HathiTrust is a Solution

The Foundations of a Disaster Recovery Plan for the Shared Digital Repository

This report serves as recommendations made by
Michael J. Shallcross,
2009 Digital Preservation Intern
University of Michigan
School of Information

Executive Summary

This report seeks to establish the framework of a Disaster Recovery Plan for the HathiTrust Digital Library. While professional best practices and institutional needs have provided a clear mandate for HathiTrust's Disaster Recovery Program, common parlance has often obscured two prominent features of such initiatives. First, a 'Disaster Recovery Plan' is actually comprised of a suite of documents which detail a range of issues, from crisis communications and the continuity of administrative activities to the restoration of hardware and data. Second, there is no conclusion to the planning process; it is instead a continuous cycle of observation, analysis, solution design, implementation, training, testing, and maintenance.

The primary goal of the present document is to provide a foundation on which future planning efforts may build. To that end, it examines the strategies by which HathiTrust has anticipated and mitigated the risks posed by ten common scenarios which could precipitate a disaster:

- Hardware failure and data loss
- Network configuration errors
- External attacks
- Format obsolescence
- Core utility or building failure
- Software failure
- Operator error
- Physical security breach
- Media degradation
- Manmade as well as natural disasters.

As this list reveals, a disaster within the digital repository refers not merely to data loss, the destruction of equipment, or damage to its environment, but to any event which has the potential to cause an extended service outage. For each scenario, the report discusses possible threats, summarizes the potential severity of related events, and then details solutions HathiTrust has enacted through direct quotations from the HathiTrust Web site and TRAC self-assessment, Service Level Agreements, and literature from service providers and vendors. Attached appendices provide relevant information and include contacts for important HathiTrust resources, an annotated guide to Disaster Recovery Planning references, and an overview of key steps in the Disaster Recovery Planning process.

The concluding section of the report provides recommendations and action items for HathiTrust as it proceeds with its Disaster Recovery Initiative. These are divided into Short (0-6 mos.), Intermediate (6-12 mos.) and Long-Term (12+ mos.) objectives and are arranged in a suggested order of accomplishment.

- Short-term goals include:
  - Describing the nature and extent of HathiTrust's insurance coverage
  - Testing and validation of current tape backup procedures
  - Improved physical and intellectual control over system hardware
  - Establishment, distribution, and maintenance of phone trees
  - Increased documentation of institutional knowledge
  - Identification of Disaster Recovery measures in place at the Indianapolis site.
- Intermediate-term objectives focus on:
  - Creation of a Disaster Recovery Planning Committee

        - Initiation of the data collection and analysis essential to the creation of recovery strategies (This section provides a high level break down of various tasks and includes the coordination of activities between the Ann Arbor and Indianapolis sites as well as with service providers and vendors.)
    - Long-term action items deal with:
        - Completion and implementation of the suite of Disaster Recovery documents
        - Initiation of staff training and tests of organizational compliance.
        - Storage of an additional copy of backup tapes at a remote third location
        - Investigation of an alternate hot site in Ann Arbor in the event a disaster renders the MACC unusable
        - Consideration of a third instance of the repository
        - Avoidance of vendor lock-in if a key supplier should go out of business.

This report demonstrates that various risk management strategies, design elements, operating procedures, and support contracts have endowed HathiTrust with the ability to preserve its digital content and continue essential repository functions in the event of a disaster. The establishment of the Indianapolis mirror site, the performance of nightly tape backups to a remote location, and the redundant power and environmental systems of the MACC reflect professional best practices and will enable HathiTrust to weather a wide range of foreseeable events. Unfortunately, disasters often result from the unknown and the unexpected; while the aforementioned strategies are crucial components of a Disaster Recovery Plan, they must be supplemented with additional policies and procedures to ensure that, come what may, HathiTrust will be able to carry on as both an organization and a dedicated service provider.

Acknowledgements

The author would like to thank Shannon Zachary for her encouragement and guidance; Cory Snavely and Jeremy York for their generous expenditure of time, energy, and knowledge; and Nancy McGovern and Lance Stuchell for access to their outstanding Disaster Recovery Planning resources. The following individuals have also been invaluable sources of advice, support, and information: John Wilkin, Bob Campe, Cyndi Mesa, Ann Thomas, John Weise, Larry Wentzel, Lara Unger-Syrigos, Bill Hall, Emily Campbell, Sebastien Korner, Jessica Feeman, Phil Farber, Chris Powell, Cameron Hanover, Stephen Hipkiss, Tim Prettyman, Rene Gobeyn, and Krystal Hall. Thanks also to Dr. Elizabeth Yakel, Magia Krause, and Veronica and Cora Fambrough. The work in this report was made possible by an IMLS Grant.

