	Infrastructure Independent

Organizing Principles

- **Collection-based Approach**
  - Archive Collection instead of Data
  - Maintain **Context** associated with Data
  - Manage Information to capture relevance of digital objects
- **Use collection generation tools to recreate on new technology**
  - Collection database
  - Collection SQL for queries
  - Presentation interface

Migration of Information

- **Data Format** - **presentation, analysis**
- **Schema** - **ontology, semantics**
- **Meta-data** - **provenance, control, features**
- **Access** - **search, retrieval**
- **Database** - **query language**
- **Archive** - **internal system tables**
- **Media** - **bitfile preservation**

Information Management

Infrastructure Levels	Language	Data Flow Systems	Data Control
Format		Presentation	
Ontologies	Schema Definition		Schema Manipulation
Access		Discovery	
Metadata	Metadata Definition		Metadata Manipulation
Database		Handling	
Archive	Collection Layout		Storage Management
Media		Storage	

Persistent Archive at SDSC Infrastructure Migration

- **Content** - **80+ TB, 8 million files**
- **Data Format** - **SGML / XML**
- **Schema** - **MCAT / Dublin Core**
- **Access** - **FTP / SRB / Web**
- **Database** - **Illustra / DB2 / UDB**
- **Archive** - **Datatree / UniTree / HPSS**
- **Media** - **3480 / 3490 / 3590**

Persistent Archive Prototypes

- **NARA**
  - Scalable persistent archive
    - Digital object definition
    - Collection instantiation
- **NPACI**
  - Federation of collections
    - Extensible schema
- **AMICO**
  - Federation of schema
    - AMICO, MARC, CDL Formats
- **Patent Digital Library**
  - Migration of digital object structure language
    - Greenbook to SGML to XML

NARA Prototype

- **Used commercially available software**

Presentation	**ASCII, XML style sheets, Raster images**
Discovery	**SQL**
Collection	**Object-relational database (Oracle on Sun Enterprise server)**
Archive	**High Performance Storage System (HPSS on IBM SP2)**

Data Collections

- **Newsgroups - 1 million records**
- **TIGER/Line 1992 (Bureau of the Census)**
- **104th Congress**
- **Electronic Access Project (EAP)**
- **Vote Archive Demo 1997 (VAD)**
- **Combat Area Casualties Current File (CACCF)**
- **Patent Data (USPTO) - 2 million patents**
- **Image collection (AMICO) - 80 GB**

Collection-based Management

- **Identification of Attributes**
  - Object-level Attributes
    - parameters, schema, presentation, etc
  - Collection-level Attributes
    - access, curation, index, domain metadata, etc
- **Packaging of Data and Metadata**
  - Capture Locality
  - Ease of Rebuild

Infrastructure

- **Archives**
  - Transparency of storage
- **Data-handling System**
  - Data Management
  - Access for Interactive and Batch Process
- **Database and Catalog System**
  - Metadata Repository

Infrastructure (contd.)

- **Digital Libraries**
  - Manipulation
  - Presentation
- **Collection Instantiation System**
  - Systems
  - Methods
  - Data and Meta-data

Describing Information Content

Information Level	Infrastructure - Scientific Data	Infrastructure - Text
Federation	Ontology	Digital Library
Data Collection	Schema	Dublin Core
Data Set	Metadata	Provenance
Features	XML	XML
Logical type	Vector bundle	Mime Type
Structure	MPI Datatype	DTD
File Format	HDF v5	Electronic record

Metadata Catalog

- **Digital Object Metadata**
  - types, formats, lineage
  - ingestion protocols and usage methods
  - domain-specific metadata, related data
  - information discovery
  - presentation information

Metadata Catalog (contd.)

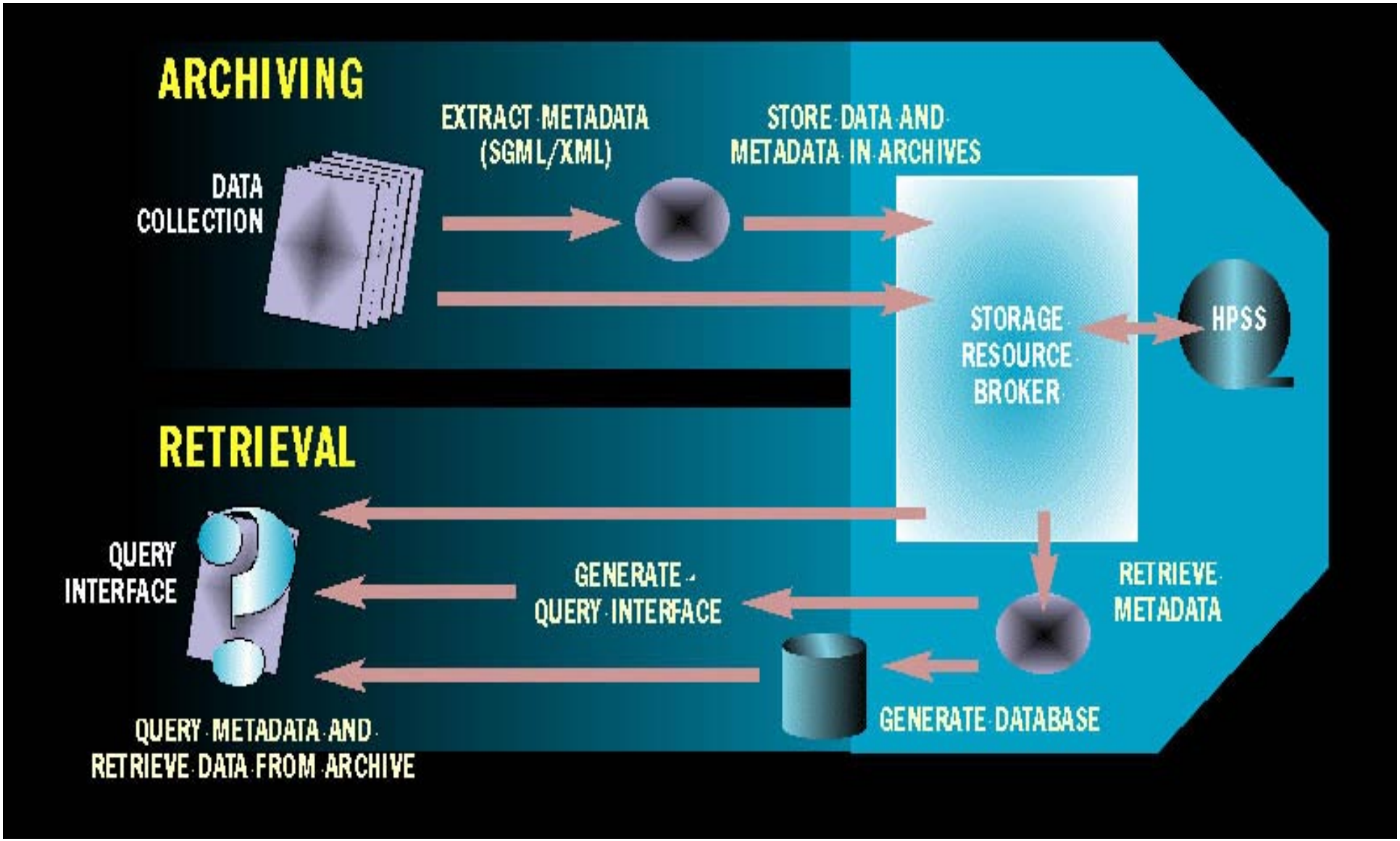
- **System-level Metadata**
  - Storage System Characteristics
  - Access and Audit Control
  - Replication, Partition, Containers
  - Location Transparencies
  - Authentication and Encryption

