Recovery of a Digital Image Collection Through the SDSC/UMD/NARA Prototype Persistent Archive[1]

Mike Smorul, Joseph JaJa, Fritz McCall, Susan Fitch Brown
Institute for Advanced Computer Studies, University of Maryland

Reagan Moore, Richard Marciano, Sheau-Yen Chen
San Diego Supercomputer Center, University of San Diego

Rick Lopez, Robert Chadduck
National Archives and Records Administration

Abstract

The San Diego Supercomputer Center (SDSC), the University of Maryland, and the National Archives and Records Administration (NARA) are collaborating on building a pilot persistent archive using and extending data grid and digital library technologies. The current prototype consists of node servers at SDSC, University of Maryland, and NARA, connected through the Storage Request Broker (SRB) data grid middleware, and currently holds several terabytes of NARA selected collections. In particular, a historically important image collection that was on the verge of becoming inaccessible was fully restored and ingested into our pilot system. In this report, we describe the methodology behind our approach to fully restore this image collection and the process used to ingest it into the prototype persistent archive.

1. Introduction

The San Diego Supercomputer Center (SDSC), the University of Maryland (UMD), and the National Archives and Records Administration (NARA) are collaborating on building a pilot persistent archive using and extending data grid and digital library technologies. The initial phase of the project developed abstractions for managing data objects, data repositories, collections, and knowledge processes. Briefly, the data abstraction specifies an encoding format that enables the presentation application to correctly interpret the stream of bits defining the data object. The data repository abstraction enables the association of a persistent (unique) identifier with each data object, independent of the local file name or where it is stored at any particular time instance. The collection abstraction defines context (expressed as metadata) associated with the archived data, while the knowledge abstraction provides semantic relationships among attributes of an object and a description of the processing steps used to create and validate the data object. Mechanisms dealing with technology evolution need to be applied at each abstraction level.

[1] Sponsored by the National Archives and Records Administration and the National Science Foundation under the PACI program.

The pilot system under development is based on an architecture that consists of a stack of three well-defined layers: the data layer, the information management layer, and the knowledge/user interface layer, requiring well-defined interfaces and protocols between adjacent layers. The data layer is responsible for managing the bits stored in various repositories, which are in general widely distributed and heterogeneous. The principal function of this layer is to hide the details about any particular repository and to assign a unique identifier to any stream of bits upon the request of the information management layer. This identifier enables the data layer to retrieve the corresponding stream of bits, regardless of its actual physical location on any of the repositories of the archive. This layer also is responsible for data replication, encryption, compression, and data integrity in response to requests from the information management layer. It is also stipulated that the responsibility for data preservation at the bit level falls upon this layer. Note that we make no assumption about whether the overall control of this layer is centralized, hierarchical, or fully distributed. In our current prototype, the data layer is implemented using the Storage Request Broker (SRB) that provides many of the necessary functionalities needed for this layer.

The information management layer deals with the semantics of the data rather than storage and bits. In particular, it maintains metadata associated with data objects, and enables the organization of objects into collections, as well as maintains information about access control privileges, data curation and stewardship. Preservation strategies at the collections level are incorporated within this layer. As in the case of the data layer, we expect the different components to be distributed and heterogeneous, and no assumption is made as to the coordination approach among these components. The MCAT (Metadata CATalog) associated with the SRB is currently used to implement the information management layer of our prototype, which is currently being extended to handle the peer-to-peer federation of multiple MCATs.

The user interface layer enables the users to gain knowledge about relationships among the members of a collection, and to quickly identify and access data objects of interest. Knowledge describing relationships between information attributes of the collections, structural relationships described by the encoding format, and relationships between the different collections, are managed by this layer.