Table of Contents

****Appendices F – I are embedded PDF files.****

Introduction

In the realm of print libraries, a disaster is a fairly unambiguous event: it is a fire, a broken pipe, an infestation of pests—in short, anything which threatens the continued use and existence of texts or the environment in which they are stored. This basic definition may also be applied to the digital library, in which a disaster refers not merely to the loss of content or corruption of data, the destruction of equipment or damage to its environment, but to any event which has the potential to cause an extended service outage. This last part proves to be the greatest difference between the print and digital worlds because there are a great many threats which can leave data intact but incapacitate the primary functions of a digital library. The daily operation of an institution such as HathiTrust involves the anticipation and resolution of a variety of problems—crashed servers, software bugs, networking errors, etc.—which only rise to the level of a 'disaster' when they exceed the capacity of normal operating procedures and/or the maximum allowable outage periods. Disaster Recovery Planning thus prompts us to develop robust strategies to mitigate and limit the effects of common problems and at the same time forces us to think the unthinkable. Nevertheless, confronting worst-case scenarios is a vital activity; the belief that an event will never happen simply because it has never happened is an invitation to the very disaster we seek to avoid. Herein lies a conundrum, in that the creation of detailed plans for every eventuality is nearly impossible and also impractical, since the results of such an endeavor would be needlessly complex as well as expensive. At its basis, then, Disaster Recovery Planning demands an astute assessment of risk so that we may weigh the costs of preparations and solutions against the costs of a potential event.

So where to begin? When the subject of Disaster Recovery Planning arises, common parlance often obscures two prominent features of such initiatives. First, a 'Disaster Recovery Plan' is actually comprised of a suite of documents which detail a variety of related issues, from crisis communications and the continuity of administrative activities to the recovery of hardware and data and the restoration of core functions. Second, there is no conclusion to the planning process or a point at which a plan is 'done'; there is instead a continuous cycle of observation, analysis, solution design, implementation, training, testing, and maintenance. The essential first step is therefore a thorough knowledge of the organization, its goals, and its mandate for a Disaster Recovery Program so that later efforts can focus on the articulation of policies and the development of solutions. As a preliminary step in this effort, this report looks to establish a basic foundation from which future planning efforts may grow.

- Goals for HathiTrust's Disaster Recovery Program

While a more formal statement of HathiTrust's goals and requirements for its Disaster Recovery Program must be elucidated, the repository's mission statement provides a good indication of its main objective in the formation of a Disaster Recovery Plan. As part of its aim to "contribute to the common good by collecting, organizing, preserving, communicating, and sharing the record of human knowledge," HathiTrust seeks "to help preserve these important human records by creating reliable and accessible electronic representations."[1] This statement clearly joins the twin imperatives of preservation and access with an additional requirement: reliability. The development and implementation of a Disaster Recovery Plan will ensure that digital objects will retain their authenticity and integrity over the long term and that partner libraries and designated users may rely on HathiTrust services (or their timely resumption) and content in the face of catastrophic events.

[1] HathiTrust. "Mission & Goals" (2009) retrieved from http://www.hathitrust.org/mission_goals on 8 July 2009.

- The Mandate for Disaster Recovery Planning in Digital Preservation

HathiTrust's mandate for a comprehensive and proactive Disaster Recovery Plan stems from a number of significant sources, among which we may include its mission and goals. The "Institutional Data Resource Management Policy" (2008) of the University of Michigan's *Standard Practice Guide* also provides an impetus for the creation of a Disaster Recovery Program. While not necessarily inclusive of the Michigan Digitization Project materials stored in HathiTrust, this document underscores how important it is that data resources "be safeguarded [and] protected" and "contingency plans […] be developed and implemented."[2] In its discussion of the latter point, the policy specifies that:

> Disaster Recovery/Business Continuity plans and other methods of responding to an emergency or other occurrences of damage to systems containing *institutional data* […] will be developed, implemented, and maintained. These contingency plans shall include, but are not limited to, data backup, Disaster Recovery, and emergency mode operations procedures. These plans will also address testing of and revision to disaster recovery/business continuity procedures and a criticality analysis.[3]

