Designing Storage Architectures for Preservation Collections
Embassy Suites Hotel at the Chevy Chase Pavilion, Washington, DC
September 17-18, 2007

Overview: The purpose of the meeting was to foster a discussion among the technical storage community, the digital preservation community and Library of Congress (LC) staff to consider a range of issues related to the physical storage of digital collections. LC convened a meeting last year, and this year is a continuation of that discussion featuring a wider range of participants. The goal of the meeting is to generate conversation and to articulate the needs of the participants, and especially to highlight the needs of the digital preservation community.

Action Items:

- Mary Baker of H.P. Labs will compile a bibliography of recent academic papers regarding file systems and storage activities to be shared with the participants.

Meeting introduction: Martha Anderson, the Acting Director of the National Digital Information Infrastructure and Preservation Program (NDIIPP) at the Library of Congress (LC), set the tone for the meetings. She discussed the variety of communities represented and shared the context of LC's interest in storage, as well as describing the NDIIPP partners and briefly discussing the environment that they are working in. She noted that not all of the LC partners are focusing on the use of storage, but that "storage" was the second-biggest line item after "staff" in a recent survey of the NDIIPP partners. She also noted that you don't have anything if you don't have a place to store the bits.

Technical Challenges facing the Library of Congress: Thomas Youkel *[need his title]* from the Library discussed the current state of the Library and the technical challenges that it faces. He described digital archiving as a relatively new challenge, but the most challenging environment for the Library.

Some of LC's archiving challenges include:

- Ingest process
- Physical and digital Inventory/Content management
- Storage (disk and tape, upgraded once every 1-3 years. LTO, T10-10000 drives)
- Hardware Technology Migration
- Automating workflow

The Library's current storage technology environment:

- Heterogeneous server environment – 200 servers – various Unix and Windows
- Monolithic storage environment – EMC – 650TB and Sun/STK 150TB – 80% used
- Tape environment – IBM Total Storage and Sun/STK SL8500
- Near Line – 40TB CAS 100% used, 500TB tape 60% used
- Retrievable data stored – 950-980TB, passing 1 petabyte shortly
- 30 of the servers have 80% of the data on them

Harvard University Library Repository Storage: Stephen Abrams, the Digital Library Program Manager at the Harvard University Library, presented information on the "Preservation and Access Repository Storage Architecture." He began by discussing current digital preservation issues at

Harvard. He noted that Harvard's primary strategies for preservation center on securing redundancy and heterogeneity, and that their primary challenge is dealing with the scaling issues. He then went on to provide details on Harvard's storage architecture. Harvard's repository is utilized for both preservation and access. They classify any given abstract piece of content as either a Public use assets (U) or an Archival Storage (A) asset.

He noted that the architecture design requires each asset to be located in 3 physical locations, and on at least 2 separate storage mediums. They utilize Sun's SAMQFS virtualization tool and perform all data replication over the network.

Much of the discussion that followed revolved around the issues of auditing and managing the data, and some discussion about the costs of continual data management ("scrubbing").

Library of Congress Hardware and Software: Henry Newman, Chief Technology Officer at Instrumental, Inc., and a consultant to the Library, gave a presentation on the hardware and software issues that the Library faced while preparing its Requests for Proposal for the National Audio-Visual Conservation Center (NAVCC).

He identified migration, data reliability and the complication of the standards process as some of the Library's main concerns.

Migration:
- How to migrate hardware and software?
- How to migrate Formats?
  Lack of integration between software layers.
- How to change one component or move to a new system?

Data Reliability:
- Silent Data Corruption
- Per-file error detection
- The need for Proactive error management

Hardware Vendor Presentations:

Cisco: Rajeev Bhardwaj, the Director, Product Management, Data Center Business Unit (DCBU) of Cisco gave a presentation that addressed the role of Storage Area Networks (SANs) in storage architectures for digital collection preservation. He discussed the benefits of a SAN in terms of its:
- Costs of long-term storage, including media and facilities
- Data integrity
- Migration to new technologies
- Security

LSI: Bill Delaney and Mohamad El Batal of the Software Solutions and HW/Systems Architecture Team at LSI gave a presentation on current directions in the storage industry, focusing on:
- Data Reliability/Integrity
- Data Storage Security (In-Flight and At-Rest)
- Power Efficient Storage
- Solid State Drive (SSD) storage

Seagate: Dave Anderson, Director of Strategic Planning, Seagate Technology, began by discussing some trends concerning disc drive capacity. He also discussed the concept of "stiction." In the context of hard disk drives, stiction refers to the tendency of read/write heads to stick to the platters, preventing the disk from spinning up and possibly causing physical damage to the media. Some hard drives avoid the problem by not resting the heads on the recording surfaces. This problem happens when the discs spin down and power up again.

