SDSC Projects

- **Part 1: BUILDING PRESERVATION ENVIRONMENTS** ***(Reagan Moore, moore@sdsc.edu)***
  - Storage Resource Broker (SRB) and collection migration technologies:
    - Name space management for resources, users, files, metadata, constraints
    - Bulk import of metadata and registration of files directly from file systems
    - Bulk registration of SRB collections into the DSpace technology
  - Goals: understanding mechanisms used to support federation/migration for geodata; understanding collection description sufficient for migration onto supporting infrastructure
- **Part 2: SOME GIS DATA ARCHIVING PROJECTS** ***(Ilya Zaslavsky, zaslavsk@sdsc.edu)***
  - Archiving spatial data /NARA projects
  - A few recent NHPRC or InterPARES supported projects: Maine GeoArchives, VanMap

UCSD

SDSC

Preservation

- Archival processes through which a digital entity is extracted from its creation environment and migrated to a preservation environment, while maintaining authenticity and integrity information.
- Extraction process requires insertion of support infrastructure underneath the digital material, characterization of the authenticity and integrity, characterization of the digital encoding format, and characterization of the display operations
- Goal is infrastructure independence, the ability to use any commercial storage system, database, or access mechanism

Preservation Communities

- **InterPARES - diplomatics**
  - Preservation of records
  - Focus: provenance, authenticity, integrity
- **NARA**
  - Preservation of records from federal agencies
  - Focus: infrastructure independence, scalability
- **State archives**
  - Preservation of submitted “collections”
  - Focus: automation of archival processes

Preservation Strategies

- **Emulation**
  - Migrate the display application onto new operating systems
  - Equivalent to forcing use of candlelight to look at 16th century documents
- **Transformative migration**
  - Migrate the encoding format to the new standard
  - Migration period is expected to be 5-10 years
- **Persistent object**
  - Characterize the encoding format
  - Migrate the characterization forward in time

Data Grids

- **Distributed data management**
  - Share data through creation of collections
  - Manage collections distributed across multiple storage systems
  - Meet patient confidentiality requirements
  - Manage wide area network latencies
  - Support access through preferred APIs
- **Provide storage repository abstractions that make it possible to migrate collections between vendor specific products, while ensuring authenticity**
  - Keeping the collection invariant while the underlying technology (OS, storage system software, access mechanisms, metadata management, etc.) evolves

Preservation Environment

- **Digital library infrastructure that supports**
  - Preservation metadata
  - Arrangement and description of items
  - Access mechanisms
- **Data grid infrastructure that supports**
  - Shared collections that are migrated forward in time
  - Management of technology evolution
  - Administrative metadata providing status of records

Infrastructure Independence

Data Access Methods (Web Browser, DSpace, OAI-PMH)

Storage Repository

- Storage location
- User name
- File name
- File context (creation date,…)
- Access constraints

Naming conventions provided by storage systems

Data Grids Provide a Level of Indirection for Each Naming Convention

Data Access Methods (C library, Unix, Web Browser)

Data Collection

Storage Repository

- Storage location
- User name
- File name
- File context (creation date,…)
- Access constraints

Data Grid

- Logical resource name space
- Logical user name space
- Logical file name space (URID)
- Logical context (metadata)
- Control/consistency constraints

Data is organized as a shared collection

SRB Data Grid Abstractions

- Logical name space for files
  - Global persistent identifier
- Storage repository abstraction
  - Standard operations supported on storage systems
- Information repository abstraction
  - Standard operations to manage collections in databases
- Access abstraction
  - Standard interface to support alternate APIs
- Latency management mechanisms
  - Aggregation, parallel I/O, replication, caching
- Security interoperability
  - GSSAPI, inter-realm authentication, collection-based authorization

Storage Resource Broker 3.3

Application

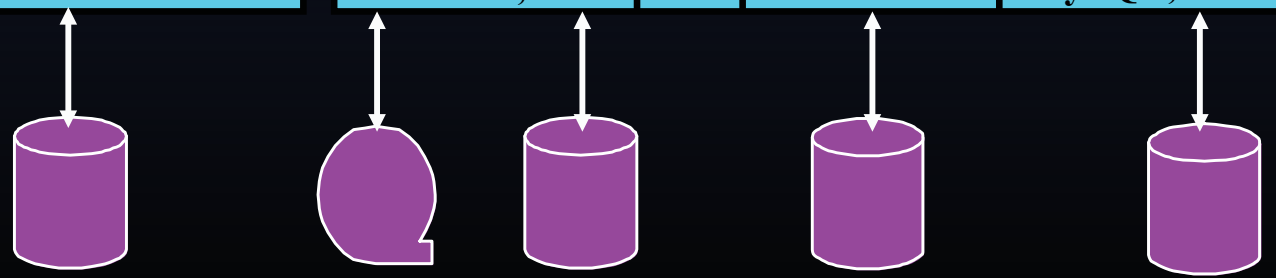
C Library, Java	Unix Shell	Linux I/O C++	NT Browser, Kepler Actors	DLL / Python, Perl, Windows	HTTP, DSpace, OpenDAP, GridFTP	OAI, WSDL, (WSRF)

Federation Management

Consistency & Metadata Management / Authorization, Authentication, Audit

Logical Name Space	Latency Management	Data Transport	Metadata Transport

Database Abstraction	Storage Repository Abstraction			
Databases - DB2, Oracle, Sybase, Postgres, mySQL, Informix	Archives - Tape, Sam-QFS, DMF, HPSS, ADSM, UniTree, ADS	ORB	File Systems Unix, NT, Mac OSX	Databases - DB2, Oracle, Sybase, Postgres, mySQL, Informix



