Office of Information Technologies

IT Architecture Initiative

Digital Asset Management: An Introduction to Key Issues

Prepared by R. David Vernon and Oya Y. Rieger
Revised 2/02
REV 2

Introduction

NOTE: The genesis of this paper was a joint effort of the Cornell University Library and the Office of Information Technology.

As part of the IT Architecture Initiative, the Office of Information Technologies (OIT) is producing a series of papers outlining directions in information technology architecture. In the spirit of RFCs (Request for Comments), the papers are intended to facilitate understanding of and open dialogue about information technology trends at Cornell, with the ultimate goal of improving the use and interoperability of information technology services throughout Cornell.

Synopsis

The goal of this paper is to highlight the critical challenges the Cornell community is facing in managing its rapidly growing "digital assets." The paper's aims are fostering a common understanding of the basic principles of digital asset management and suggesting closer collaboration among several stakeholders to devise scalable, common solutions.

This paper includes:

- The Storage Challenge at Cornell

- Primary Digital Asset Management System Issues

- Action Items

- Closing Thoughts

What is Digital Asset Management?

Digital asset management is the systematic management of digital data, such as text, image, audio, and video files, so that they can be reused and re-purposed. It aims to maximize the value of these assets by facilitating easy storage and retrieval while protecting and, at times, enhancing their utility. A digital asset is any form of salient information that plays a role in an institution's efficiency and effectiveness. Some examples of digital assets are:

- Reports

- Scientific data

- Databases

- Image repositories

- Web sites

- Distance Learning Courses

Some goals of digital asset management are:

- Enable ownership control (rights management) and security.

- Ensure the authenticity and integrity of documents.

- Create reusable content that can support both short- and long-term use.

- Ensure effective management of assets to maximize efficiency, productivity and profitability.

- Protect the integrity of data (storage and transmission requirements).

- Ensure the longevity of data (archiving).

With the increasing dependency on digital information, institutions are recognizing that effective management of these assets is critical. As the definition of digital asset management continues to evolve some use the phrase interchangeably with "persistent archives," "digital preservation" or "digital archiving."[1] However, for the purposes of this paper, the authors interpret digital asset management as the broadly scoped challenge of deploying highly valued repositories for digital information.

The Storage Challenge at Cornell

With the explosive growth in the use of data intensive applications and Internet resources, there is an unprecedented demand for reliable digital archives to store and distribute Cornell's "digital assets." To address this growth within Cornell a growing number of departments are developing services that could be placed under the general rubric "digital libraries." Some worthy of note include:

- Cornell Lab of Ornithology's archive of natural sounds.

- Cornell Computer Science department's digital library research initiatives.

- Cornell University Library's proposed central digital asset management services.

- Courses stored at Cornell Computing Services.

- Site for Science (http://nsdlib.nsdl.cornell.edu/nsdl/portal/index2.html).

---

[1] The main goals of digital preservation are to ensure content (bit stream) integrity, protect technical context, maintain provenance data, enable usability (locate, retrieve, use). Although "digital preservation" and "digital archiving" are sometimes used interchangeably, "archiving" historically has a slightly different meaning. Digital archiving refers to initiatives assumed by institutions with a mandated responsibility to maintain digital information for legal, fiscal, evidential, or historical purposes. In addition, the ultimate goal of a persistent archive is to preserve not only the bits associated with the original data, but also the context that permits the data to be interpreted.

All of these projects focus on a common need: *the secure and reliable [2] storage of diverse forms of digital information that allows "best" use of and access to those assets over time*.

Primary Digital Asset Management System Issues

Implementers of digital collections often struggle with common challenges and questions. These include but are not limited to:

- Selection of hardware and hardware limitations.
- Managing the persistence of assets within the data repository.
- Cataloging (metadata) processes.
- Control and security of digital assets.
- Rights management.
- Content creation.
- Interoperability among digital libraries.
- Long-term funding.

Hardware and Related Server Applications

A broad range of hardware and software tools are used to structure and store digital assets.