Metadata Catalog (contd.)

- **Schema-level Metadata**
  - Metadata that defines schema attributes
  - Ontology, Indexing, Relationships between attributes
  - Evolution Mechanisms
    - data
    - schema
  - Federation and Integration Information

NARA Prototype

- **Demonstrate ability to ingest, archive, recreate, query, and present a digital object from a 1 million record Usenet collection (RFC1036)**
  - 2.5 GB of data
  - 6 required fields
  - 13 optional fields
  - User defined fields (over 1000)
- **Determine resources required to scale size of collection**


ARCHIVING
EXTRACT METADATA (SGML/XML)
STORE DATA AND METADATA IN ARCHIVES
DATA COLLECTION
STORAGE RESOURCE BROKER
HPSS
RETRIEVAL
QUERY INTERFACE
GENERATE QUERY INTERFACE
RETRIEVE METADATA
GENERATE DATABASE
QUERY METADATA AND RETRIEVE DATA FROM ARCHIVE

Ingestion Procedure

- **Digital object creation**
  - Define meta-data
  - Define object structure (XML DTD)
- **Collection definition**
  - Define collection-level meta-data
  - Define collection structure / schema (XML DTD)
  - Define mapping from Object DTD to Collection DTD
- **Create containers**
  - Digital objects, encapsulate with meta-data
  - Collection, aggregate digital objects for storage

Instantiation Procedure

- **Dynamically generate collection creation scripts**
  - Use Object DTD, Collection DTD, Mapping between DTDs and target system attributes to create DDL for database table creation
- **Dynamically load collection into database**
  - Use Object DTD, Mapping between DTDs, DDL, and target system attributes to create load scripts
- **Dynamically create SQL for querying database**
  - Use Collection DTD, DDL, and target system attributes to generate SQL

Ingestion Demonstration

- **Resources**
  - SGI Indigo, Ethernet, Oracle on Sun Enterprise, HPSS on SP2 to support a 1 million object collection
- **Usenet digital object creation - meta-data mining**
  - 1 hr 39 min
- **Archival storage of digital objects over Ethernet**
  - 1 hr 29 min
- **Archival storage of original data**
  - 1 hr 2 min

Instantiation Demonstration

- **Retrieval of objects and generation of load scripts**
  - 2 hr 40 min
- **Database creation**
  - 4 hr
- **Database indexing**
  - 4 hr
- **Query**
  - 1 sec

Migration to New Technology

- **If database infrastructure changes, create new versions of**
  - Generator program
  - Loader program
- **If DTD for defining structure of objects or collection changes, create interface program for mapping old DTD structure tags into new DTD structure tags**
  - Dynamically apply interface program as collection is read from archive

Digital Object Definition - Lessons Learned

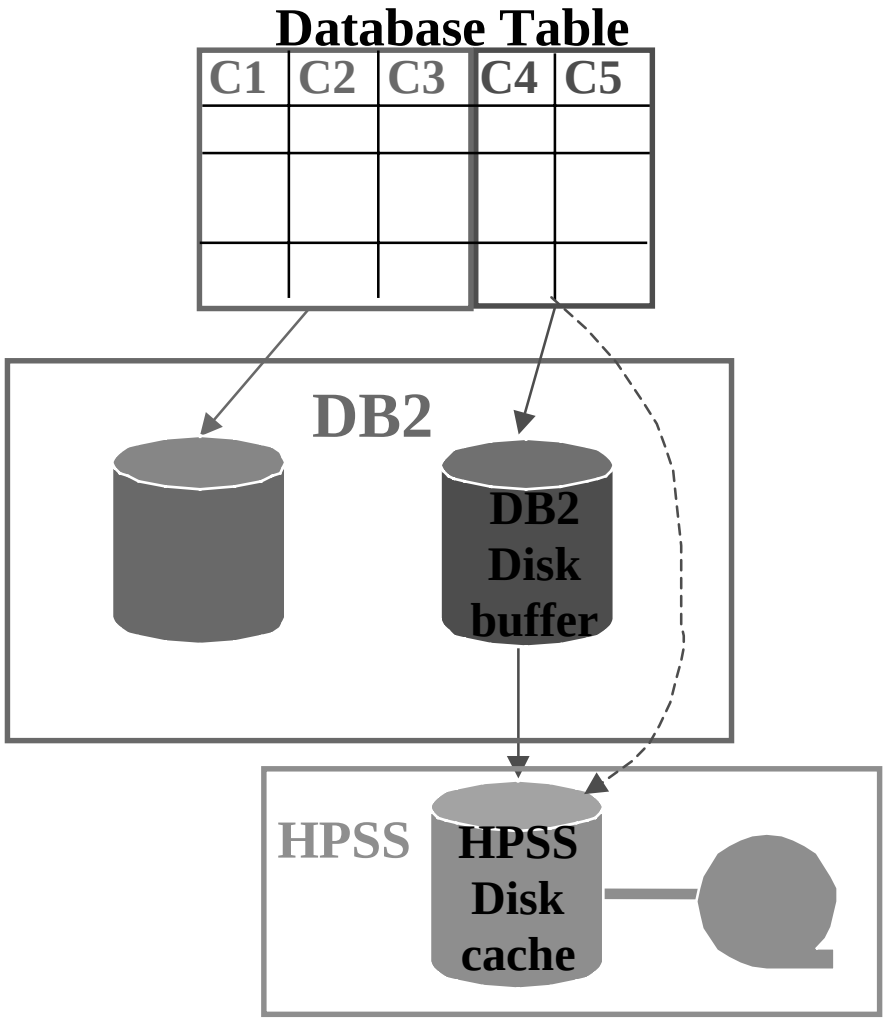
- **Unique encapsulation of digital object**
  - Beware of objects embedded within objects
- **Duplication of digital objects**
  - Usenet messages posted to multiple bulletin boards
- **Replication of meta-data within an object**
  - (objects embedded within objects)
- **Use of unexpected character sets**
  - Can cause bad meta-data
- **Quality assurance is necessary**
  - Automated meta-data validation