UCSD

SDSC

Examples of Extensibility

- **Storage Repository Driver evolution**
  - Initially supported Unix file system
  - Added archival access - UniTree, HPSS
  - Added FTP/HTTP
  - Added database blob access
  - Added database table interface
  - Added Windows file system
  - Added project archives - Dcache, Castor, ADS
  - Added Object Ring Buffer, Datascope
  - Adding GridFTP version 3.3
- **Database management evolution**
  - Postgres
  - DB2
  - Oracle
  - Informix
  - Sybase
  - mySQL (most difficult port - no locks, no views, limited SQL)

Examples of Extensibility

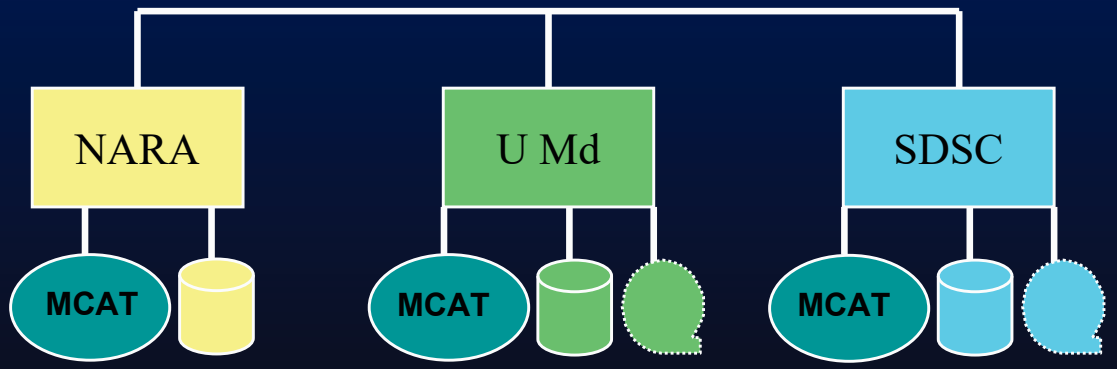
- **The 3 fundamental APIs are C library, shell commands, Java**
  - Other access mechanisms are ported on top of these interfaces
- **API evolution**
  - Initial access through C library, Unix shell command
  - Added iNQ Windows browser (C++ library)
  - Added mySRB Web browser (C library and shell commands)
  - Added Java (Jargon)
  - Added Perl/Python load libraries (shell command)
  - Added WSDL (Java)
  - Added OAI-PMH, OpenDAP, DSpace digital library (Java)
  - Added Kepler actors for dataflow access (Java)
  - Adding GridFTP version 3.3 (C library)

National Archives and Records Administration - Research Prototype Persistent Archive

Demonstrate preservation environment

- Authenticity
- Integrity
- Management of technology evolution
- Mitigation of risk of data loss
  - Replication of data
  - Federation of catalogs
- Management of preservation metadata
- Scalability
  - EAP collection
  - 350,000 files
  - 1.2 TBs in size

Federation of Three Independent Data Grids



Principle copy stored at NARA with complete metadata catalog

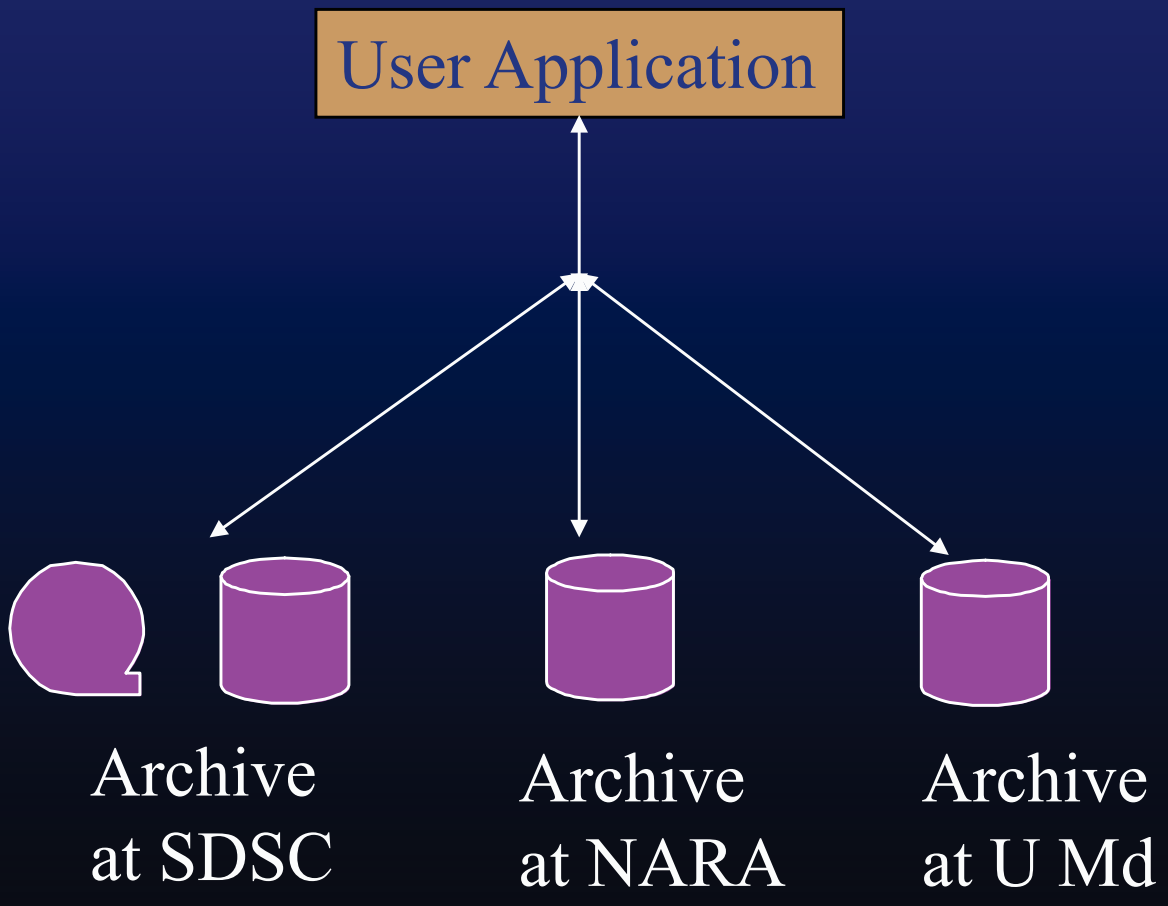
Replicated copy at U Md for improved access, load balancing and disaster recovery