Software strategies range from simple "flat file" copies of digital assets structured by the directory tools of a given computer operating system to advanced and specialized databases. These digital library or data-storage applications not only efficiently record data on media, but also often provide advanced cataloging and digital asset management tools. In addition, within commercial application offerings, there is a range of application focuses, for example, some are developed primarily for the management of digital video assets, others for digital images, and others for traditional text repositories. At Cornell, departments currently leverage a mix of these commercial and locally developed applications. [3]

The types of storage hardware most commonly used include high-speed disk drives, magnetic tape, and optical media. These storage devices have varying life expectancies. Digital assets stored on them must be refreshed[4] periodically to assure the information's integrity over time. Though there are not any universally applied standards on media life, most institutions assume a less than 10-year life span for magnetic media, and a less than 100-year life span for high quality optical media. Pragmatically, the life of any storage device is more likely to be influenced by the use environment in which it exists. Often used optical media stored in harsh environments may fail before archived magnetic tapes stored in optimal conditions. Simply stated, no mainstream system in use at Cornell today is a "write once and forget about it forever" storage architecture. *All users who are concerned about persistence of their digital assets should be cognizant of their respective storage system's limitations and requirements for data refreshing.*

---

[2] It should be noted that the term "reliable" also implies "persistent" storage of data. One approach for persistent storage is being developed jointly by the U.S. National Archives and Records Administration and the San Diego Supercomputer Center. The approach focuses on storing the information objects that make up a collection and identifying their metadata attributes and behaviors that can be used to recreate the collection. See also:
R. Moore et al. Collection-Based Persistent Digital Archives, D-Lib Magazine, March 2000, Volume 6 Number 3 [Part 1]
 <http://www.dlib.org/dlib/march00/moore/03moore-pt1.html>; Number 4 [Part 2]
 <http://www.dlib.org/dlib/april00/moore/04moore-pt2.html>.
[3] Need list of applications. Video Bank, etc ... ALSO INCLUDE LIBRARY APPLICATIONS

[4] Refreshing involves copying content from one storage medium to another.

A myriad of architectures based on the above hardware can be used to create digital libraries. Primary among these are:

- RAID (Redundant Arrays of Inexpensive Disk) hard disk systems.

- Robotic magnetic tape-based storage systems.

- Robotic optical disk storage systems (CD / DVD).

- Storage Area Networks (SAN's).

- Combinations of the above hardware systems forming hierarchical data storage systems.

RAID arrays are by far the most commonly used solution. These systems are simple, cost effective, and allow very fast storage and access to digital information. However, the life expectancy of the hard disk drives that comprise RAID arrays is fairly short. Most system administrators expect a functional life of less than 5 years and many plan on less than 3! The ultimate cost of long-term storage of data on only RAID disk systems can be higher than initially assumed.

Tape-based systems are an order of magnitude less expensive (per unit of data stored) than spinning disk systems. In addition, in optimal environments tape life can easily exceed 10 years.[5] The disadvantage of tape-based systems is increased length of time required to access data from the tape when compared to magnetic and optical disk alternatives. To date, the most common deployments of tape-based storage at Cornell are used to back-up disk storage or to archive static data. An example of this type of system is the CIT EZ-Backup service.

Optical media fills a niche between magnetic disk and tape. While it is not clear that optical storage systems have a lower total cost of ownership than disk systems, they offer excellent access speeds when compared to tape storage architectures and far superior media life compared to tape-based or disk-based systems.

Storage area networks are most often groups of disk arrays linked via a dedicated high speed fiber network. Theoretically this storage area "network" acts as a common backplane to integrate and present multiple storage devices to servers and clients attached to the network as a singular storage resource. Pragmatically the complexities of metadata associated with data positioning of bitstreams on multiple disk platters and issues of security have kept true SAN's from gaining large market shares. Many companies, such as IBM and EMC have products that are marked under the general rubric of a SAN resource, but deviate from a "pure" SAN and are more closely related to large integrated RAID disk systems.

Storage Area Networks should not be confused with Common File Systems used by many institutions to share data. Common File Systems are applications running on servers that allow multiple host accesses to a file via an open data network. Examples of Common File Systems Include: Distribute File System (DFS), Andrew File System (AFS), Network File System (NFS), Novell Netware, and Microsoft's file serving suite.