Impact on Persistent Archive

- **Collection-centric Archives**
  - Collection Meta-data Management
  - Storage of Context
- **Container Technology**
  - Aggregation of Digital Objects within Collections
- **Mechanism for Migration of Archive Meta-data**
  - Can apply same technology used to migrate collection meta-data

DB2/HPSS Integration

- **Collaboration with IBM TJ Watson Research Center**
  - Ming-Ling Lo, Sriram Padmanabhan, Vibby Gottemukkala
- **Features:**
  - Prototype, works with DB2 UDB (Version 5)
  - DB2 is able to use a HPSS file as a tablespace container
  - DB2 handles DCE authentication to HPSS
  - *Regular* as well as *long (LOB)* data can be stored in HPSS
  - Optional disk buffer between DB2 and HPSS



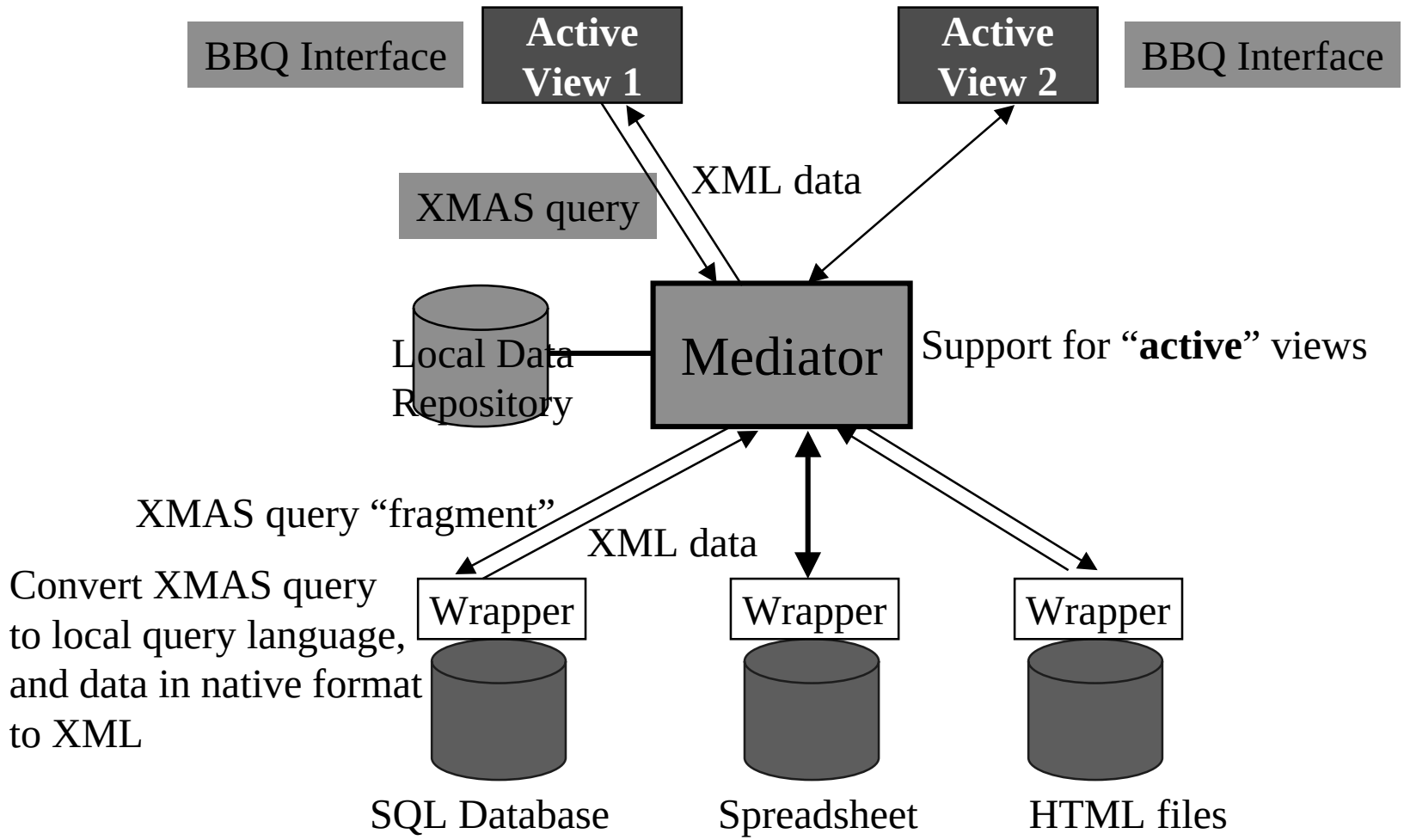
Future Development

- **Generalization of programs to generate and load new database to work with arbitrary DTDs**
  - Goal is to use one set of programs for all data collections
- **Generalization of presentation technology**
  - Use of XML style sheets to define presentation format
  - Demonstration of presentation of digital object as archived
    - Multiple presentation styles for digital objects
  - Support for semi-structured data
    - Multiple views of collection
    - Annotations applied to data collection or digital objects within collection - MIX
    - Browsing mechanism for collection - BBQ