Deep Archive at SDSC, no user access, but complete copy

UCSD

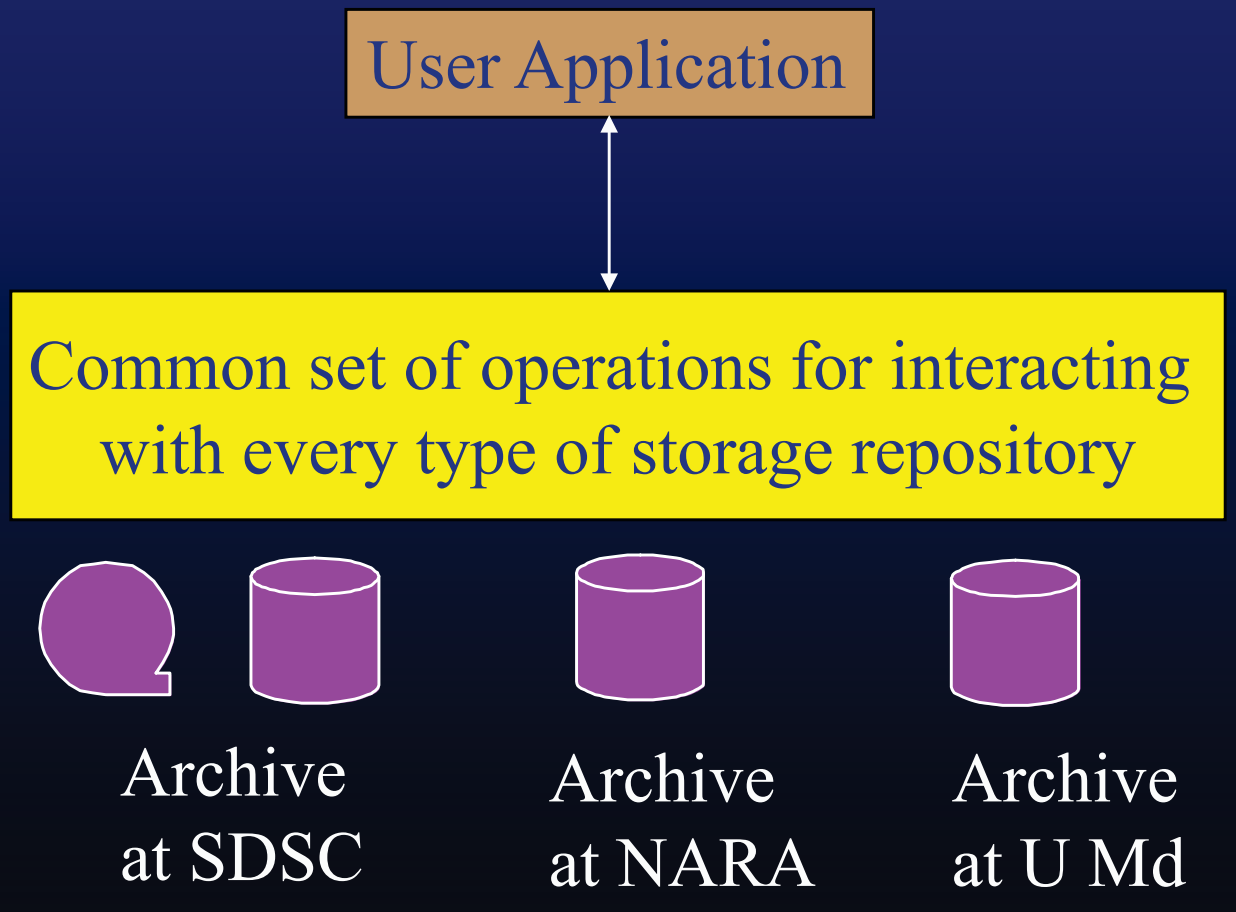
SDSC

Accessing Multiple Types of Storage Systems



Standard Data Access Operations

- Remote operations
  - Unix file system
  - Latency management
  - Procedures
  - Transformations
  - Third party transfer
  - Filtering
  - Queries
- Collective operations
  - Replication
  - Fault tolerance
  - Load leveling