We are in the process of building a prototype persistent archive based on the architecture described above. The pilot consists of “grid bricks” or node servers at SDSC, the University of Maryland, and NARA, glued together through the SRB middleware. A separate MCAT catalog (implemented in Oracle databases at SDSC and NARA and an Informix database at Maryland) is set up at each of the grid bricks. The prototype currently manages several terabytes of NARA selected collections. The largest such collection is a historically important image collection that was in danger of becoming inaccessible because of technology changes. In the next sections, we briefly introduce the collection and our methodology for fully restoring and ingesting this collection into our pilot system.

2. A Brief Description of the Image Collection and its Historical Importance

As part of NARA's Electronic Access Project (EAP), approximately 124,000records were digitized during 1997 through 1999. The images constitute a rich historical collection of photographs, drawings, maps, charts and textual documents. This collection includes digital images of:

- Watercolor sketches by John J. Young from an 1859 exploration of the Utah territory;
- Civil War maps, plans, engineering drawings, diagrams;
- Photographs of civil works projects in northwestern states, 1900-1952;
- Groundbreaking photographs by Lewis Hine documenting child labor abuses for the National Child Labor Committee, 1908-1912;
- Photographs and documents from a 1921 survey of Blackfeet Indians;
- Original sketches drawn by artist Charles Alston to highlight the participation of African Americans during World War II;
- Photographs of the Kennedy White House;
- United States Information Agency reports on U.S. involvement in the war in Vietnam and on the impact of race relations in the U.S. on American foreign policy;
- Letters from Albert Einstein to the U.S. Navy's Bureau of Ordnance, 1943-1944;
- Photographs of the Three Mile Island incident;
- Rare images from World War I;
- An 1867 census of freedmen and their descendants of the Cherokee Nation;
- Photograph albums from the Engineering and Research Center of the Bureau of Reclamation, 1903-1972.

This collection represents a significant cross-section of NARA holdings, and was made available on line during 1997 through 1999. However only access files and thumbnail images were made available over the web due to the relatively large space requirement for the time. As of early 2002, several serious issues regarding the preservation of the master digital image file collection had arisen, which we describe in the next section. We also discuss the steps taken to fully recover the collection and ingest it into our pilot system.

3. Restoration Process and the Ingestion into the Pilot Persistent Archive

We describe in some detail the status of the collection prior to the restoration process, followed by the steps taken to fully restore and ingest the collection into our pilot system.

Media Description

The digital images were stored on 5.25" Write-Once, Read Many (worm) media, with the expectation that the media would have a lifetime measured in tens of years. The image collection was stored on more than 450 2.6-Gb double-sided WORM cartridges. The cartridges were written using Windows 95 native drivers. Thumbnail and access images were stored on 44 CDROMS in JPEG and GIF formats. The original scanning project organized the images into 44 different batches.

In addition to the images, metadata was written to CDROM in several formats. The first format was a series of text files containing an export of the entire NARA Archival Information Locator (NAIL) database. The NAIL export contained both single and multi-line entries labeled with an upper-case identifier followed by white space. Descriptions of the codes were supplied in a separate MicroSoft Excel spreadsheet. Second, a single text file with tilde-separated fields containing descriptions from the NARA website was provided. Last, two sets of item tracking databases from the original scanning project were supplied in MicroSoft Access format.

Media Risk

Since the OptiMF2 filesystem on the WORM disks was no longer under active maintenance there was a high risk that the data would no longer be readable even though the media were in fine condition. NARA only had one workstation left from the original scanning project that was able to read the cartridges. This workstation's operating system only had the ability to use one external driver and was not able to use a robot to bulk process the files. The Liberty Systems' Optisys LLC sells a driver for reading these cartridges, but had stopped producing new versions of the OptiDriver for any operating systems newer than Windows 98. Clearly, this is a typical scenario confronted by digital preservation through technology obsolescence. In this case, the media was fine, but the storage repository software had become obsolete through reliance on an operating system that had a limited life.

Ingestion Requirements

Due to the obsolete nature of the OptiMF2 file system, procuring drivers for modern operating systems was impossible. NARA was able to locate contact information for Liberty Systems, LLC and verify the OptiDriver V3.08 / OPT-AC-2 would read the media.