He also discussed object-based storage devices (OSD), which are designed specifically for the archive market. OSD is designed to put space management technology in the drive itself, which allows it to self-mirror critical data, and distinguish data from unused space. OSDs allow users to put access control and encryption at the individual object level.

One participant noted that excess functionality on drive itself may inhibit migration to introduction of new technologies.

EMC: Kem Clawson, Chief Technology Officer, Federal Systems Division, EMC Corporation discussed emerging storage services. He identified the need for systems and processes that allow users to adapt, focusing on technology life cycle management. He also noted that he believes that storage services will emerge as utilities, with a primary example being Amazon's S3 storage service.

One participant noted that the problem with the current storage services is that they won't accept liability for the data, and that if they did, they'd be out of business. Clawson suggested that those policies would continue to evolve in the future as the industry became more robust.

Sun Microsystems: Robert M. Raymond from the Tape Drive Engineering Storage Group at Sun Microsystems discussed the benefits of tape storage. He noted that for long-term preservation purposes he was most concerned with developments in holographic storage. He noted that holographic storage has had a 40 year development history with no major commercial products yet, and while it offers a promising archive life, there is little field history on its reliability, so real failure mechanisms are unknown.

Summary of general meeting themes: Mike Handy of the Library next summarized some of the general themes running through the meeting to this point. He noted that last year's program could have been characterized at the highest level as the "disc vs. tape meeting," and while that was very useful for last year's participants, he was hearing different things at this meeting. He noted how the Library is embarking on a continual process of migration after building a large archival system, and that at this year's meeting he was hearing about some of the complexities of actually undertaking that challenge. He also noted that he had not expected storage characteristics to be represented in terms of their carbon footprint.

He also presented the preliminary results of a questionnaire that had been distributed to the participants prior to the meeting. The questionnaire asked the participants take a few minutes to rate a series of questions on a scale of 1 to 5 to help the conveners frame the discussion towards questions of interest.

Questions receiving high responses:

- How does the storage software keep track of what is stored?
- Does software or hardware do any of these things: checksums, import/export, rules management, inventory?

- What happens if the software itself corrupts the data?
- Can I make more than one copy at a time? How do I know that the copies are identical?
- Are my files changed in any way when they are stored? Will they change when I retrieve them?
- What happens when your primary vendor fails?
- What are other people doing?

Open Discussion: The presentation of these questions prompted a wide-ranging discussion about the challenges of establishing and maintaining archival storage systems. Some observations from the participants:

- Open systems should be driven at the client level through virtualization.
- Interfaces and technologies will change over time and need to be migrated. How are we going to do rolling procurements and migrations going forward?
- Automate planning before something is going to be obsolete so you can do the migration before that happens.
- Too much burden on the vendors in terms of backward compatibility and backwards testing. When you make decisions, value companies that provide investment protection for the materials.
- Where do the vendors see much of the machine intelligence residing in the future?
- Standards process has become fractured. Government has to get involved in rounding up the standards process.
- When you try to put systems together you see a replication of functionality in many of these components, and latency in a system in cumulative.
- Cultural heritage community is interested in building the simplest, most redundant system possible, with less complexity in all the pieces.
- Where does the intelligence reside? It depends on the function you're trying to provide. We get lots of requests to add intelligence. If we add intelligence, does it add heterogeneity? If I add heterogeneity it helps my customers.
- The Library doesn’t have any classified data, so we're not going to encrypt anything. We're also not going to compress data either. The application side is doing compression, but we're not doing any storage compression.
- Most of the archives you're dealing with are not going to be dealing with legacy hardware. Three years from now, somebody is going to supply me with larger tape drives, and I'm going to want to migrate my old data on tape.
- We would like to know that nothing will stand between us and the user 100 years from now that was created by the systems that we use. The data has no barriers, and encryption is a barrier. Not all compression is bad, but compression can be a barrier.
- From the key management perspective, some people need access control. How are you going to put that into the framework for system management that we live with today, and are there changes that need to happen at the Open System and POSIX level?
- Is there an off switch for the encryption? Many people want encryption, but for archival purposes, we don't want it.
- If you have a system featuring one command that wipes data, that’s not suitable for an archive device.
- We're hearing that there are some archival storage systems that don't need encryption, but there are some that need it as a core requirement. Address the problem in a generic way. Do you really want to be solving the problem in your own way so you can't leverage what has been done by others? It's there but you don't have to use it.
- Archives have a lower rate of retrieval than for an online system. Our goal is preservation of our

assets.