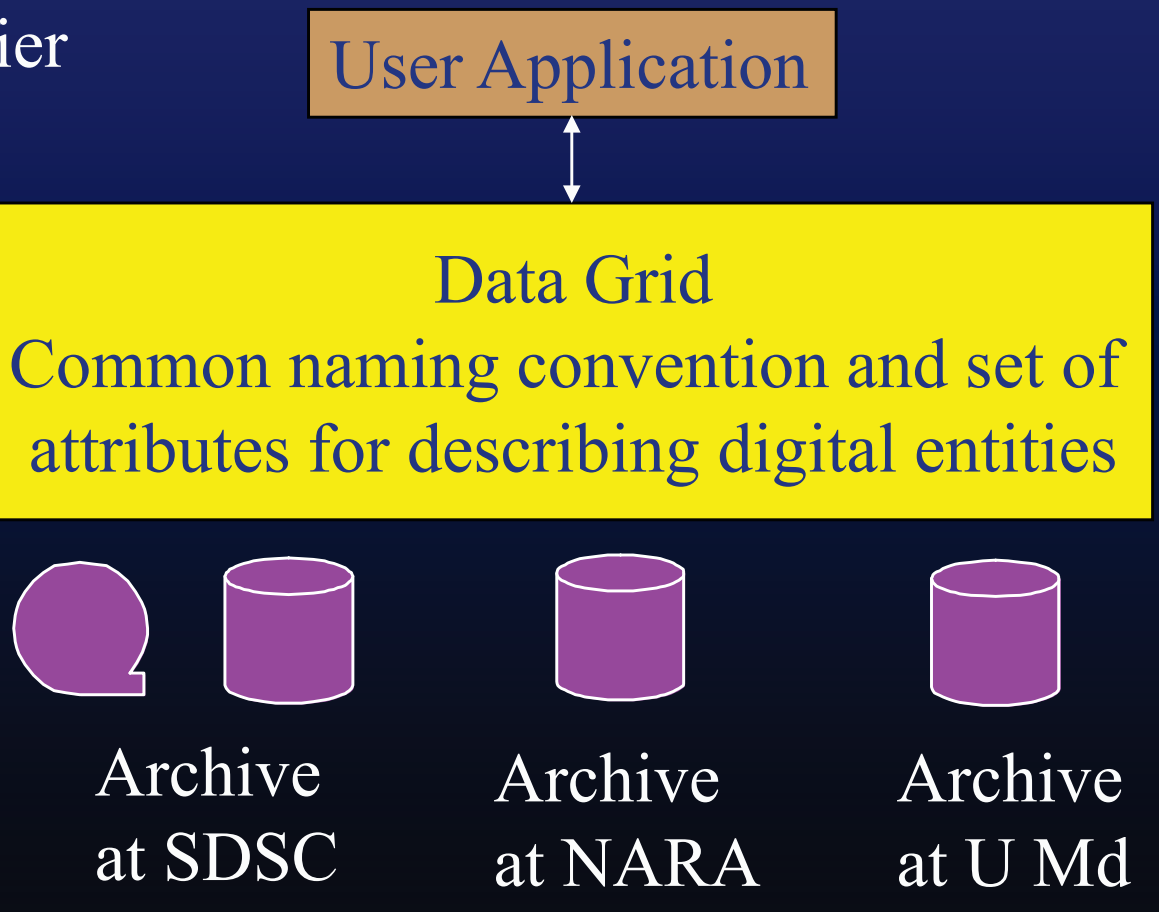


Building a Distributed Collection

Logical name space
- Location independent identifier
- Persistent identifier

Collection owned data
- Authenticity metadata
- Access controls
- Audit trails
- Checksums
- Descriptive metadata

Inter-realm authentication
- Single sign-on system



Storage Resource Broker Collections at SDSC (11/2/2004)	GBs of data stored	Number of files	Number of Users
Data Grid	Ź	Ź	Ź
NSF/ITR - National Virtual Observatory	53,858	9,536,698	80
NSF - National Partnership for Advanced Computational Infrastructure	24,738	5,754,890	380
Hayden Planetarium - Evolution of the Solar System visualizations	7,201	113,600	178
NSF/NPACI - Joint Center for Structural Genomics	5,228	652,031	50
NSF/NPACI - Biology and Environmental collections	8,851	33,340	67
NSF - TeraGrid, ENZO Cosmology simulations	121,550	1,096,947	3,247
NIH - Biomedical Informatics Research Network	6,002	4,107,508	214
Digital Library	Ź	Ź	Ź
NLM - Digital Embryo image collection	720	45,365	23
NSF/NPACI - Long Term Ecological Reserve	253	8,436	36
NSF/NPACI - Grid Portal	2,211	51,227	407
NIH - Alliance for Cell Signaling microarray data	856	62,291	21
NSF - National Science Digital Library SIO Explorer collection	2,080	808,901	27
NSF/NPACI -Transana education research video collection	92	2,387	26
NSF/ITR - Southern California Earthquake Center	91,040	1,791,494	62
Persistent Archive	Ź	Ź	Ź
UCSD Libraries archive	128	204,828	29
NARA- Research Prototype Persistent Archive	166	316,813	58
NSF - National Science Digital Library persistent archive	3,571	26,908,350	122
TOTAL	328 TB	51 million	4,900

DSpace Familiar As:

- **Simple user-friendly front end providing:**
  - Digital content ingestion
  - Indexing, search and discovery
  - Content management
  - Dissemination services
- **Jointly developed by:**
  - MIT Libraries
  - Hewlett-Packard (HP)