While data backup procedures and a host of risk management practices are already an integral part of HathiTrust's operation, the repository now looks to formalize the other strategies suggested by the "Institutional Data Management Policy." Beyond the example laid out by this document, HathiTrust's mandate for Disaster Recovery derives from the professional literature detailing best practices in the field of digital preservation. The *Reference Model for an Open Archival Reference System* identifies Disaster Recovery as an essential component of its "Archival Storage" function and highlights the importance of such plans in achieving the goal of long-term preservation of a digital archive's holding. As outlined in the OAIS document, "the Disaster Recovery function provides a mechanism for duplicating the digital contents of the archive collection and storing the duplicate in a physically separate facility."[4] HathiTrust has successfully met this requirement by performing nightly tape backups and establishing a mirror site at Indiana University in Indianapolis. The *Trusted Repositories Audit & Checklist: Criteria and Checklist* (2007) is even more explicit in its requirement that repositories document their policies and procedures with "suitable written disaster preparedness and recovery plan(s), including at least one off-site backup of all preserved information together with an off-site copy of the recovery plan(s)."[5] Professional best practices as well as internal needs and goals thus provide the mandate which underlies HathiTrust's development of a formal Disaster Recovery Plan.

- Disaster Preparedness in the Design and Operation of HathiTrust

One of the primary goals of HathiTrust is to provide "transparency in all of its operations, including its work to comply with digital preservation standards and review processes."[6] Nowhere is this commitment more clear than in its efforts to anticipate and mitigate risks which could threaten the

---

[2] University of Michigan. "Institutional Data Resource Management Policy" (2008) *Standard Practice Guide,* retrieved from http://spg.umich.edu/ on 8 July 2009.
[3] Ibid.
[4] Consultative Committee for Space Data Systems. *Reference Model for an Open Archival Information System* (2002) p. 4-8.
[5] OCLC and CRL. "Section C3.4" *Trusted Repositories Audit & Checklist: Criteria and Checklist* (2007) p. 49.
[6] HathiTrust. "Accountability" (2009) retrieved from http://www.hathitrust.org/accountability on 25 June 2009.

contents and functions of the Shared Digital Repository. As a first step in addressing the disaster preparedness requirement in section C3.4 of the TRAC Criteria and Checklist,[7] this document serves two purposes. First, it provides an overview of the policies, procedures, resources and contracts that enable HathiTrust to address the challenges and threats endemic to the field of digital preservation. Material is therefore cited directly from the HathiTrust Web site (http://www.hathitrust.org), the most recent version of HathiTrust's review of its compliance with the minimum required elements of the TRAC Criteria and Checklist,[8] and relevant literature provided by key vendors and service providers.[9] Second, this report examines HathiTrust's current level of disaster preparedness and defines current and forthcoming efforts in its development of a dynamic and proactive Disaster Recovery Program. Per the recommendations of the TRAC Criteria and Checklist, this document records the measures and precautions already in place in regards to "specific types of disasters" that could befall HathiTrust. These events include hardware failure, data loss, network configuration errors, external attacks, core utility failure, format obsolescence, software failure, physical security breach, and manmade as well as natural disasters. While a formal, written plan detailing individual roles and responsibilities in the repository's response to each of these scenarios is still forthcoming, the evidence gathered in this report reveals that crucial elements of a Disaster Recovery Plan are already in place within HathiTrust.[10]

- Essential HathiTrust Business Functions

As the development of the Disaster Recovery Plan proceeds, it is important to bear in mind that its goal is not merely the restoration of hardware and data but also the recovery and continuity of essential repository functions. The following list represents core functions that need to be addressed by HathiTrust's Disaster Recovery Plan and as such should not be considered a comprehensive representation of the repository's functions. By directing planning efforts toward specific functions (rather than the organization's activities as a whole), HathiTrust may prioritize and focus its recovery responses and resources to ensure that the most essential functions go back online first. Subsequent discussion of Disaster Recovery strategies and risk management solutions in this report are presented under the assumption that the continuity of these functions is a primary objective. The prioritization of these functions remains to be determined by an appropriate authority.[11]

---

[7] "Repository has suitable written disaster preparedness and recovery plan(s), including at least one off-site backup of all preserved information together with an off-site copy of the recovery plan(s). The repository must have a written plan with some approval process for what happens in specific types of disaster (fire, flood, system compromise, etc.) and for who has responsibility for actions. The level of detail in a disaster plan and the specific risks addressed need to be appropriate to the repository's location and service expectations. Fire is an almost universal concern, but earthquakes may not require specific planning at all locations. The disaster plan must, however, deal with unspecified situations that would have specific consequences, such as lack of access to a building." OCLC and CRL. *Trusted Repositories Audit & Checklist: Criteria and Checklist* (2007) p. 49.