- Any single asset is probably going to sit there forever.
- Most of the archival systems are not required to handle huge use spikes. Want a way to get things out of the archive and into the access stream when necessary.
- Back to the question of distributed function and more devices getting more intelligence. I've seen many systems developing problems for different reasons. Firmware/microcode seem to have the most problems. We don't know why data becomes corrupt (cosmic rays/hand of God). Are we really getting better at writing and debugging this microcode?
- Are we going to be able to detect errors with the firmware, having firmware do checking throughout the whole datapath?
- By increasing intelligence, you have the ability to check the data more regularly, and you have something that can check things at each point.
- Archiving is actually a very small part of the market. We're going to have to build our systems out of cheap mass-market components. We're going to get the components that the mass-market wants, and we're going to have to deal with them.
- I'd like every item in the chain to do a lot more checking. File systems ignore what the discs are telling them what went wrong. Not having that be the case would be a big step forward. The application that is doing the archiving need not trust the devices underneath.

Introduction of NDIIPP Partners: After a break, the NDIIPP partners in attendance introduced themselves and briefly described their work.

Richard Moore, San Diego Supercomputer Center (SDSC): In the process of completing a pilot project to develop some procedures and ways of working with the LC to establish trust as a third-party repository. We've used a couple of different collections, one static, one dynamic. Storage is one of the simplest issues. How do we provide trusted access to archivists? What are the checksum and verification procedures?

Larry Carver, University of California, Santa Barbara, National Geospatial Digital Archive (NGDA): Collecting and sharing digital mapping data, any kind of object that has a coordinate on the earth's surface. Preserve it and link the various archives together with our partners at Stanford and UTENN, and to provide access through spatial searching.

Terry Moore, University of Tennessee: Testing different storage infrastructures for the NGDA.

Mike Smorul, University of Maryland: Tools for NDIIPP, integrity checking and the ingest part. What type of metadata do you need to package...trusted way to audit large numbers of files, and provide a way for the files to be audited by a third-party without any materials...

Jim Kutzner, Public Broadcasting System (PBS): Public television stations exploring how to design an archive for long-term preservation of public television programs being produced in a ‘born-digital’ environment.

David Rosenthal, Stanford University Lots of Copies Keep Stuff Safe (LOCKSS) project: We build tools that allow libraries to build collections of items they find on the web. Government documents, theses and dissertations. Allow libraries to use relatively unreliable technology to take advantage of libraries cooperating with each other.

Helen Tibbo, University of North Carolina VidArch project: Preservation of contexts around video. ACM and NASA collections, now collecting video from YouTube, especially those related to the 2008 elections. Looking at appraisal techniques to see what intellectual material would need to be preserved to explain videos so that they make sense.

Jonathan Crabtree, University of North Carolina, DataPASS project: Quantitative data, formats, migration. Exploring ways to distribute the storage.

Lucy Lowell, National Science Foundation: Speaking on behalf of the Cyberinfrastructure program at NSF, who are looking to make a large investment in data archives.

Taylor Surface, Online Computer Library Center (OCLC): Web Archives Workbench, a toolkit that provides machine assistance for archiving government publications. “Hub and spoke” model of moving data from one repository to another.

Evan Owens, Portico: Published scholarly information, scholarly journals. Ten million articles currently under contract. Primarily interested in ingest and interchange, moving from publishers systems to our system.

Martin Halbert, Emory University MetaArchive project: Inter-institutional, low-cost digital preservation strategies for digital archives. Technology needs around a low-cost environment.

Keith Johnson, Stanford University: Involved in the Stanford digital repository and the NGDA project.

Patricia Cruse, California Digital Library: Building a web archiving service that will enable libraries to build collections of web content. Crawling the dot.gov domain for the last four years, have lots of data. Also preserving other assets coming out of the university system.