Technology organizations often combine disk, tape, and optical storage systems in order to get the best results for large online and archival storage projects. Hierarchical data storage systems automatically migrate data from one storage media to another to take advantage of a given technology's benefit. For example, highly active data remains on relatively high cost disk, while static data is migrated off to low cost tape. While these systems at first seem to offer the best of all worlds, their complexity and management overhead should not be underestimated! Although there is no clear rule for when it makes sense to deploy a hierarchical storage system, those with storage requirements of greater than 5 TB might want to consider this solution.[6]

---

[5] If tapes are accessed often and the environment is harsh, tape life can be dramatically reduced.

[6] See also: R. Moore et al. Collection-Based Persistent Digital Archives, D-Lib Magazine, March 2000, Volume 6 Number 3 [Part 1]
 <http://www.dlib.org/dlib/march00/moore/03moore-pt1.html> ; Number 4 [Part 2]
 <http://www.dlib.org/dlib/april00/moore/04moore-pt2.html>.

Data Persistence

Ensuring short- and long-term persistence of digital assets requires attention to the applications used to structure the data as well as the storage media. If the original application or utility doesn't exist in the future, even well preserved assets may be useless. National standards bodies are currently promulgating data type standards[7] and projects exist that intend to structure assets into archives that assure future interpretation.[8]

Pragmatically, given the pervasive nature of many data file types fear of lost interpretation tools may not be as valid an issue as was once believed. It is unlikely that future file interpretation will be problematic for digital archives containing assets formatted in standard formats such as ASCII, .gif, .tif, .mpeg, .jpeg, .ps, and .html.

A more likely and insidious problem would be loss of access to the database structure used to organize and "file" the digital assets. Many digital asset storage "databases" used to file assets on tapes, disk and optical media do so in a *proprietary format*. In short, without the vendor's software that placed the given file on a tape or disk, you cannot retrieve the information. This fact bears consideration by anyone trying to architect a long-term digital asset management architecture. Unfortunately, most major corporations, such as IBM and Microsoft to name but two, are unwilling to assure open access to source code[9], and often will not even escrow the application(s) to assure users access in the case of the company's demise. This control over the storage application is one of the reasons the largest data archives in the world use a consortium derived storage application that assures access to needed software and stores data on media in a means that open source code tools can extract.[10]

Metadata

Metadata has become a buzzword in various communities that design, create, and preserve information systems. Expanding the boundaries of educational and cultural institutions, metadata is now seen as one of the essential components of success for any organization that owns a digital resource.

What is Metadata?

> "Without metadata, content is just bits."
> *Fortune*, November 27, 2000

Metadata is a structured description of an information object or a collection of information objects. Some examples of information objects are:

- Books
- Journal articles
- MS Word files
- Student registration databases
- Image files
- Video clips
- Course web sites

---

[7] http://www.diffuse.org/standards.html

[8] http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html

[9] Source code is the human readable form of a given computer application from which the machine-readable form has been generated.

[10] http://www.sdsc.edu/hpss/hpss1.html

Metadata describes a range of characteristics that are pertinent to text, audio, video, image, and multimedia.

The history of metadata goes back to card catalogs familiar from libraries, museums, and archives. In the traditional context, each entry in the card catalog (a 3x5 card) enabled users to locate material by title, author, classification number, or a subject heading – providing both intellectual and physical access to a collection. The advent of the Web and the proliferation of multimedia content have significantly broadened the scope and functionality of metadata. Here are some examples of how people are using metadata to facilitate their work and research:

- Professor Jacob has developed a digital image collection of wheat to support his research project. For each image, he carefully records metadata, including an image ID, wheat name and type, size of crop, time image is created, and location so that he can organize and manage his collection by these information elements.

- Professor Nickel is using the *Family Welfare* statistical database created by one of her Cornell colleagues to support her child poverty case study. She is relying on the technical documentation to interpret raw data found in the database. The technical documentation includes various metadata elements, such as information on how the data were collected and tallied, and a description of each variable included in the study.

- The copyright status of images used in CyberTower[11] study rooms are managed through a rights management database, which documents the copyright clearance process. This metadata is essential for protecting the Cornell Adult University from any copyright violation claims, in addition to securing the faculty's right to the images created by them.