NARA supplied a HP SureStore 80EX robot for bulk processing of 16 worm cartridges at a time. UMIACS supplied an older PC running Windows 98. Due to time constraints and software limitations on the ingestion workstation, it was decided to stage the worm data to temporary network attached storage rather than ingest directly into the SRB. Temporary storage was provided through an SMB mounted volume from the UMIACS grid brick. Ingestion into the SRB was then performed on the grid brick.

Media transfer between NARA and UMIACS was arranged to process batches of 64 cartridges. Chain of custody was recorded on paper forms required by NARA. This paperwork in addition to log files generated during the ingestion process provides a solid chain of custody for the archive.

Several scripts were written in PERL to run on the ingestion workstation. These scripts copied data from worm media to temporary storage and provided an interface to record physical labels from the worm media.

Each 16-cartridge batch was processed and ingested into the SRB separately. Each batch recorded the physical label of each worm cartridge, directory listings of the original worm media, and log files of the ingestion process.

The OptiDriver presented the entire robot as a single DOS driver letter with the logical label on each cartridge side appearing as a directory under the DOS drive. Contents for each cartridge side could be accessed under the corresponding label. Attempting to retrieve any file under a volume would cause the robot to automatically load the appropriate cartridge side.

Ingestion Pipeline for master images

The ingestion pipeline involved a two-step process. First images were loaded from the robot onto temporary network attached storage. Several scripts were used to automate the process and generate appropriate log files during the copying. Recorded information included directory listings of the media, timing information and records of success or failure of a transfer.

In a second step, the images were registered and loaded into the SRB data grid. If a driver for the OptiMF2 filesystem had been available for a modern file system, it would have been possible to create a SRB server that could talk directly to the original storage system, without having to stage the data onto the temporary disk cache. An expectation is that the degree of assurance associated with the migration effort would be greater when all operations are performed and automatically audited by the data management system. UMIACS was able to ingest 1-2 robot loads per day into the SRB using the intermediate disk cache. The time consuming process was reading data from the worm media. Between 5 and 6 hours was usually required to complete a batch. Operator time was about 20-30 minutes per batch to load/unload the robot including recording of disk labels.

Once the images were in the disk cache, SRB bulk operations were used to rapidly register the master images and all generated log files. Registration consists of the creation of a logical file name for each digital entity within the MCAT catalog. The images were then stored under SRB management control by making a copy in a SRB vault. Since all data written to a SRB vault is owned by the SRB data management system, access controls are managed by the MCAT catalog for each image, to enforce preservation policies. Each of these processing steps generated preservation metadata for each digital entity. In the SRB data grid, the preservation context is managed as metadata attributes on the logical name. The administrative metadata maintained as part of the preservation context included the mapping from the logical file name to the physical file name and storage device where the SRB-managed image was stored, the access controls, audit trails to record accesses on the images, and checksums. Ingestion from temporary storage into the SRB was limited by the speed of the grid bricks used to implement the SRB vault, and the temporary storage media.

The layout of the objects within the SRB preserved the original organization of the WORM media. If needed, we would be able to reconstruct the collection in its original form.

Ingestion of thumbnails

Ingestion of the thumbnail and access images was manageable from any Linux workstation with access to the UMIACS grid brick. A PERL script was written to extract the logical disk label of the CDROM, gather directory listings from the CD, and prompt for any information on the physical CD label. The script read directly from the CD and loaded the thumbnail and access images into the SRB without any temporary storage. The logical name space in the SRB was organized as a collection hierarchy that replicated the worm directory structure, with appropriate directory listings and ingestion log files added to allow reconstruction of the original layout.

Ingestion of metadata

All the metadata supplied by NARA was uploaded into the SRB in the native format that was supplied to UMIACS. The NARA-metadata directory contains copies of the raw metadata. In addition to preserving the raw metadata two efforts were made to assign descriptive metadata to each object to allow for faster retrieval.