Ralph McEldwoney, National Computer Security Center: U.S. government organization within the National Security Agency (NSA) that evaluates computing equipment for high security applications to ensure that facilities processing classified or other sensitive material are using trusted computer systems and components. 10,000 researchers around the country who use resources primarily for unclassified research. Hold 4 petabytes of data for our program. We project that we'll be up to 10 petabytes. Have to store, similar to the library. We don't have to store this data forever, but are interested in many of the same issues. Storage initiative project. HPC for parallel file systems. Encryption of data at rest. Lifecycle management tools. Undetected bit errors.

Stephen Abrams, Harvard University JSTOR-Harvard Object Validation Environment (JHOVE) project: Framework for doing format-specific identification, validation, and characterization.

Open Discussion: General Discussion that followed touched on the following points:

- Format migration varies tremendously by what format you're looking at. Problems of graphics are different from the problems of texts. Publishers and vendors spend lots of time improving images for publication. Automated migration will give you a mediocre result, though you can optimize a bit, depending on how important the materials are.
- "Early mass migration" was done on items at risk of obsolescence. There is a certain advantages to doing late migration or migration on demand.

- Formats are not changing that fast now, and it may be more on the order of 20-25 years vs. 5 years.
- A question we should ask is where responsibility for that migration lays. For mass-media forms it makes sense for an archive to get in that business. For scientific (or other esoteric) data, it should clearly be the responsibility of their community.
- The efficient way is to migrate formats and technology at the same time.
- Strong case for not migrating formats ever. Most of this stuff will never be archived, so building it into a form that it will make it archival is a waste of money.
- Can't imagine moving hardware and software at the same time. Allow the user to dynamically migrate something.
- We don't want to touch those bits and move them to some other form.
- There is also the problem when you have so much data that you can't think about migration. Mostly we just want to know that it's the same data.
- Technical metadata stored with the files themselves. Main way to handle it is to store that data with the archive.
- Most published content now is "published" content, it has to have open source renderers. They don't do a perfect job, but they do an adequate job of rendering all the formats in use for publishing. Source level backwards compatibility is also strongly enforced. From a preservation standpoint, everything is stored in sourcecode control systems, so you can reconstruct items.
- The burden of maintaining backwards compatibility is a burden on vendors from an economic perspective.
- Depends on what kind of formats you're talking about. Not that hard to read an old graphics program. Word processing is a little more difficult, but still exists.
- The notion of maintaining a renderer seems like a powerful notion. Where does that leave you in being able to convert that data to some more generic format that could be used by some the newest application?
- A change in open source software means that you might have to rebuild your entire linux kernel.
- If you consider the storage media as a format, we should also consider open source versions of the file system Seems essential that we have the ability to recreate the file system in the future.
- Open formats of what we write have been very successful. We're considering obsoleting our very first on-disc format for SAMQFS, and the problem is maintaining the backwards compatibility. The question is if anybody is still using it.

End of Day One.

Day Two, Tuesday September 18, 2007

Review of Day One: Mike Handy from the Library opened the proceedings by presenting a set of observations from the previous day:

- He noted the efforts on the part of the vendors to put intelligence into every component, and to provide a tool set to allow users to manage everything across the network. He found it slightly problematic to embed intelligence at a level that cannot be observed or accessed, noting that it might make it harder to pinpoint problems with so many "intelligent" components.
- It appeared that only oddball items would provide an intractable problem for format migration.
- As far as costs go, he noted that it was "cheap to make a baby, expensive to raise one."
- He had more questions for the participants regarding authentication: what are the mechanisms for determining whether someone has altered the data?

Software Vendor Presentations:

EMC: David Warner, the Senior Solution Specialist, The Americas, Records Management Technologies, Content Management & Archiving Division, discussed the challenges of enterprise content management. He described *enterprise archiving* as a common set of middleware services for collecting, classifying, retaining, migrating, securing and discovering all types of information. When using a federated approach, the content lives in the production systems and gets managed "in place" by a number of common services, including classification, search, policy enforcement, and disposition.

He was most concerned about technology processes related to Federated Records Management (RM), including the immutability of external content, the ability to enforce retention/holds, and the continuous monitoring of external content.