- The Bureau of National Affairs Web site uses HTML Meta Tags[12] so that their Web pages can easily be located. Using metatags, the staff assigns key words to describe the contents of the Web site. These keywords, which serve as metadata, are used by Internet search engines to increase precision.

- BJTV, a local television channel, uses content rating metadata to disclose the nature of its programs. This practice allows families to filter content and block inappropriate material from their children.

Metadata Types

Metadata relays information about content, context, quality, condition, and structure of information objects. The most common categories of metadata are:

- Descriptive

- Structural

- Administrative (technical or preservation)

It is important to note that there is a significant level of overlap among these classes, and that the types continue to evolve responding to changes in the information environment. In addition to these classifications, we see the emergence of specialized metadata sets to support certain purposes. For example, the advent of technology-mediated instruction introduced a new category of metadata called "instructional metadata."

---

[11] Cornell Adult University's CyberTower Web site offers an array of "study rooms," each representing a short course taught by a Cornell faculty member. This subscription-based system targets Cornell alumni.

[12] *HTML Meta Tags (http://WWW.Stars.com/Search/Meta/Tag.html)* are used to specify to Web search engines how to index your document. Metadata for the Web is still in its infancy and continues to evolve. There are not yet any commonly accepted standards. Consequently, the vast majority of Web tools (such as search engines) lack any common infrastructure for specifying or using the properties of Web content.

Descriptive metadata facilitates resource discovery by providing intellectual access points such as title, author or creator, subject terms, and location. This category of metadata greatly contributes to the interoperability of systems by providing common frameworks and data dictionaries. For example, one can search the Cornell Online Library Catalog using a Library of Congress Subject Heading, which is a controlled vocabulary for descriptive metadata.

Examples of Descriptive Metadata[13]

Journal Article from a Web Site

Title: Autobiography of W. J. Stillman
Creator: Stillman, William James
Subject:
Identifier: Atlantic Monthly, v. 85, issue 507 (Jan. 1900), pp. 1-16
Identifier (URI): http://cdl.library.cornell.edu/cgi-bin/moa/moa-cgi?notisid=ABK2934-0085-3
Publisher: Atlantic Monthly
Date (W3C-DTF): 2000-01
Type: text
Format: image/tiff

Student Activities Database

Title: Student Activities Database
Creator: Cornell University, Office of the Dean of Students
Subject (LCSH): Student activities--New York (State)--Ithaca--Databases.
Description: A database locally maintained by the student affairs staff of the Office of the Dean of Students, and networked in that office (Day Hall, B32)
Type: dataset; service
Format: html, xml, pdf
Date: 1998-01

Economics 101 Web site

Title: Economics 101

Creator: Drake, William E. (Professor)
Date (Valid): Spring 2001
Description: Prerequisites: Completion of Economics 100.
Description (TOC): Syllabus; Course readings; Quizzes; Chat room; Library resources
Identifier (URI): http://www.nyed.edu/courseinfo/Ec101
Subject: math; algebra; calculus

---

[13] These descriptive metadata examples are based on Dublin Core Metadata Initiative, which is used to supplement existing methods for searching and indexing Web-based metadata <http://dublincore.org/>.

Structural metadata is used to navigate an information object. It assists in identification and retrieval of structural sections, such as chapters, sections, illustrations, video segments, etc. Structural metadata also identifies relationships among collections. For example, one can note that there are three versions of a certain file, in .GIF, MS Word, and .PDF formats. The role of structural metadata is growing as the current computing capabilities can use this information in an automated way to search, manipulate, and interrelate information objects.

Examples of Structural Metadata

Journal Article from a Web Site

Pages: pp. 34-56
Footnotes: pp. 55-56
Table 1: p. 36
Illustration 1: p. 37

Student Summer Activities Database

Database Fields: student activity title, place, number of students enrolled, counselor, budget for the activity, starting and ending dates.
Relational Fields:
Related Databases:

Economics 101 Web site

Section Titles: syllabus, course readings, quizzes, chat room, library resources
Prerequisites: Economics 100

Administrative metadata supports short- and long-term management of information. A related but broader category, preservation metadata[14] aims to document an information object's lifecycle to facilitate its maintenance and development. Examples of data elements in the administrative metadata category include media specifications, file format, compression ratio, authentication data, security, and maintenance information. Administrative metadata helps to document an information object's use, function, and history. Information on intellectual property rights, such as contractual terms related to the document's use and distribution, is also included in this category.