The first effort to attach metadata to the ingested objects loaded a complete set of metadata to each image's user-defined metadata fields. Data was mined from the NAIL export and web data. Each metadata item was added to the MCAT catalog as user-defined metadata. In this process, a value/attribute pair is created in the MCAT catalog for each metadata item for each digital entity. This inefficient method was chosen over creating our own schema and using a mechanism such as the DAI interface to ingest metadata so as to maintain compatibility between the schemas at NARA and SDSC. Introducing a new schema would make future replication a non-trivial task and would possibly limit the metadata to certain databases. The long-term solution to this problem is addressed by the Extensible Metadata Catalog, which supports definition of new table structures.

The second attempt at ingesting the metadata took a two-pronged approach. First we attached metadata from the EAP tracking databases to each object in the user-defined MCAT attributes, then we populated a separate set of tables designed to provide a browsing interface to the NAIL export. The second set of tables was created outside of the MCAT. This would allow SRB-only access to search the more limited EAP tracking interfaces in addition to providing a more intuitive browse interface with data from the NAIL export. This effectively stored the administrative metadata in the SRB catalog, and the descriptive metadata in a separate access catalog.

Scripted processes

The scripts for processing the worm media, ingesting thumbnails, and processing metadata used off the shelf components. On windows, ActivePerl was used to write two scripts, one to transfer data from the worm media and the other to record written labels. Since the OptiDriver presented a file system view of the entire worm robot, standard file and directory operations could be used to copy data and gather metadata about the media. A script to process the thumbnails first scanned the CDROM for metadata, and then used the SRB Perl access interface to ingest images and metadata. The scripts used to process the NAIL metadata were divided into two modules. The first script extracted relevant information from the NAIL metadata and the second ingested the result into either an Informix database using the Perl DBI interface, or into the MCAT using the SRB Perl interface.

Manual processes

In the preservation of the EAP collection, a strong attempt was made to automate each task, and minimize the number of manual steps that had to be performed. As noted above, even though scripts were used to manage each stage of the preservation processes, we still were forced to do some manual steps:

- Because we could not build an SRB server to control interactions with the OptiMF2 file system, we manually recorded information written on the WORM media. On modern storage systems, this information may be accessible through calls to the storage repository. Information stored in the structure of the file system was automatically recorded.
- Exception handling. When a media was identified as being unreadable, manual steps were used to identify a viable copy. Once the digital entities are registered into the data grid, the ability to access a replica becomes automatic, as part of the access process, when the first copy is not available.
- Descriptive metadata. The review of the descriptive metadata requires either an expert to decide whether the semantics are relevant, or an expert system that is able to apply semantic clustering across the terms that are used to determine relevance.

The amount of time required for ingesting the collection was under an hour a day during the data load. This includes loading the robot, recording disk labels, and loading CDROMs. The only process that could not be shortened is recording physical disk labels. Given a larger robot and possibly a CD stacker, it would have been possible to completely automate the ingestion process.

Preparation time took about two weeks after the necessary components were identified. This included writing the scripts and assembling the ingestion pipeline.

Collection Presentation

Three separate collection browsers were developed to show different levels of collection abstraction. The first browser presented a view of the data arranged identical to its arrangement in the SRB. This was useful for visually verifying the media ingestion and ensuring its layout within the archive was correct. The second interface presented a batch view of the data. NARA processed the data in 44 separate batches with each image being assigned an identifier within its batch. This interface allows rapid access to individual images if its batch and identifier were known. The third interface organized the data into collection and record group hierarchies defined by the NAIL database. This allowed for content based browsing of the collection.