Sun Microsystems: Rick Matthews, the Senior Staff Engineer for Archiving, Backup and Storage Software provided some observations on how software can help secure the reliable and economical retention of data. He noted that an effective retention solution would be affordable; distribute multiple copies of stored data to increase reliability; and separate reliability metadata.

He noted that all systems will change in the future, so users need to factor in change when designing and working with systems. His concerns centered around standards, especially the need to track format standards, and the risk that proprietary format standards would somehow endanger data retention.

IBM: Michael Factor from the IBM Haifa Research Lab discussed his current research on Preservation DataStores. He noted that he had little concern for the preservation of the bits, but had quite a bit of concern about all technologies related to the obsolescence of formats and software. The Preservation DataStores research (part of the European Union's CASPAR project) is:

- OAIS-based
- Independent of the underlying physical storage layer (tape, disk, etc.)
- Generic, independent of the type of stored data
- Scaleable (e.g. global namespace)
- Offloading functionality to the storage layer

He described it as a new storage paradigm that would:

- Physically co-locate the information object (AIP)
- Execute data intensive functions at the storage component (fixity computations and validation, data transformation)
- Handle provenance events internally (have the storage keep track of whatever happens to the data)
- Support the loading and execution of external transformations
- Maintain referential integrity (update links during migration)
- Ensure readability of the data by a different system in the future (global self-described formats)
- Support media migration
- Support a graceful loss of data (amount of information lost is proportional to the amount of bits lost)

Hitachi Data Systems: Andres Rodriguez, the Chief Technology Officer of File Services for Hitachi Data Systems discussed cluster security in terms of encryption for data at rest in systems designed for long time storage of archival data, largely focusing on a simplified key management scheme that leverages a distributed storage architecture. He provided an overview of several storage encryption regimes, and discussed some of the difficulties with key management. The Hitachi Content Archive Platform uses the cluster itself as a way to do shared key management by dividing the keys across the entire cluster (which he called "secret sharing"). He then described some best practices for secret sharing and encryption under this system.

Open Discussion: General Discussion that followed touched on the following points:

- The T10 technical group of the International Committee for Information Technology Standards (INCITS INCITS - INternational Committee for Information Technology Standards) defined an open protection standard which adds protection information to each data block, allowing multiple check points along the data path. What would the software guys do to make this accessible from the software?
- If you have T10 DIFS, does that mean that you have to do checksums anyway? It's a diagnostic tool, not a replacement for the application doing the data integrity checking itself.
- What if you're getting this problem in the iNODE, or the file system superblock, places that have nothing to do with your data? Can't do checksums on superblocks.
- This allows you to determine what part of the data is gone. If it's lost on one sector, the checksum means nothing.
- How can we not afford to do a full end-to-end check of your full system? We should not rely on the diagnostic tool to do all the work.
- Preservation archive libraries are not in the mainstream when it comes to some of these issues. The top security issue for archive libraries is the authentication of the data. Encryption is often seen as a "needless encumbrance."
- The problem is when you have a doubt in the future about the authenticity of the data, which is why you need a complete and comprehensive audit trail. The challenge is asserting the audit trail after you move data from medium to medium.
- What society needs from the Library is a tamper-evident record of history.
- If you need to do format migration for access, that's fine.
- Instead of "authentication," we would use the term fixity. Depends on where you're getting your content from. Only one publisher we deal with is currently capable of doing a firm handshake about their data. Most of the people we deal with can't do anything that hard.
- The British Library uses a secure timeserver as part of their authentication process. Something external that will validate the process.
- Software people think that the hardware solutions are reliable and that their software is reliable, but it is all dependent on the metadata and the policies above the hardware and software levels.

Threat Model Presentation: David S. H. Rosenthal, chief scientist of the LOCKSS (Lots of Copies Keep Stuff Safe) Program at the Stanford University Libraries, gave a presentation on "Engineering Issues in Preserving Large Digital Collections."

He noted the need to start designing archiving software that has a clear view of the threats to IT systems. He described insider abuse is the clearest single threat. He pointed out how it was impossible to deliver a tamper-proof record, especially when the government is an insider in the system. He presented a black box experiment that hypothesized about creating a system that would preserve a petabyte of data for a century with a 50% chance that every bit will survive, which he imagined as the type of system that the LC would want to build. He imagined the bits inside the box like isotopes, undergoing a small process of decay. Suppose the Library wants to fund an independent test lab to see if the servers are this good? They could build an exabyte system and watch it for a year, and they might see 5 bit flips in a year. He proposed that even if systems could be built that were reliable enough to meet this target, they probably couldn’t be affordably built or monitored.