[8] *Hathitrust Digital Library Review of Compliance with Trustworthy Repositories Audit & Certification: Criteria and Checklist Minimum Required Elements*, revised May 20, 2009. Available at http://hathitrust.org/documents/trac.pdf

[9] Contact information for relevant University of Michigan departments and service providers as well as for external vendors may be found in Appendix A.

[10] A list of resources related to disaster recovery and the planning process may be found in Appendix D (Annotated List of Disaster Recovery Planning Resources).

[11] This list of essential HathiTrust business functions was developed in conjunction with Jeremy York.

- Ingest
  - Ingest digital objects (SIPs) via GRIN—the Google Return Interface (or a modified ingest portal for local content)
  - Validate ingested content with GROOVE—the Google Return Object-Oriented Validation Environment (or a modified version for localized ingest)
- Archival Storage
  - Preserve indefinitely digital objects and metadata (AIPs) in the Shared Digital Repository (includes ensuring the integrity and authenticity of materials). This function addresses the needs of partner libraries as well as individual users.
  - Record changes to and actions on items while they are in the repository
  - Maintain a persistent object address for items within repository
- Dissemination
  - Provide access to digital objects for users
  - Allow for the text searches through a variety of fields
  - Enable large scale full-text searches
  - Permit the creation of public and private content collections
  - Disseminate digital objects (DIPs) to users (via the page-turner access system and data API)
  - Distribute datasets and HathiTrust APIs to developers
  - Research and develop additional applications and resources for HathiTrust
- Administration
  - Provide transparent and up-to-date information to users and the general public via http://www.hathitrust.org/
  - Communicate information and coordinate activities amongst partner libraries and HathiTrust boards and committees.
- Data Management
  - Update and manage the Rights and GeoIP databases
  - Build and maintain Collection Builder and Large Scale Search Solr indexes
  - Determine appropriate user access to texts via database queries
  - Sync content with the Indianapolis site and backup content to tape

HathiTrust's Disaster Recovery Strategies

- Basic Requirements for Disaster Recovery

Roy Tennant has identified three requisite components of a digital Disaster Recovery Plan: (1) the use of an effective data protection system (i.e. RAID), (2) redundant power and environmental systems, and (3) regular backup of information to tape and, ideally, to a remote mirrored site.[12] HathiTrust has incorporated all these elements into its design and operation. Its Isilon IQ storage cluster provides a high degree of data redundancy with its N+3 parity protection; the Michigan Academic Computing Center provides fully redundant power and environmental systems for HathiTrust infrastructure; and nightly tape backups and the replication of data to a fully operational mirror site located at Indiana University in Indianapolis with the same levels of power and environmental conditioning provide multiple copies as well as geographic distribution of content.

  - "HathiTrust is intended to provide persistent and high availability storage for deposited files. In order to facilitate this, the initiative's technology concentrates on creating a minimum of two synchronized versions of high-availability clustered storage with wide geographic separation (the first two instances of storage will be located in Ann Arbor, MI and Indianapolis, IN), as well as an encrypted tape backup (written to and stored in a separate Ann Arbor facility).

    Each of these storage or tape instances is physically secure (e.g., in a locked cage in a machine room) and only accessible to specified personnel. Each separate storage system is also equipped with mechanisms to provide mirrored management and access functionality, and employ 100% data redundancy in an effort to prevent data loss."[13]

Details on parity protection and the HathiTrust server environment are available below (see Scenario 1 and Scenario 5, respectively).

- Disaster Recovery Strategy #1: *Redundancy between the Ann Arbor and Indianapolis Sites*

HathiTrust's first line of defense in the event of a disaster is its hot mirror site in Indianapolis. While ingest of material is restricted to the Ann Arbor location, both sites possess two web servers, a MYSQL database server, and an Isilon IQ storage cluster (currently composed of 21 'nodes,' servers composed of Central Processing Units as well as storage). During normal operations, this arrangement allows HathiTrust to balance a high volume of web traffic across both sites such that individual user requests may be handled by either site in a transparent manner. Should the tolerances for failure be exceeded at a site (as in a disaster situation) the failover capability buit into the HathiTrust architecture enables the remaining site to provide access to the designated community without noticeable service disruptions. As noted in the May 2009 HathiTrust Update, with the full operation of both locations, "We are now ensuring that users do not feel the effects of single-site outages, such as routine maintenance,

[12] Tennant, Roy. "Digital Libraries: Coping with Disasters." *Library Journal*, 15 November 2009. Retrieved from http://www.libraryjournal.com/article/CA180529.html on 13 July 2009.
[13] HathiTrust. "Technology" retrieved from http://www.hathitrust.org/technology on 15 June 2009.

by taking advantage of site redundancy."[14] However, because ingest takes place only in Ann Arbor, the loss of key components there would inhibit the repository's ability to acquire new content.