UCSD

SDSC

Use SRB as filestore for DSpace bitstreams

Simple User Interface

DSpace

Content
- Ingestion
- Discovery
- Dissemination

+

"Unlimited" Storage

SRB

Uniform interface to storage
- Distributed
- Heterogeneous

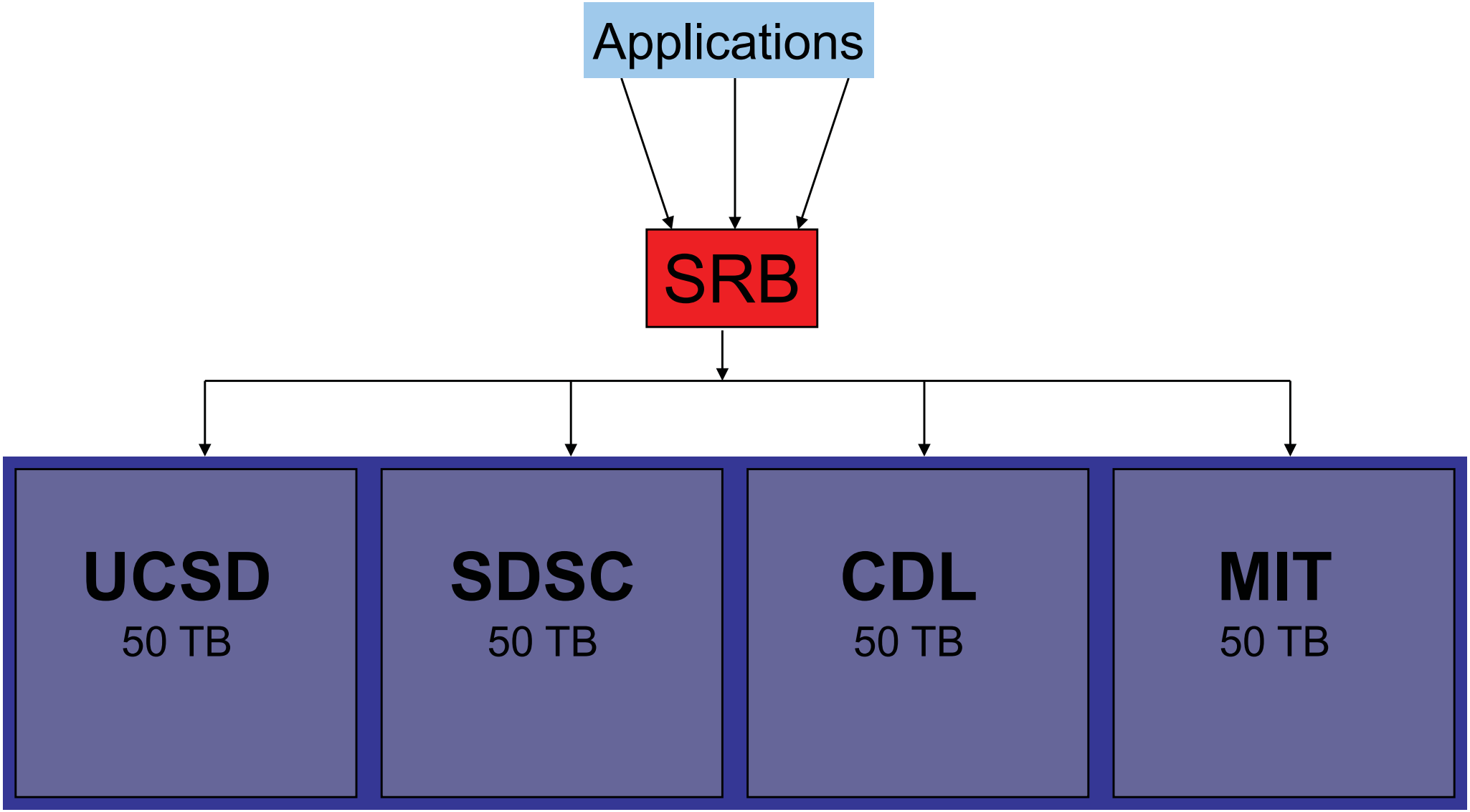
STEPS:

•Replace DSpace file system calls with SRB access calls

•Employ METS based Archival Information Package (AIP)

•Enable exchange of data and metadata between independent DSpace and SRB systems

•Validate authenticity of exchanged content



200 TB

Single Logical Resource

Archival Processes

- **Archival form = Original bits of the digital entity + Archival context:**

Preservation Function	Type of information
Administrative	Location, physical file name, size, creation time, update time, owner, location in a container, container name, container size, replication locations, replication times
Descriptive	Provenance, submitting institution, record series attributes, discovery attributes
Authenticity	Global Unique Identifier, checksum, access controls, audit trail, list of transformative migrations applied
Structural	Encoding format, components within digital entity
Behavioral	Viewing mechanisms, manipulation mechanisms

Persistent archive processes

Archival Process	Functionality
Appraisal	Assessment of digital entities
Accession	Import of digital entities
Description	Assignment of provenance metadata
Arrangement	Logical organization of digital entities
Preservation	Storage in an archive
Access	Discovery and retrieval

Archiving and accessing spatial data

- Archival forms for spatial data
  - Variety of data types and models (raster, vector, 3D…)
  - Different spatial registration mechanisms
  - Data structures with different amounts of "intelligence"
  - Data quality; multi-scale… typical geographic issues…
- Infrastructure independence
  - Lots of proprietary formats and management systems, vendor-locked… several emerging XML encoding standards …
  - Demo of an XML-based online GIS with NARA Herbicides collection
  - Web services on top of various spatial information systems; standard encoding of spatial processing functionality

Some metadata issues

- What constitutes context of geospatial data
- Adding geospatial metadata to archival metadata?
- Feature-level vs layer-level metadata
- Data quality
  - Depends on feature type, measurement procedures, transformations and conversions, etc…
  - DRG (1995-98 the "scanning phase", but maps developed over much longer periods of time)
  - Common lineage at large scales (USGS topo quads)– for DRG, DEM, DLG, DOQ data series; a "lineage map" ?
- Organization:
  - multi-scale, multi-resolution
    - 7.5' (1:24,000) – 30x60' – 1x2degree (DRGs)
  - Multi-hierarchies
    - 7.5x7.5' – 30x30' – 1x1degree (DEMs) … Alaska and Hawaii different
  - Multiple data contributors

Encoding examples: DTD Fragments for Raster Data

From GLCF (U of Maryland)