He proposed that systems be designed with a clear view of the threats. He proposed that random decay of the bits is not that big of a threat. Under the old threat model, these were perceived to be the most important threats:

- Media failure
- Hardware failure
- Software failure
- Network failure
- Obsolescence
- Natural Disaster

He identified items that he perceived to be even more important threats:

- Operator error
- External Attack
- Insider Attack
- Economic Failure
- Organization Failure

He identified some rules of thumb for managing threats, noting that you can get safer data but higher cost from:

- More replicas
- More independent replicas
- More frequent audits of replicas

The implications of this are to:

- Make replicas as cheap as possible
- Make replicas more independent, greatest possible heterogeneity.
- Make auditing much cheaper and less intrusive.

He noted that “operator error” was the most likely of all the threats, but also the one that was the most underreported. He also identified insider abuse as another significant threat. He suggested that these problems might be mitigated to some extent by a no-fault reporting system similar to the Aviation Safety Reporting System (http://asrs.arc.nasa.gov/). He then described some of the capabilities and benefits of the LOCKSS archiving system that he is working on.

Open Discussion: General Discussion that followed touched on the following points:

- The software itself is also a correlation. When you design your auditing software, you have to assume that other things are happening on your system.
- Need engineering knowledge to ascertain the minimum number of copies you need to get the right amount of safety. There are rules of thumb, but you could over-engineer the system.
- If you don't have an economic model, you'll end up over-engineering.
- Should be building systems that are more resistant to operator failure.
- More copies, correlation coefficients. How do you model that? Collect anonymous data, even if it’s in a modest database that attempts to gather that information.
- Consumer drives are cheap because they manufacture them in enormous volumes, which sets up correlations between bad firmware, etc.

- Reliability that people expect is unbelievably challenging.
- Are there different levels of reliability based on the economics?
- How are you going to know what's going to be important in 50 years?
- In IT, everything is gold and platinum, all costing the same.
- Costs vs. value, moved away from an a priori determination of value. Through curatorial activity selectors have always made decisions. There is a difference between general collections and special collections.
- Amazon S3 and similar services don't accept responsibility for the data. But as a component, that's ok. They've established the cost of basic bit storage, and you can go and buy it from the market after establishing the price.
- Storage tier is not our responsibility as archivists. Should be able to send the data to S3-like entities that will be offered as services.
- Because it's difficult to evaluate the services, these services might be taken over by Potemkin services.
- How to create the financial incentives to allow them to be more audited, similar to health inspections of kitchens?

Disk and Tape Storage Cost Models Presentation: Richard Moore, the Division Director for Production Services at the San Diego Supercomputer Center (SDSC) gave a presentation on a recent report they have prepared on their realistic cost estimates and projections for an at-scale provider to "store bits." He began by providing some caveats about the details of their report:

- They are dealing with sustainable storage (annual cost w/ media/technology refresh & data migration, not write-once and put on a shelf)
- Based on SDSC experience only (include UCSD's indirect costs – will vary by institution)
- Based on SATA disk and enterprise-class tape systems
- Cannot be specific about vendor costs or burdening, but relative fractions are reasonable
- This is a snapshot as of Jan 2007 - will decline w/ time
- Paper focuses only on single-copy "bit storage" costs
- "Bit storage" is only a fraction of the cost to preserve data.

He then discussed SDSC's three-stage (ingest/storage/use) model for a digital preservation environment, while presenting a brief description of the SDSC storage infrastructure. He noted that SDSC's archive shows exponential growth with a consistent doubling period of approximately 15 months. He noted that the price of storage is getting to be a significant cost element that they really need to constrain.

SDSC's Cost Estimates Include:

- Annualized capital costs of the media (including disk controllers)
- Other annualized capital costs
    - Disk: File system servers, SAN gear
    - Archive: Silos, tape drives, disk cache, file system servers
- Hardware maintenance and software licenses (annual)
- Facilities costs – space, utilities (annual)
- Labor to maintain & administer systems, migrate data (annual)
    - Disk: 3 FTE's to administer disk storage & SAN
    - Archive: 3 FTE's to administer archival systems
- Annual costs normalized by:

    - Total SATA disk deployed (~1.8 PB SATA)
    - Current volume of data stored on tape (~5 PB)
- Sustainable rate - $/TB/year
    - Assumed to be long-term storage w/ migration costs

They depreciate tape cartridges roughly every other cycle, but he noted that the media costs are not the dominant costs, as additional capital infrastructure is required. They have 3 full-time people to administer the disc storage.