Generalizing Persistent Archives

- **MIX - Presentation transparency**
  - Mediation of information with XML
  - Support for semi-structured data
- **Scalable archival storage systems**
  - Cost/performance
- **SRB - Location transparency**
  - Access to heterogeneous systems
  - Access to remote systems
- **MCAT - Collection transparency**
  - Dynamic recreation of collection
  - Extensible Schema support

MIX: Mediation of Information using XML



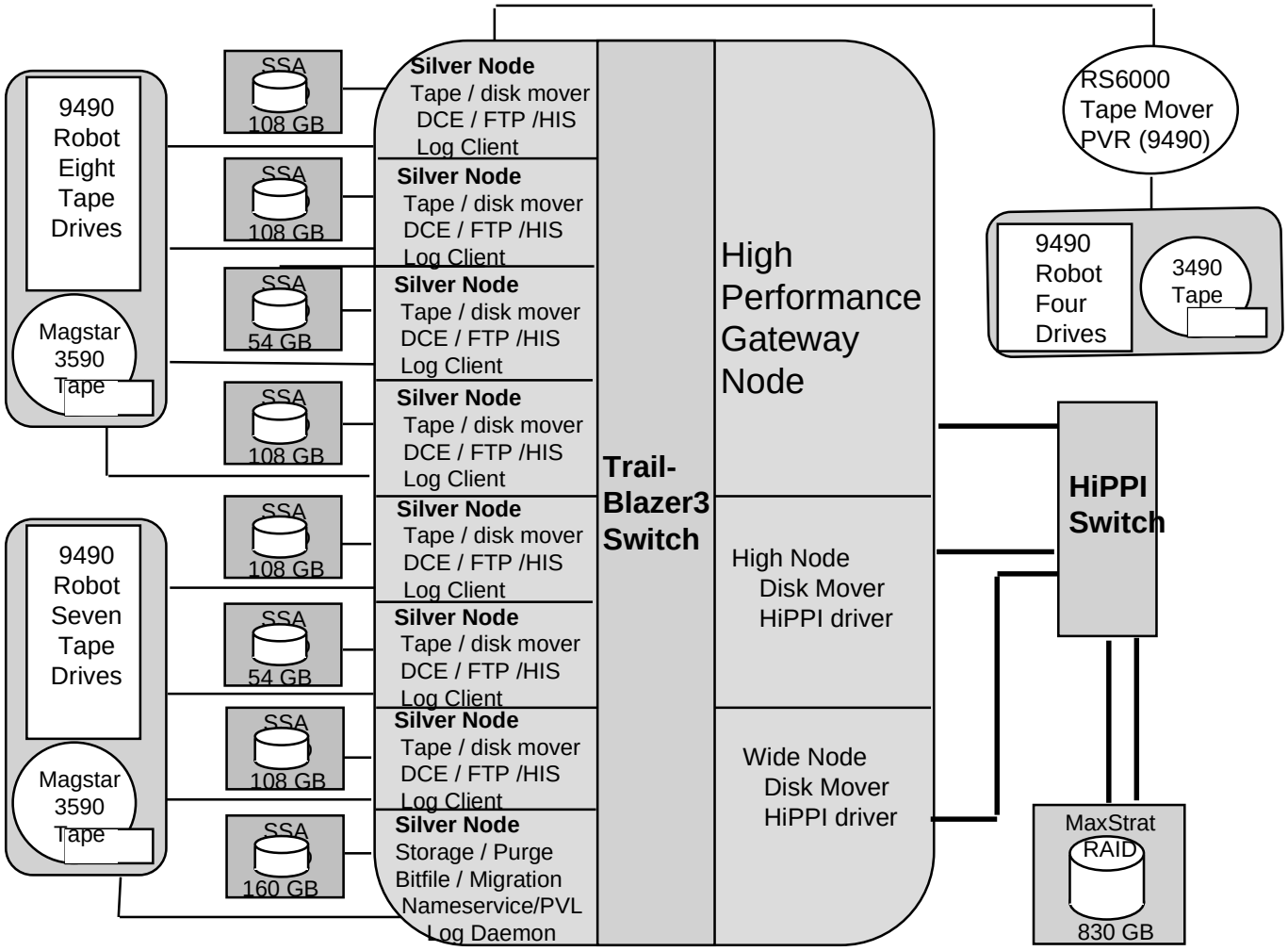
High Performance Storage

- **Provide access to tertiary storage - scale size of repository**
  - Disk caches
  - Tape robots
  - Manage migration of data between disk and tape
- **High Performance Storage System - IBM**
  - Provides service classes
  - Support for parallel I/O
  - Support for terabyte sized data sets
  - Provide recoverable name space

State-of-the-art Storage: HPSS

- **Store Teraflops computer output**
  - Growth - 200 TB data per year
  - Data access rate - 7 TB/day = 80 MB/sec
  - 2-week data cache - 10 TB
  - Scalable control platform
    - 8-node SP (32 processors)
- **Support digital libraries**
  - Support for millions of data sets
  - Integration with database meta-data catalog

HPSS Archival Storage System



HPSS Bandwidths

- **SDSC has achieved:**

Node-HPGN	90 MB/s
Texas Memory Box	80 MB/s
Max Strat disk	60 MB/s
SSA Raid	20-30 MB/s