Examples of Administrative Metadata

Journal Article from a Web Site

File Format: GIF
Compression: LZW
Storage Media: Sun Microsystems – CIT Storage Facility
System Administrator:
Change History: Moved from UNIX to Sun Microsystems in June 2000.
Copyright: Journal is public domain. Rights on digital version belong to Cornell University Library.

Student Activities Database

Access Restrictions: Use of the database is limited to the Student Activities staff.
Database Software: MS Excel Version 2.1 Windows
Location: Day Hall, Equipment Room, G1
System Requirements: Windows NT
System Administrator: Joan Break
Backup Cycle and Method: Daily through EZ-Backup

Economics 101 Web site

System Requirements: CourseInfo Version 1.0
System Administrator: CIT/ATS
Access Limitations: Requires a password and a user ID
Copyright Clearance: Two images required copyright limitation (with a link to the Copyright Database).
Use Rights: Content owned by Professor Drake – permission needs to be granted to for the site or part of it to be incorporated.
Mime Types: HTML files, style sheets, .GIF images, MPEG3 video
Content Updates: 1/5/2001, 3/6/2001, 6/6/2001

Role of Metadata in Digital Asset Management

> "Digital assets are only as good as the metadata describing them."
> *Web Techniques*, April 2001

Digital asset management relies on the existence of descriptive, structural, and administrative metadata to support the following processes:

---

[14] The Open Archival Information System (OAIS) Reference Model initiative aims to articulate the system functionality and components to preserve any type of information over any length of time. This model presents a preservation metadata model with a high level framework for entities, functions, and associated administrative activities <http://ssdoo.gsfc.nasa.gov/nost/isoas>.

- Enable usability and efficient distribution of digital assets to staff and users.

- Facilitate the reliable storage and retrieval of digital assets.

- Ensure authentication of digital content through digital signatures, public key encryption, etc.

- Control unauthorized access to digital content.

- Protect intellectual property rights.

- Ensure longevity of digital assets in the face of rapidly changing technical and organizational infrastructures.

It is too risky to rely on institutional memory for tracking the development and use of information systems. There needs to be a common and reliable location for project teams to record information such as the technical specifications of a system, maintenance procedures, access rights, etc. Digital asset management metadata serves this purpose and assists in keeping track of legal requirements, privacy concerns, or other proprietary interests. This metadata also insures that potential users can make an informed decision about whether data is appropriate for their intended use. With the advent of e-commerce, usage-tracking data (e.g., number of logins) is becoming an integral part of digital asset management metadata.

Metadata Standards and Structures

Metadata standards offer common principles such as data dictionaries (controlled vocabulary / subject headings for a certain discipline) and allow information to be easily and consistently located and managed. For example, they enable interoperability by allowing users to search across a group of image collections using certain access points. Different communities propose, design, and maintain metadata standards to support their specific needs. The main goal of adhering to a national or local standard is to enable interoperability. For example, if a common metadata standard is implemented for Cornell course Web sites, users can search across these sites to locate courses that are related to a certain topic or to find a site that includes audiovisual illustrations. Full potential of metadata is achieved only if an institution adheres to a standard and when the information and retrieval systems take advantage of this framework. A successful example of this vision is demonstrated by the Multimedia Educational Resource for Learning and Online Teaching (MERLOT). Designed primarily for faculty and students in higher education, MERLOT facilitates the sharing of teaching-learning materials by organizing them in a searchable database, which is based on the Instructional Management Systems metadata standard.[15] Equally important to standards is the structure used to input metadata. XML (Extensible Markup Language) is emerging as a universal format. Like HTML, XML uses tags and attributes to record metadata embedded in information resources.

Several groups are engaged in metadata standards development, including:

- *Metadata for Education Group*
  http://www.ukoln.ac.uk/metadata/education/
  The Metadata for Education Group (MEG) serves as an open forum for debating the description and provision of educational resources at all educational levels across the United Kingdom.