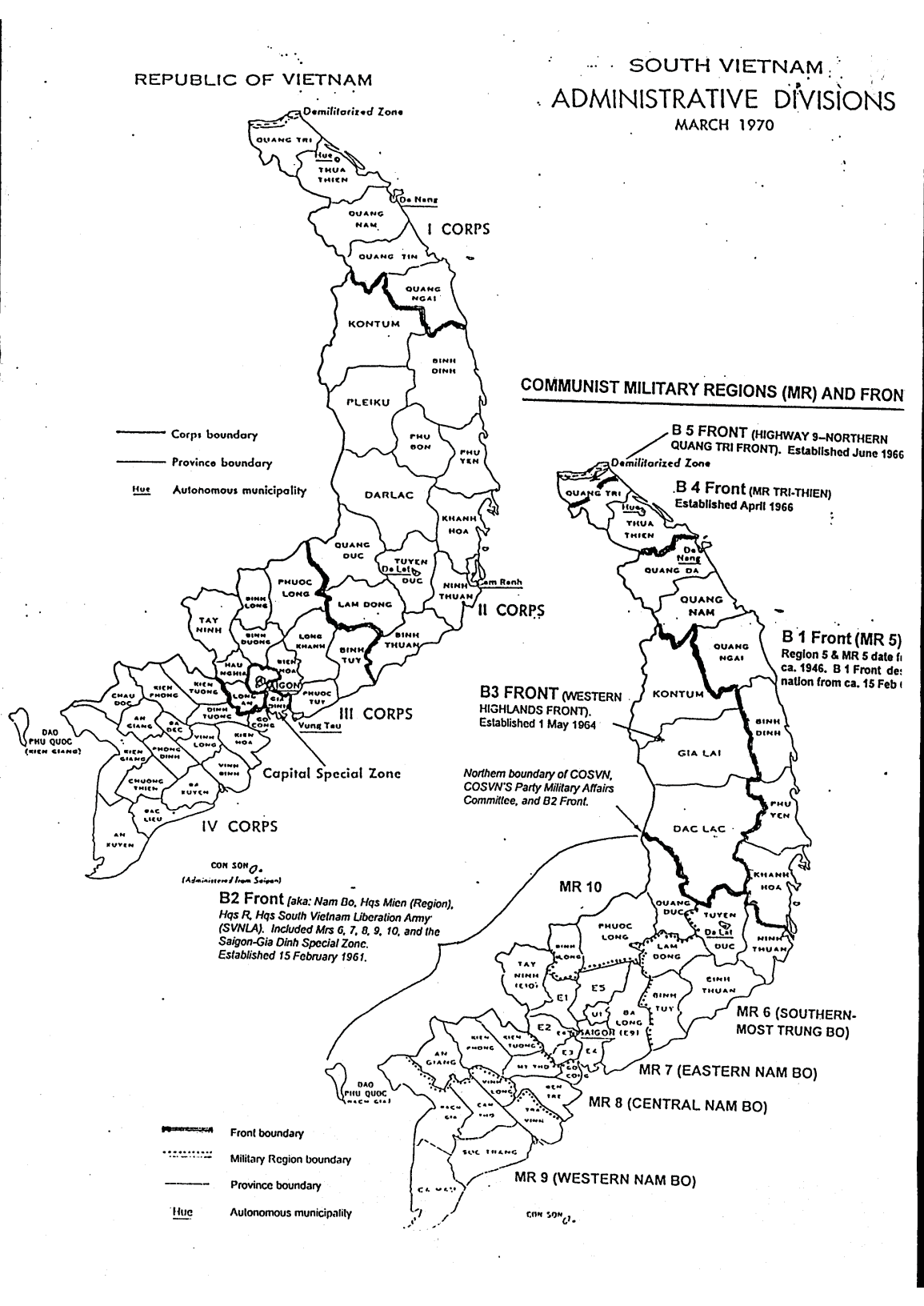
```
<!ELEMENT GRANULE (GLOBAL_CONTEXT, GRANULE_DATA,
                   PROCESSING?, SIGNATURE?)>
 CONTEXT  (SATELLITE | LANDSAT | LANDSAT_SATELLITE | DRG | DOQ_STD | DOQ_MN) #REQUIRED
<!ELEMENT GLOBAL_CONTEXT (SENSORING?, DATA_SET, PROJECTION, COORDINATION,
                          WRS?, REFERENCE?, COVERAGE)>
<!ELEMENT SENSORING (SATELLITE?, SENSOR?, SOLAR_LOCATION?)>
<!--        The valid values for SATELLITE element are: LANDSAT4, LANDSAT5, -->
<!--        LANDSAT7, NOAA11, NOAA12, NOAA13, NOAA14, NOAA15 and TERRA.     -->
<!ELEMENT SATELLITE (#PCDATA)>
<!--        The valid values for SENSOR element are: TM, MSS, AVHRR, ETM+   -->
<!--        and MODIS                                                       -->
<!ELEMENT SOLAR_LOCATION (SOLAR_ELEVATION, SOLAR_AZIMUTH)>
<!--        Both SOLAR_ELEVATION and SOLAR_AZIMUTH have format of real(3.2) -->
<!ELEMENT SOLAR_ELEVATION (#PCDATA)>
<!ELEMENT DATA_SET (DATA_SET_NAME, DATA_SET_ID?)>
<!--        The valid values for DATA_SET_NAME and associated CODE are:    -->
<!--          1   :   LANDSAT_THEMATIC_MAPPER                              -->
<!--          2   :   LANDSAT_MULTISPECTRAL_SCANNER                        -->
<!--          3   :   ADVANCED_VERY_HIGH_RESOLUTION_RADIOMETER_GAC         -->
<!--          4   :   CENTRAL_AFRICAN_REGIONAL_PROGRAM_FOR_THE_ENVIRONMENT -->
<!--          5   :   UNITED_STATES_COASTAL_MARSH_HEALTH                   -->
<!--          6   :   GLOBAL_LAND_COVER                                    -->
<!--          7   :   LANDSAT_ENHANCED_THEMATIC_MAPPER                     -->
<!--          8   :   GLOBAL_LAND_COVER_DERIVED_FROM_AVHRR-1KM             -->
<!--          9   :   GLOBAL_LAND_COVER_DERIVED_FROM_AVHRR-8KM             -->
<!--         10   :   GLOBAL_LAND_COVER_DERIVED_FROM_AVHRR-1DEGREE         -->
<!--         11   :   CONTINUOUS_FIELDS_TREE_COVER_PROJECT                 -->
<!--         12   :   AVHRR_DERIVED_FROM_GAC-8KM                           -->
<!--         13   :   DIGITAL_RASTER_GRAPHICS                              -->
<!--         14   :   DIGITAL_ORTHOPHOTO_QUADRANGLES_STD                   -->
<!--         15   :   DIGITAL_ORTHOPHOTO_QUADRANGLES_MN                    -->
```