- **Striping required to achieve desired I/O rates**

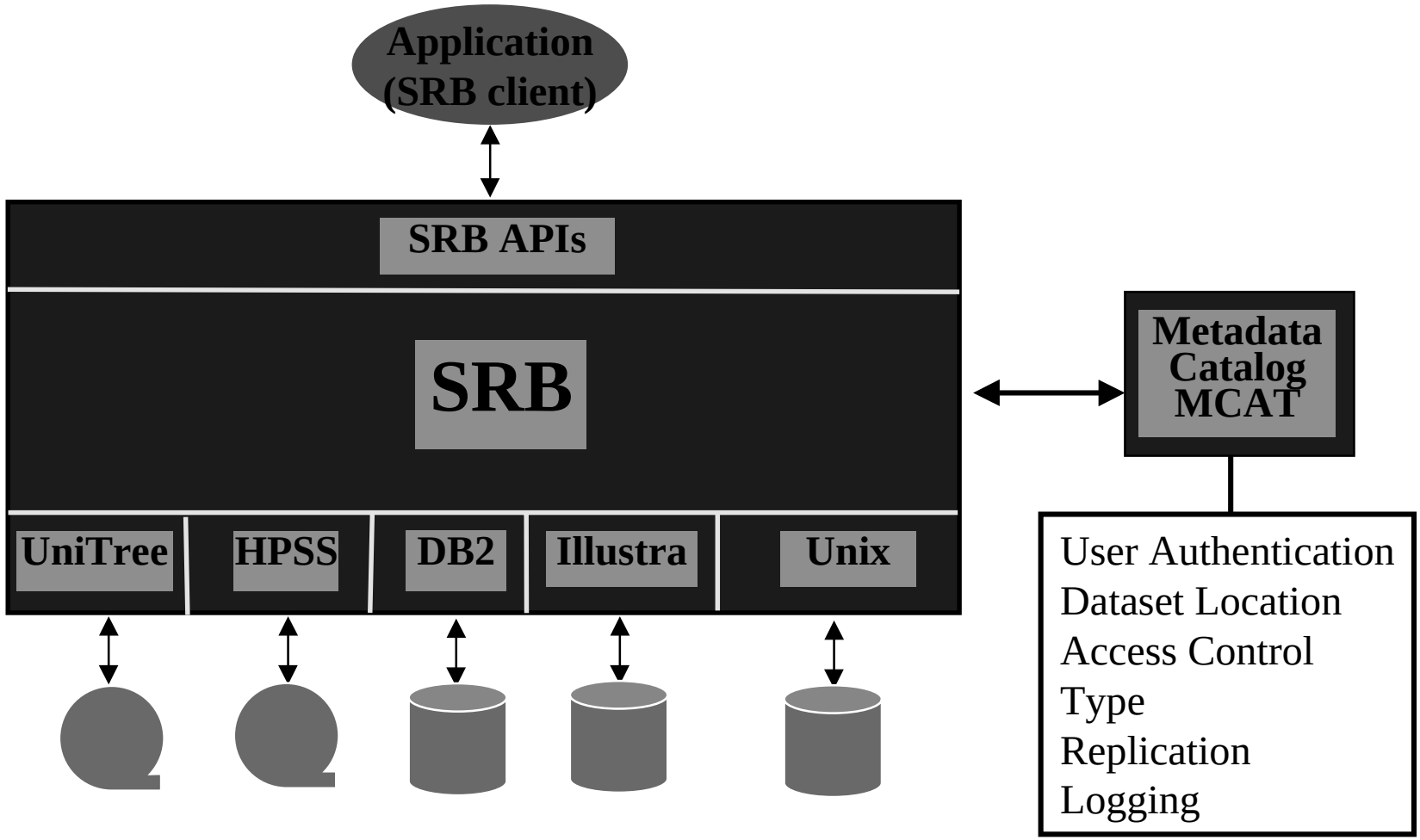
SRB/MCAT

- **System developed at SDSC**
- **Resource Transparency provided by the Storage Resource Broker**
  - access, location, system transparency
- **Information Repository provided by the MetaCatalog**
  - collection-level metadata