One question they were looking to answer in their report was to determine how costs scale with the size of the storage infrastructure:

- Economies of scale are significant as one moves up to "at-scale" [large] installations ($/TB/yr decreases)
- Once infrastructure is "at-scale," economies of scale slow down and the cost ($/TB/yr) levels off with installation size
- A portion of the cost elements (software licenses) are fixed with installation size => decreasing $/TB/yr for these elements
- With large installations, the net $/TB/yr will level off and then slowly decline
- Weak economies of scale in people and servers
- Software license fees go down as scale goes up.
- Media costs are ~30% of the integrated 'bit storage' costs and total capital is ~50% of costs for both tape and disk

He noted that they continue to expect storage costs to go down, and that if the costs continue to scale, they should be fine. The area they need to concentrate on the most are the items that will not scale.

He then discussed some new technologies in storage, including MAID (Massive Array of Idle Discs), which he said had a capitol cost comparable to disc. MAID has lower operations cost and extended useful lifetimes. He also expected the costs of disc and tape storage to eventually converge. He briefly discussed some of the emerging commercial storage environments such as Amazon's S3 service, suggesting that SDSC's costs were similar to those of the commercial providers.

During the question period, he noted that migration costs are factored in the annualizing of all the media. He noted that they do migration all the time, moving from one disc to another on the file system.

FACIT storage architecture presentation: Terry Moore, the Associate Director of the Logistical Computing and Internetworking Laboratory and the Center for Information Technology Research at the Computer Science Department at University of Tennessee, presented on "Guiding ideas of the FACIT Storage Architecture." FACIT is the Federated Archive Cyberinfrastructure Testbed, a project UTENN is working on with the University of California, Santa Barbara. The goal is to create a testbed for exploring a different approach for doing archive federation.

He noted that archivists that are managing data for preservation have to deal with the "hand-off" problem:

- Repeated handoffs between institutions
- Repeated migrations across storage media and storage systems
- Repeated migrations across archive systems

Their tests are working towards securing a “handoff process” that can be sustained. He provided an introduction to their logistical networking (LN) research, a “bits are bits” infrastructure for storage.

He noted that data has logistics, and that after it gets big it gets difficult to manage, and that's a logistical problem. Data logistics is the management of the movement and positioning of digital data, and computing resources it requires, in order to enable people to take action at a given time and place to achieve some purpose. He noted that the logistical challenge of data intensive collaborations requires storage to be everywhere.

The LN hypothesis is to build a generic storage stack similar to the generic network stack by standardizing on a common protocol. For networking it has been the Internet Protocol (IP). For the LN people it’s IBP (the internet backplane protocol). Other properties can then build off that design point.

Open Discussion: The general discussion that followed was driven by questions submitted by participants during the course of the meeting:

What tools can be provided to curators that will help them keep the most valuable stuff?
- Tools for archiving the web and building tools that will characterize the content in some way. Build tools to go through massive sets of data.
- Feasible to do natural language analysis, but it’s impossible to replace curatorial assessment.

How can vast amounts of data be indexed?
- Not a problem of indexing per se, the effect of scale on our current practices. Can we afford tens of thousands of curators at scale?
- We have librarians who are trying to catalog our web archives by trying to find ways to use full-text indexing and find the tolerance for their lack of precision. Might be better to just get close enough.
- Tremendous interest in automated indexing, but no clear idea of how to deploy it.
- Different expectations for different sort of stuff. People don't expect the same precision with web sites as with books. The excitement is on searching across corpora.

Are other libraries and archives measuring costs similar to SDSC?
- Can you speculate about scaling down?
- General infrastructure costs, some scale factors like that. The invisible supporting infrastructure.
- Going beyond the costs that you looked at, what does it cost to do the integrity checking, etc? Would it be possible to estimate costs on that?
- Pilot project with LC has looked to develop some procedures.