- *Dublin Core Metadata Initiative (DCMI)*

  http://www.dublincore.org/
  Identifies a small, simple set of metadata elements that can be used by any community to describe and search a wide range of digital resources. DC, IEEE, and IMS joint effort for the development of instructional metadata.

- *IMS Learning Resource Metadata Specification*

  http://www.imsproject.org/metadata/index.html
  Aims to define the necessary metadata elements to support widespread reuse, discovery and

---

[15] http://www.merlot.org/

sharing of learning materials via the Internet. The goal is to simplify the process of locating learning materials on the Web.

- *Content Standard for Digital Geospatial Metadata (CSDGM)*
  http://www.fgdc.gov/metadata/contstan.html
  The objectives of the standard are to provide a common set of terminology and definitions for the documentation of digital geospatial data.

- *The Digital Object Identifier (DOI®)*

  http://www.doi.org/

It provides a framework for managing intellectual content, for linking customer with content suppliers, for facilitating electronic commerce, and enabling copyright management for all types of media.

Control and Security of Digital Assets

For many digital library deployments it is important to consider access and control issues. While limiting local access to a known and limited group can be trivial, providing only authorized access to a global community, and then maintaining some form of control over the assets once accessed, can be a daunting challenge.

Along with local authentication and authorization strategies, national initiatives such as evolving Public Key Infrastructure services[16] must be understood by any department interested in providing regulated access to a global community. Without such tools to authenticate and in turn authorize access, stored digital assets may easily become accessible to unintended parities. For sensitive or copyrighted information this may constitute a considerable liability for the archive's owner.

In addition to the need to control initial access to the digital asset resource, many owners of the assets will require control or at least future identification of their assets once extracted. To accomplish this, many organizations mark the digital assets with digital "water marks" or digital "finger prints." For example, owners of an image can digitally fingerprint the image to provide irrefutable proof that the image is theirs—a valuable utility if another is trying to take credit for the image. These fingerprints are often "invisible" bit streams imbedded into the larger bit streams of the asset.

Though it is beyond the scope of this paper to provide detail on evolving text, sound, image and digital video fingerprinting and watermarking technologies, it is important to note that the systems are not trivial and require significant command of the evolving technology.

Interoperability of Digital Libraries

Campus organizations that create systems to store digital assets do so in parallel with a larger Cornell, national and global community. In an ideal world tools used to extract information from one of these digital stores would work with another. This can only be accomplished if there is a universally agreed upon open standard and process for the retrieval, aggregation and presentation of assets stored. However, no such universally accepted digital library standard exists, and at Cornell, and throughout the world, there remains a mix of digital storage strategies.

This does not mean that departments at Cornell should ignore the challenge of creating extensible and interoperable architectures that will allow broad access. Minimally, departments should:

- Examine metadata suites that are consistent with peer group schemas.

- Explore resources available from the Cornell University Library.

- Understand research being performed within the Cornell Digital Library Research group.

---

[16] See http://www.cit.cornell.edu/oit/authentication.html

The key is to understand the added value "standard and open" asset management architectures might offer a given department. If value is there, the department should work diligently to implement.

Funding and Institutional Accountability

One final, key element in the deployment and long-term maintenance of a digital asset library is often left unsaid—funding. All too often institutions create repositories for digital assets without addressing the need for long term, immutable funding streams to assure the ongoing operation costs. The best designed storage systems will fail as soon as the funding to replace aging hardware or maintain staff runs out. All departments worried about persistent storage *must* create base funding streams that are assured for the expected life of the archive.

Action Items

There are several areas on which CIT, Cornell University Library, Cornell Digital Library Research Group and other Cornell departments could collaborate that offer value to the Cornell community at large. These are:

- Foster a common understanding of the challenges.

- Foster willingness to work cognizant of, and when appropriate, in concert with other campus digital asset management initiatives.

- Sponsor Digital Asset management forums to facilitate open exchange of information.

- Outline a "best practice" storage process.

- Form a metadata registry where Cornell Metadata strategies could be openly shared to assure best interoperability between Cornell digital repositories.