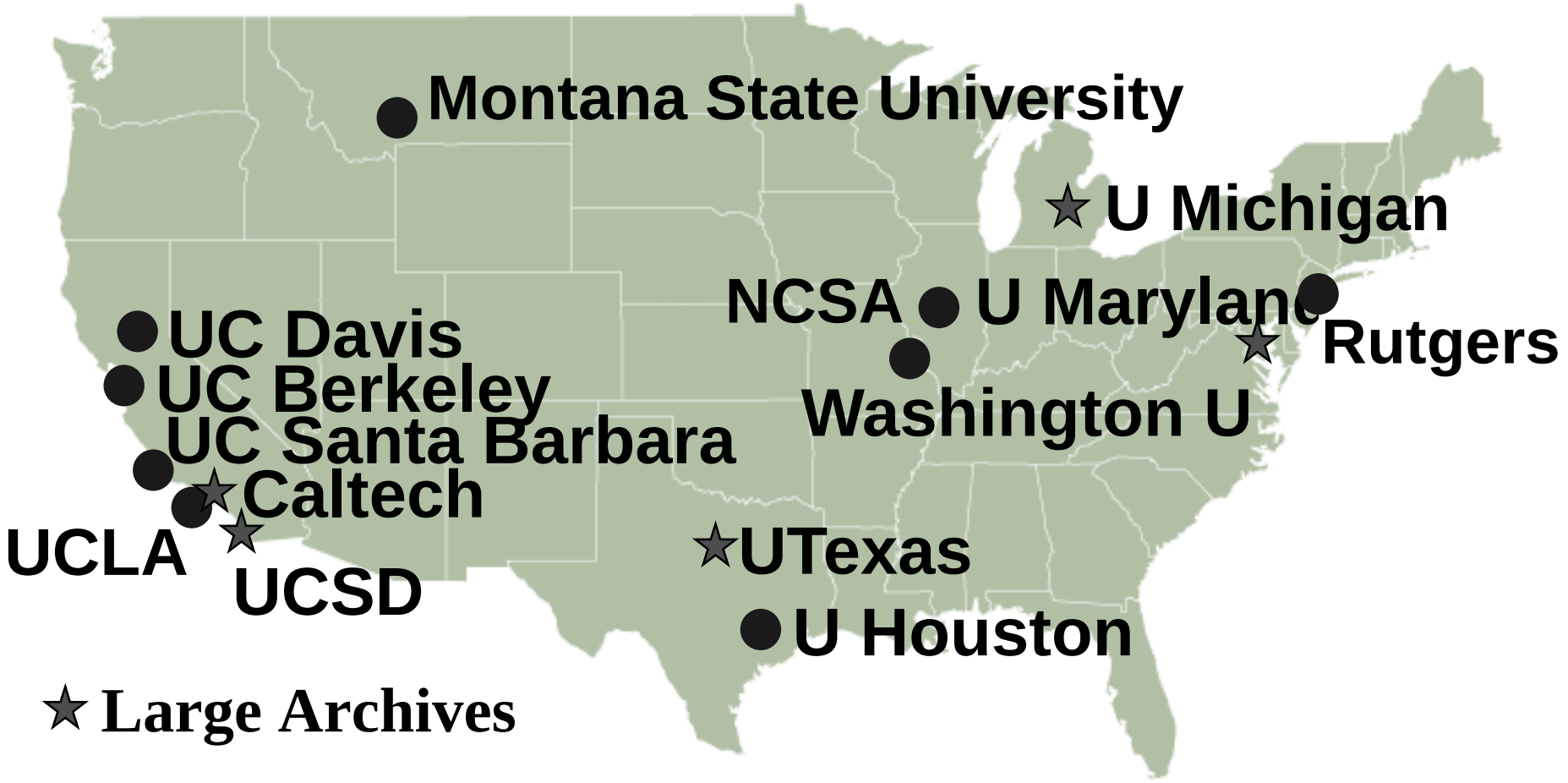
MCAT Architecture

- **Management of Meta-data Catalog**
- **Several Types of Collection-meta-data**
  - logical structures
  - attribute clustering information
  - token attributes
  - linkages
- **Automatic Query and Update across Federated Catalogs**

SRB Software Architecture



14 Installed SRB Sites



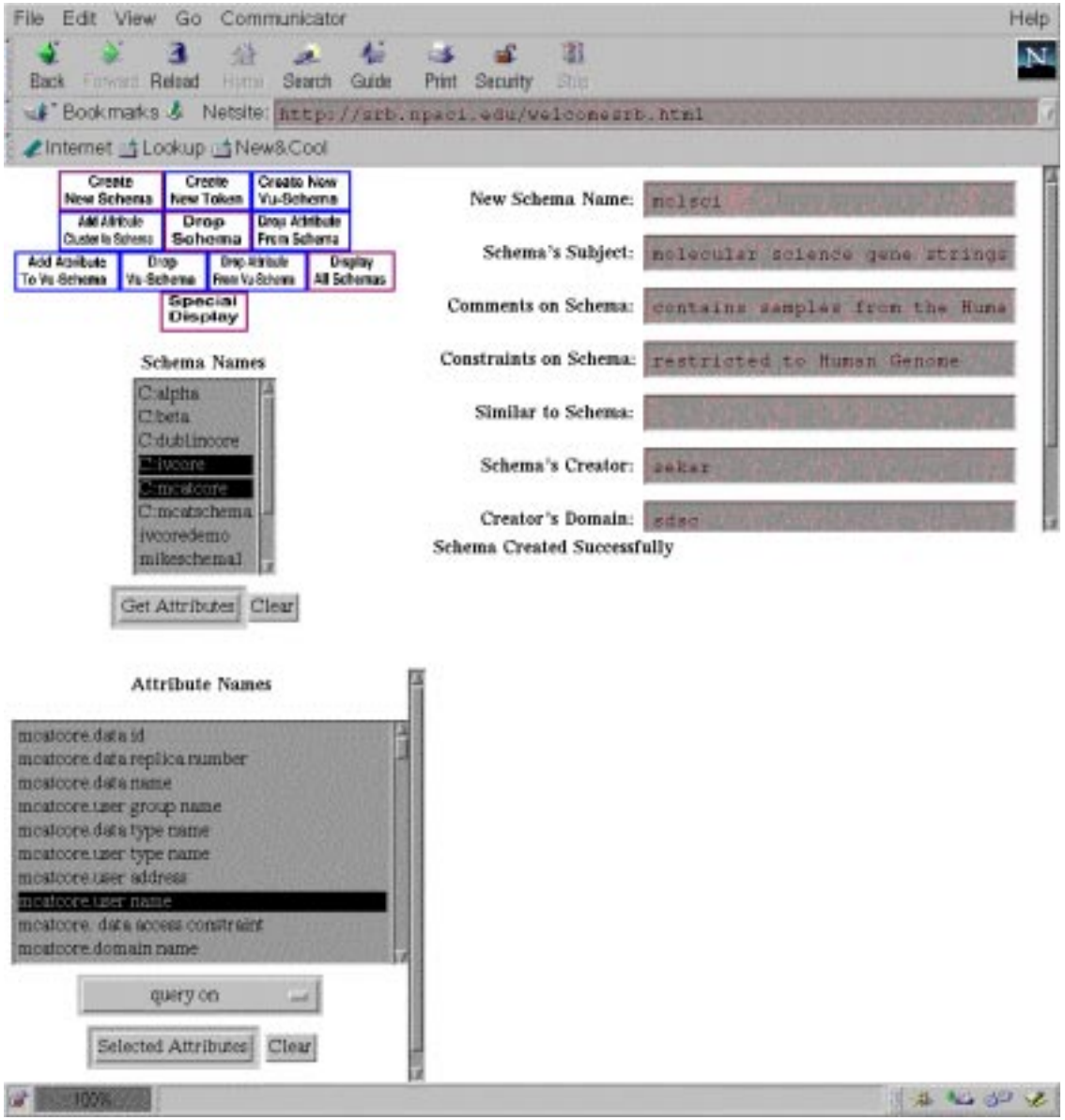
SRB / MCAT Features

- **Support for *Collection* hierarchy**
  - allows grouping of hetero-geneous data sets into a single logical collection
  - hierarchical access control, with ticket mechanism
- **Replication**
  - optional replication at the time of creation
  - can choose replica on *read*
- **Proxy operations**
  - supports *proxy (remote)* ***move*** and ***copy*** operations
- **Monitoring capability**
- **Supports storing/querying of *system-* and *user-defined* "metadata" for data sets and resources**
- **API for *ad hoc* querying of metadata**
- **Ability to extend schemas and define new schemas**
- **Ability to associate data sets with multiple metadata schemas**
- **Ability to relate attributes across schemas**
- **Implemented in Oracle and DB2**