Can we really get information on failure rates or are we going to have to continue generating our own tools?
- Any answers that one vendor could give you will be false in another environment. I understand the goal behind that question, but not possible to do it in a way that would pertain to any environment.
- Moving the smarts into the device is the right thing to do.
- Are people willing to pay 2-3 times as much to get documentation on the error codes?
- Trying to develop a threat model, it's helpful to get something more/less reliable about the failure modes. In some ways you can get that information, but in others not.
- We look hard at the reliability data for the components, pretty basic math. Components have

embedded firmware all over the place. Regardless of who caused the problem, we can find out where in our system the problem lays so we can report back.
- There are different ways to have hardware fail. If you understand this, you can mitigate your design.
- With our large customers, we'll take them through the drives, we have tested millions of them. We prove to our customers that it meets their reliability criteria.

Is anyone working on storage technology that lasts longer than 3 years?
- Disc drives are designed to last 3 years. Mean Time Before Failure (MTBF) is not guaranteed beyond 3.
- There are parts in this path that will be obsolete in where they are in the technology/standards cycles.
- Business aspects to the question. It would take huge funding in order to come up with a technology to replace magnetic disc and tape.
- We put five year warranties on our products. We don't want to be designing and building every component. It really inhibits our flexibility to look for component suppliers that will guarantee the lifetime of our products.
- We have a class of customers out there that always want to integrate some widget in the product so that they can compete. We have a lot of customers with disc arrays that we still service. Many customers think that 3-5 years is the ideal period because you get upgradability.
- Right answer is to live with the answer. Hardware is going to flow through the system, focus on making that flow as minimally disruptive as possible.

How do you inventory what you have and know what is in your preservation archive?
- One approach is to have an enterprise content management layer manage it. Trying to develop file crawling tools, e-discovery requirements.
- Depends on the "what." Might be the blocks that make up the files. We do keep this information at different layers. Information going up and down the layers, from one inventory and from what shape and form. Segmented in different ways in the stack. Good interfacing between layers is what's important.
- Aggregating all that information into one place for storage purposes is risky. Have to audit that against the copies of the items located in other locations.
- When you have all sorts of groups contributing data, how would you define what's on there?

Describe the difference between a hash and a checksum?
- Take input, run it through algorithm and get output, essentially impossible to take this output and get original input (hash)
- There is some possibility that you take random input and you get the same hash out. SHA-1 is ok against deliberate corruption, MD5 is not.

What does the user community feel is the biggest unanswered problem for the preservation community?
- Sustainable business models. LC thinks that it has a sustainable model, but it may not be enough to get their mission accomplished. Everybody else's model is suspect. Driving the cost down is the only way we can afford to do it.
- Funding a task force to see who can afford to pay.
- A federation of archives will help ameliorate the cost issues. For federations to be formed, it will take a mind-shift by archivists, to allow somebody else to control their materials.

- Richard Moore's numbers were strongly suggestive that you have to be as big as SDSC to get the economies of scale.
- How do we develop into a trusted institution in the digital preservation/storage arena?

There are institutions such as financial institutions that keep their data reliably for 10 years or more. Can we find out what makes their data reliable for their time horizon?

- Valuable to try and tap into their expertise
- Life sciences probably has more of this, has to be portable but secure. They don't need it spinning on the disc drive.
- Banks are better financed.
- Comes down to financing. Life sciences budget for migration costs, etc. in the development of the drug as part of a long-term budget process.
- Storing it on tier one storage, first class, most expensive fibre channel drives.
- Chronic problem in the not-for-profit academic sector, have to pick and choose, find the right overlap. Stole a lot of ideas from the documentation industry. That's what's going to happen in the cultural preservation community. The money is in the corporate sector, and we can use 80% of what they're doing. Difference is understanding what we need to take from those communities.
- Clearly need for tiers. Access density and response time.
- They may have the flexibility to spend more on gear, economic analysis that tells them that they consider it their competitive advantage.
- Email is interesting, but not best match for cultural heritage institutions. I go to lots of digital library meetings, hoping that they could go out and learn about things that the vendors take for granted.

Questions/ideas for the next meeting:

- What are the agendas for us to tackle next year. Time to start getting input from the vendor community.
- Useful to send something out about what the Library is trying to explore with these meetings
- Useful for libraries and archives to have the vendors give a product roadmap. Technology statements, tape roadmap, what the risks are and what the targets are.
- Start the next meeting by having a technology discussion. You're betting the future, and vendors only have so much outlook.