4. Prototype Persistent Archive

Once the EAP image collection had been loaded into the SRB data grid, it then became possible to demonstrate multiple preservation scenarios. The SRB data grid provides data management capabilities for the following tasks:

- Migration of a collection onto another type of storage repository. An example is creating a replica of each digital entity in the IBM High Performance Storage System, in addition to maintaining a copy on a Grid Brick commodity disk system.
- Migration of a collection to another site to minimize risk of loss due to disasters. A copy of the EAP image collection was made on the NARA grid brick and on the grid brick at SDSC.
- Federation of independent authoritative catalogs of the preservation metadata. As part of the development effort, a new release, version 3.0, of the Storage Resource Broker was developed that supports the interchange of metadata between independent MCAT catalogs. This makes it possible to replicate not only the image content, but also the image context between two sites. Version 3.0 was released on October 1, 2003, and will be used to replicate the MCAT catalogs.
- Preservation of authenticity through management of the consistency between the image content and image context. Since each archival process generates preservation metadata, a major challenge in authenticity is being able to prove that the image context correctly reflects all operations that have been performed upon the image content. The SRB data grid technology explicitly manages as metadata attributes the state generated by operations on digital entities. The goal is to show that the consistency mechanisms are adequate to assert authenticity.
- Automation of archival processes. The EAP collection is accessible through the SRB data grid by the access mechanisms used to automate processing. The traditional interfaces provided by the SRB are C library calls to enable manipulation of the registered digital entities by applications, shell commands to enable processing from scripts, as well as web-based interfaces for

interactive manipulation. The University of Maryland developed as part of this project a Perl interface to the SRB to enable processing preservation tasks from Perl scripts. The Perl interface was implemented as a "swing" library on top of the C library calls.

5. Summary

The automation of archival processes is one of the essential capabilities that are needed to ensure the preservation of massive data collections. Without the ability to process hundreds of thousands to millions of digital entities, it will not be possible to preserve the digital content that is being generated. By working with the EAP collection, comprising three sets of 124,000 image files (master, access, and thumbnail), associated descriptive and technical metadata, and 1.2 TeraBytes of data, we demonstrated that it is possible to use scripts to control the processing of an entire collection. As an outcome of the project, the entire collection became accessible, both the high-resolution images as well as the thumbnails that were originally accessed through a web interface. The project also enabled the replication of the collection onto multiple types of storage repositories, as well as across multiple sites, to ensure that a disaster at any one site would not put the collection at risk.

6. References

1. NARA Persistent Archives, R. Moore, R. Marciano, R. Rajasekar, M. Wan, J. JaJa, and P. Berkman, NPACI Technical Report, June 2003.
2. R. Moore, A. Merzky, "Persistent Archive Concepts", Global Grid Forum Persistent Archive Research Group, Global Grid Forum 8, June 26, 2003.
3. Arcot Rajasekar, Michael Wan, Reagan Moore, George Kremenek, Tom Guptil, "Data Grids, Collections, and Grid Bricks", Proceedings of the 20th IEEE Symposium on Mass Storage Systems and Eleventh Goddard Conference on Mass Storage Systems and Technologies, San Diego, April 2003.
4. R. Moore, "Common Consistency Requirements for Data Grids, Digital Libraries, and Persistent Archives", Grid Protocol Architecture Research Group, Global Grid Forum, Tokyo, Japan, March 5, 2003.
5. R. Moore, C. Baru, "Virtualization Services for Data Grids", Book chapter in "Grid Computing: Making the Global Infrastructure a Reality", John Wiley & Sons Ltd, 2003.
6. R. Moore, "The San Diego Project: Persistent Objects", Proceedings of the Workshop on XML as a Preservation Language, Urbino, Italy, October 2002.
7. MySRB & SRB - Components of a Data Grid, Arcot Rajasekar, Michael Wan and Reagan Moore, *The 11th International Symposium on High Performance Distributed Computing (HPDC-11)* Edinburgh, Scotland, July 24-26, 2002.
8. R. Moore, "Preservation of Data, Information, and Knowledge," Proceedings of the World Library Summit, Singapore, April 2002

9. Preservation of Digital Data with Self-Validating, Self-Instantiating Knowledge-Based Archive, B. Ludäscher, R. Marciano, R. Moore, *ACM SIGMOD Record*, 30(3), 54-63, 2001.