Best Practice Storage Process

Currently there are no well-articulated guidelines for expected media life, recommended magnetic and optical media storage procedures, data refresh cycles, or minimum cataloging/metadata strategies. While the Cornell Policy office does have a Records Retention Policy[17] this policy ONLY notes the class of information to be stored and the length of time for the respective classes. Principle trust of the current records retention policy is preservation of University activities based on content rather than the storage media. Regardless of format (print, digital, microfilm, etc.), preservation must be ensured if the information is judged to be "permanent" for legal, historical, or financial reasons. All permanent materials do not need to be maintained centrally. The office of origin has the responsibility for retaining records as a part of the existing recording system. It is interesting to note that in some cases university policy currently requires "permanent" data retention! Without complementary policy recommendations for the structure and care of the equipment on which this data is stored, it is unlikely such ambitious data retention goals will be achieved. To this end it is strongly recommended that Cornell expand its data retention policy to include

---

[17] The current records retention policy was created in 1977 and created prior to the massive growth in and decentralized nature of digital information storage at Cornell. As the current language is ambiguous in regards to many issues outlined, its review may be warranted. The University Archivist is responsible for 1) designating which official university records are archival; and 2) affecting the transfer of archival records from the office in which they originated or were received to the University Archives at such times and in the manner and form prescribed by the Archives and subject to the appropriate retention and disposition schedules that are outlined in University Policy 4.7, Retention of University Records. The University Archives is a component of the Library's Division of Rare and Manuscript Collections. University policy does not make a distinction between paper-based and electronic records.

technical and metadata requirements as well. A good example of a broader data retention policy is the current National Archives and Records Administration draft recommendations.[18]

Metadata Registry

While there is little chance that a single campus-wide metadata "standard" could ever be achieved, there is value in noting the different schemas departments devise. Creating a central location where these targeted strategies are noted, provides an opportunity to recommend changes that may improve interoperable access to the distributed digital asset stores. Even in those cases where the nature of the digital library requires esoteric extensions to a core metadata set, information could more readily be exchanged to the community at large on the availability of the extended classification system. The library may well be the logical owner of this registry/education process. The Cornell University Library has a newly formed Metadata Working Group to provide a forum to discuss metadata issues facing the library. The key goal of the group is to exchange information about metadata and its application to all library functions.[19]

Campus-Wide Digital Asset Management Service

Clearly there is a core set of digital assets that have great value to the University. It is also clear that there is a lack of campus wide systematic thinking applied to the long-term storage of these assets. This is not to imply that all data is at risk, only that there is an exponential increase in the number of digital assets being created and the number of departments and individuals responsible for maintaining this information. It may well be that for a certain class of digital assets a central, structured, broad-based digital asset management service may be of great value. This library would be held to the highest standards of persistent storage science and assure all on campus access to a world-class service. While all departments may not require such a system, many are likely to see its advantage when faced with the issues outlined earlier. Currently both the Library and CIT have large, structured digital asset management systems. As has been an ongoing theme in past partnership discussions, it may well make sense for these groups to share expertise, compromise when appropriate, and create a digital asset management system capable of providing a much needed secure and persistent storage service for the campus at large.

Closing Thoughts and Recommendations

Plainly there is a growing demand for and deepening concern over the preservation and control of the growing store of digital assets at Cornell. The old saying: "An ounce of prevention is worth a pound of cure" is certainly applicable to Cornell's digital asset management and storage needs. Cornell does not currently have a universal plan for digital asset management. The first and most important step to rectify this shortcoming is open communication and friendly collaboration across all Cornell departments that are currently engaged in developing digital asset management resources. Fortunately, there is great expertise in this area at Cornell. The Computer Science department is world renowned for its work in digital library application development. The Cornell University Library not only maintains some of the finest online digital image collections in the world, but also has an active preservation research agenda. A unified effort, combined with the world-class expertise available from within the Cornell family will place us well ahead of many of our peer institutions.

While there are significant administrative and financial challenges to face today, failure to address these needs now, regardless of a central or distributed strategy, will inevitably cost the University far more in the future. Simply stated, this is a problem that will not go away. It requires prudent university investment for the foreseeable future. To this end the Cornell University Library and OIT will continue to work in partnership to provide leadership throughout Cornell and champion the development of required services.

---

[18] http://www.univco.cornell.edu/policy/RR.for.html
[19] http://metadata-wg.mannlib.cornell.edu:1702

Original Date of Publication: December 2001