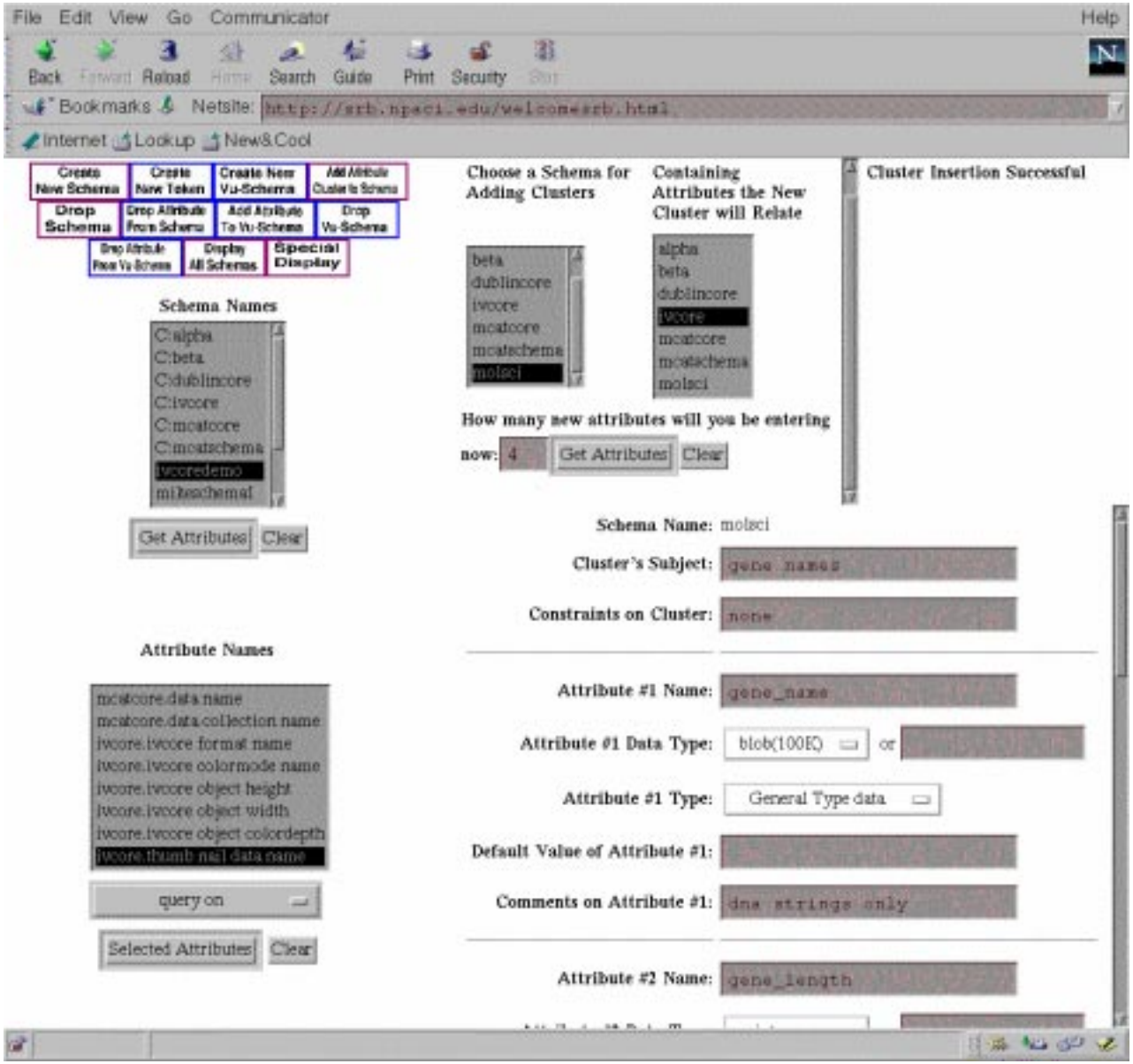
MCAT Schema Integration

- **Publish schema for each collection**
  - Clusters of attributes form a table
  - Tables implement the schema
- **Use Tokens to define semantic meaning**
  - Associate Token with each attribute
- **Use DAG to automate queries**
  - Specify directed linkage between clusters of attributes
  - Tokens - Clusters - Attributes