DTD Fragments - 2

```
<!ELEMENT PROJECTION (PROJ_NAME, ELLIPSOID, PROJ_PARA?)>
<!--        The valid values for PROJECTION_NAME and associated CODE are:  -->
<!--          0    :   GEOGRAPHIC                                          -->
<!--          1    :   UNIVERSAL_TRANSVERSE_MERCATOR                       -->
<!--          3    :   ALBERS_CONICAL_EQUAL_AREA                           -->
 ……
<!ELEMENT ELLIPSOID (ELLIPSOID_NAME, ELLIPSOID_AXIS?,
                     ELLIPSOID_OFFSET?)>
<!--        The valid values for ELLIPSOID_NAME and associated CODE are:   -->
<!--          0    :   CLARKE_1866                                         -->
<!--          1    :   CLARKE_1880                                         -->
<!--          2    :   BESSEL                                              -->
……..
<!ELEMENT COORDINATION (MAP_ZONE?, DATUM?, COORD_UNIT_NAME, POLYGON)>
<!--        MAP_ZONE is the basis for the coordinates of this granule. This -->
<!--        is required if, and only if, the map projection is              -->
……..
<!ELEMENT POLYGON (UPPER_LEFT_CORNER, UPPER_RIGHT_CORNER,
                   LOWER_LEFT_CORNER, LOWER_RIGHT_CORNER)>
<!--        All following coordinates must be in the projection specified   -->
……….
<!ELEMENT REFERENCE (REF_PNT_COORD?, REF_PNT_OFFSET_PIXEL?,
                     ORIENTATION?)>
<!--        Here is the X and Y coordinate used to geographically          -->
<!--        reference the image to the ground. Expressed in the projection -->
<!--        specified by PROJECTION_NAME and in units specified by         -->
<!-- ============================COVERAGE================================ -->
<!--        If this granule covers a range of time, this is the starting    -->
<!--        date and time in the range. If this granule covers a specific   -->
<!--        point in time, START_DATA_TIME and END_DATA_TIME should have    -->
<!--        the same value.                                                 -->
<!ELEMENT COVERAGE (START_DATA_TIME, END_DATA_TIME, PLACE_NAME?)>
```

```
<!-- ============================================================== -->
<!--                          GRANULE_DATA                          -->
<!-- ============================================================== -->
<!ELEMENT GRANULE_DATA (FILE_ATTR, GRAN_PREVIEW, HEADER_FILE*, DATA_FILE+)>
<!--       SIZE :              Total size in bytes of all files of this   -->
<!--                           granule, excluding header files, and other -->
<!--                           ancillary files.                           -->
<!--       MISSING_DATA :      Percentage of this granule that has data   -->
<!--                           that has data missing. It is expressed in  -->
<!--                           Real(1.2) format.                          -->
<!--       INTERPOLATED_DATA : Percentage of this granule that has data   -->
<!--                           interpolated. It is expressed in Real(1.2).-->
<!--       OUT_OF_BOUNDS_DATA :Percentage of this granule that has data   -->
<!--                           that is out of bounds. It is expressed in  -->
<!--                           Real(1.2) format.                          -->
<!ATTLIST GRANULE_DATA ORIENTATION (UPPER_LEFT_RIGHT | UPPER_RIGHT_LEFT
                                  | BOTTOM_LEFT_RIGHT | BOTTOM_RIGHT_LEFT
                                  | UPPER_LEFT_BOTTOM | UPPER_RIGHT_BOTTOM
                                  | BOTTOM_LEFT_TOP | BOTTOM_RIGHT_TOP)
                                                #IMPLIED
                     SIZE               CDATA    #REQUIRED
                     CLOUD_COVER        CDATA    #IMPLIED
                     MISSING_DATA       CDATA    #IMPLIED
                     INTERPOLATED_DATA  CDATA    #IMPLIED
                     OUT_OF_BOUNDS_DATA CDATA    #IMPLIED
>
 !-- ============================DATA_FILE================================== -->
<!ELEMENT DATA_FILE (BAND_PREVIEW?, BAND_IMAGE)>
<!--       The ID attribute is used to identify the band number            -->
<!ATTLIST DATA_FILE     ID   CDATA  #REQUIRED  >

<!-- ============================================================== -->
<!--                           SIGNATURE                            -->
<!-- ============================================================== -->
<!ELEMENT SIGNATURE (AUTHOR?, WORK_UNIT?, CONTACT?,
                     LAST_MODIFIED_DATE?, COMMENT?)>
```



More archival format examples

NARA Herbicides Collection

Herbicides Collection - 1

From EBCDIC tapes:

```
6507213207565    260404040 040000{0000D0000000{048{ {0000000{0000000{0000000{0000000{
6507243207565    260606060 060000{0000D0000000{072{ {0000000{0000000{0000000{0000000{
6507253207565    260606060 060000{0000D0000000{072{ {0000000{0000000{0000000{0000000{
6507263207565    260606060 060000{0000D0000000{072{ {0000000{0000000{0000000{0000000{
6507273207565    260606060 060000{0000D0000000{072{ {0000000{0000000{0000000{0000000{
6507283207565    260505050 050000{0000D0000000{060{ {0000000{0000000{0000000{0000000{
6507293207565    260404040 040000{0000D0000000{048{ {0000000{0000000{0000000{0000000{
6508022022365    060202020 010000{0000C0000000{012{ {0000000{0000000{0000000{0000000{1A
    AS890255                           000{000{
6508022022365                                                                      1B
    AS940140                           000{000{
6508042022365    060202020 006000{0000C0000000{007B {0000000{0000000{0000000{0000000{1A
    AS925205                           000{000{
6508042022365                                                                      1B
    AS970065                           000{000{
6508062022365    060202020 004000{0000C0000000{004H {0000000{0000000{0000000{0000000{1A
    BS290320                           000{000{
6508062022365                                                                      1B
    BS275298                           000{000{
6508073207565    260202020 020000{0000D0000000{024{ {0000000{0000000{0000000{0000000{1A
    YT080110                           000{000{
6508073207565                                                                      1B
    YT110060                           000{000{
6508113207565    260202020 020000{0000D0000000{024{ {0000000{0000000{0000000{0000000{
6508123207565    260202020 020000{0000D0000000{024{ {0000000{0000000{0000000{0000000{
6508151022465    020202020 008000{0000C0000000{009F {0000000{0000000{0000000{0000000{1A
    YD350155                           000{000{
6508151022465                                                                      1B
    YD450150
```