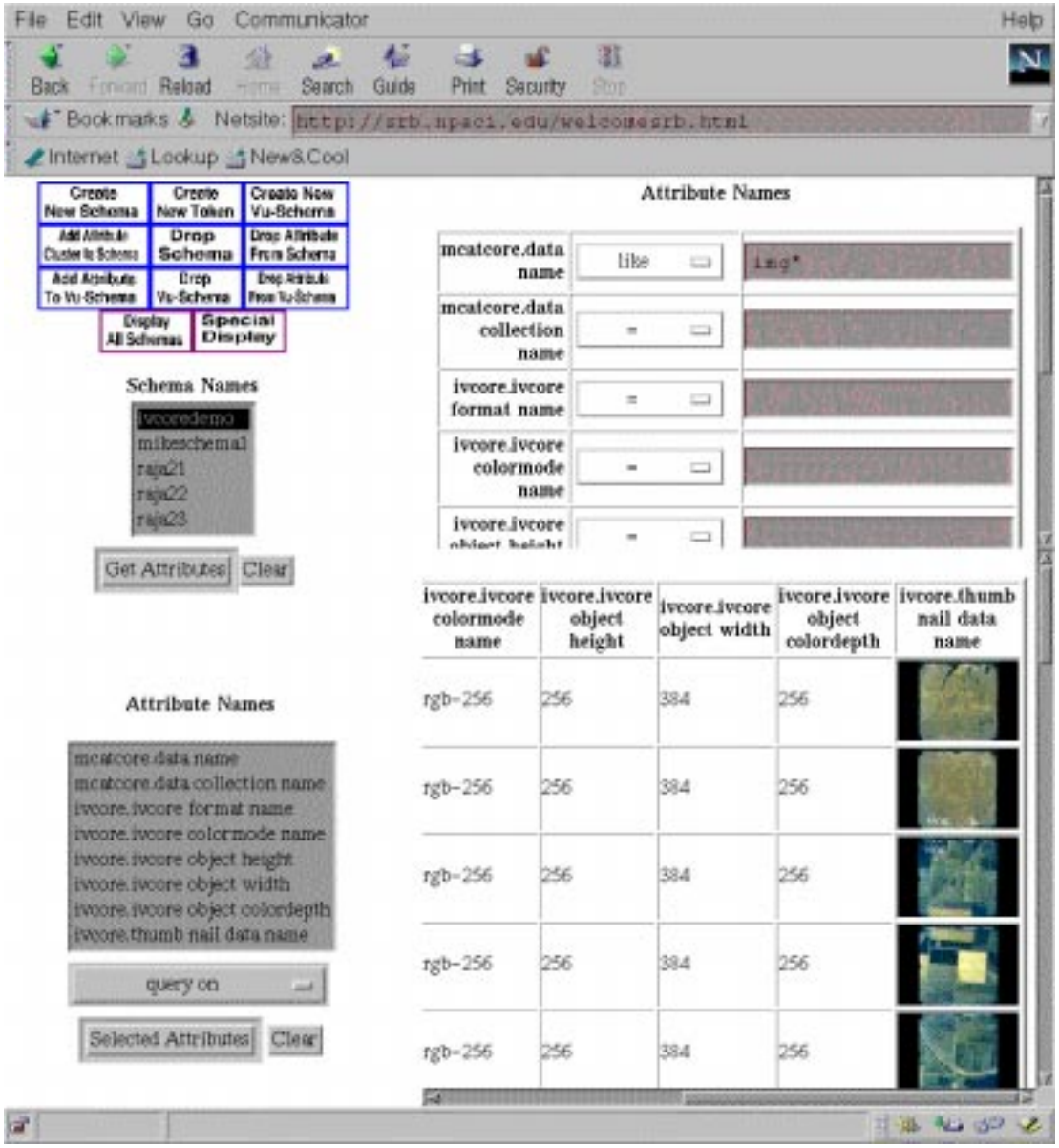
Publishing A New Schema



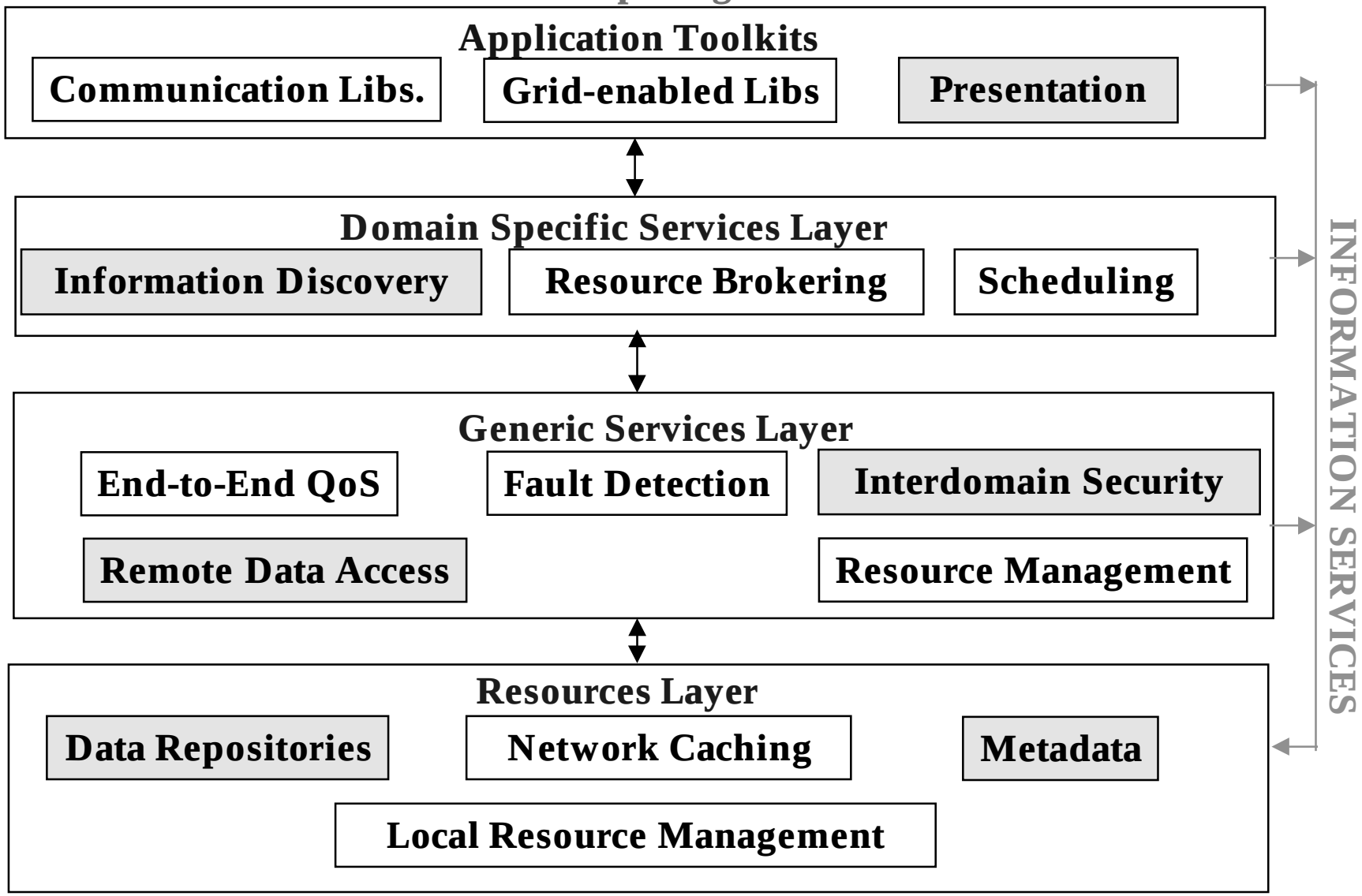
Adding Attributes to the New Schema



Displaying Attributes From Selected Schemas



Data Intensive and High-Performance Distributed Computing



Further Information

http://www.npaci.edu/DICE