Herbicides Collection - 2

Converted to XML:

```
<YEAR><yearnum>66</yearnum>
<MONTH><monthnum>01</monthnum>
<DATE><datenum>01</datenum>
<MISSION><num>206866</num>
        <RUN><code>A</code>
        <ctz>3</ctz><multi></multi><prov>27</prov>
                <aircrafts>
                        <scheduled>02</scheduled><airborne>02</airborne><productive>02</productive>
                </aircrafts>
                <agent>O</agent><gal>02000</gal><hits>0</hits>
                <aborts>
                <maintenance>0</maintenance><weather>0</weather><battle_damage>0</battle_damage><other>0</other>
                </aborts>
                <type>D</type><area>024</area><rsult></rsult>
                <UTM>
                        <utmid>1A</utmid>
                        <utm_coor>YS240780</utm_coor>
                </UTM>
                <UTM>
                        <utmid>1B</utmid>
                        <utm_coor>YS290630</utm_coor>
                </UTM></RUN>
        <RUN><code>B</code>
        <ctz>3</ctz><multi></multi><prov>27</prov>
                <aircrafts>
                        <scheduled>02</scheduled><airborne>02</airborne><productive>02</productive>
                </aircrafts>
                <agent>O</agent><gal>02000</gal><hits>0A</hits>
                <aborts>
                <maintenance>0</maintenance><weather>0</weather><battle_damage>0</battle_damage><other>0</other>
                </aborts>
                <type>D</type><area>024</area><rsult></rsult>
```

Herbicides

Long-term digital records preservation

- National level agenda:
  - Archives of Australia, UK, NARA
  - Archival formats for databases, binary data, png and jpeg, etc., but no specific GIS guidelines (Australia)
- At the state level
  - Usually GIS preservation policies are not specific
    - E.g. Maryland's GIS preservation policy… mentions lack of standard GIS preservation formats, doesn't consider frequencies
    - Maine GeoArchives project (later…)

Database preservation

- A recent ERPANET research report:
- Key considerations for database preservation:
  - Appraisal should consider the whole information system (purpose, design, context), and costs
  - Archive snapshots, or archive data marked for deletion
  - Defined isolated “archivable” parts in a federated database, and relationships between them
  - Archiving data types, check constraints (referential integrity is critical)
  - Description is often difficult
  - Preservation must extract data from their native environments, while guaranteeing authenticity. Must be automated
  - Include access considerations from the start

Database preservation - 2

- Observations/case studies:
  - In most cases, data exported into XML, flat files, or a mixture of the two
  - One of case studies (Antwerp) mentions preserving GIS data as GML
  - All case studies followed the migration strategy, which appears to be much more favored for preserving databases than emulation

Challenges / research issues

- Preservation should be a collaborative and distributed (due to the nature of geographic data collection) effort between data providers and archive providers. How such collaboration is organized? How will distributed archive architectures look like?
- How revisions in the data (at what schedule) are appraised and propagated to archives, both organizationally and technically?
- What are the archival forms for different types of geospatial data?
- What are the archival metadata standards specific for spatial data?
- What is the right combination of snapshot and event-based archiving for different types of data and different update schedules?
- What is the right combination of proprietary and open GIS formats and data handling techniques balancing long-term preservation and easy access needs?
- Whether and how particular access/visualization interfaces should be preserved along with geospatial data?
- How integrity of geospatial records should be verified, and at which levels (i.e. logical consistency, semantic integrity)?
- What is the optimal level of redundancy in archiving geospatial data?

tGIS queries

- When did Feature X exist or cease to exist?
- What existed at Location A at Time T?
- What happened to a given feature or location between Time T1 and T2?
- Did Event A exist before or after Condition X (or Event B)?
- What patterns exist between Events A-B-C and Features X-Y-Z?
- Given data for Feature Y at Time T1 & T3, what was the likely state of this Feature at Time T2?
- What will be the likely state of Feature X at Time T?
- What is the predicted outcome following Event A after Time T?

Maine GeoArchives

- Goals: operational GeoArchives prototype, archiving GIS records that have permanent value; developing related standards
- Funding: NHPRC
- Multi-agency and multi-municipality archiving
- Existing system: Oracle + ArcSDE; complete FGDC metadata for all layers; several sample layers to archive

Interesting issues:

- Adequacy of the current model
- Archiving frequency (what is expected accuracy)
- Feature-level appraisal (“informational value”) and metadata
- Change management
- Enhancing archival metadata (with provenance/lineage, data quality, data types, etc.)
- One model doesn’t fit all layers
- Archival format
- Preserving access mechanisms

VanMap

- VanMap: federates information from multiple Vancouver city departments, displays interactive maps (Oracle Spatial, Autodesk map viewer)
- Requirements:
  - Non-proprietary system for managing GIS data
  - Ability to import GIS data from a proprietary system into the preservation environment
  - Ability to query and display the GIS data in a similar fashion
  - Ability to archive web pages, “preserve look and feel”
  - Appraisal mechanism (“information value”?)
- Challenges:
  - Different update frequencies
  - Interactive browsing and mapping archived data
  - Preservation Model: snapshots, recording changes, a hybrid
  - Retaining relationships between spatial and related report data
  - What to do with hyperlinks to web pages