HathiTrust utilizes Isilon System's SyncIQ Application Software to synchronize data at the Indianapolis site with newly ingested or updated material from the Ann Arbor site. The sync to Indianapolis runs on 24 separate subsets of the data and each one runs every 2 hours, with the exception of Sundays. In other words, subset 1 runs at midnight on Monday, subset 2 runs at 2 a.m., and so on. The maximum time for data to be replicated from Ann Arbor to Indianapolis would therefore be three days plus the run time of the sync process (which tends to take less than three hours.)[15]

- "SyncIQ is an asynchronous replication application that fully leverages the unique architecture of Isilon IQ storage to efficiently copy data from a primary cluster to one located at a secondary location."[16]
- "All nodes [… in both the source and target Isilon IQ clusters] concurrently send and receive data during replication jobs in real time, without impacting users reading and writing to the system."[17]
- "A robust wizard-driven web-based interface is fully integrated into [… Isilon's proprietary] OneFS management tool to control all the functionality, including scheduling, policy settings, monitoring and logging of data transferred and bandwidth utilization."[18]
- "Only files that have changed will be replicated to the target clusters. This will optimize transfer times and minimize bandwidth used."[19]
- "In the event the secondary system is not available due to a system or network interruption, the replication job will be able to roll back and restart at the last successful copy operation."[20]
- "Upon a critical failure or loss of network connection, an alert will be sent to all recipients configured to receive critical alerts."[21]

- Disaster Recovery Strategy #2: *Nightly Automated Tape Backups*

HathiTrust's ability to recover from a disaster is also ensured by the nightly automated tape backups performed by the Tivoli Storage Manager (TSM) client application installed on the ingest servers connected to the HathiTrust storage cluster and managed by Michigan's ITCS TSM Group. The *TSM Backup Service Standard Service Level Agreement*[22] outlines the obligations and responsibilities of both the service provider and HathiTrust:

[14] HathiTrust. "Update on May 2009 Activities" (2009) retrieved from http://www.hathitrust.org/updates_may2009 on 2 July 2009.
[15] Snavely, Cory (Head, UM Library IT Core Services). Personal email on 13 July 2009.
[16] "Backup and Recovery With Isilon IQ Clustered Storage," 2007 p. 11
[17] Ibid.
[18] Ibid.
[19] Ibid.
[20] Ibid.
[21] Ibid
[22] Please refer to Appendix F (*TSM Backup Service Standard Service Level Agreement*).

- "The progressive incremental methodology used by Tivoli Storage Manager only backs up new or changed versions of files, thereby greatly reducing data redundancy, network bandwidth and storage pool consumption as compared to traditional methodologies based on periodic full backups."[23]
- "ITCS is responsible for all of the central server hardware, tape hardware, networking hardware, and related components. ITCS is also responsible for hardware maintenance as well as software maintenance, administration, and security audits on the central (non-client) TSM servers." (TSM Backup Service SLA, sec. 4.1)
- "ITCS provides 7x24 on-call monitoring and support, and strives to keep the servers up in production at all times. The target up-time is 99.9% of the time. The TSM hardware design is modular and should allow us to take pieces out of service without affecting customers. Whenever possible, system maintenance will be performed during standard weekend maintenance windows as defined by ITCS." (sec. 4.2)
- "In an emergency, customers can contact tsm@beepage.itd.umich.edu (this will go to the on-call staff's pager in real time). (sec. 4.6)
- "ITCS is responsible for physical security. Machine access audits, OS security, and network security on the TSM server end are also the responsibility of ITCS." (sec. 4.9)
- "The service […] includes data compression, data encryptions, and data replication." (sec. 1.0)
- "ITCS will maintain at least two TSM sites and will mirror data between the sites to provide redundancy in the event of a disaster. Currently those sites are the Arbor Lakes Data Facility (ALDF) at 4251 Plymouth Rd. and the Michigan Academic Computing Center (MACC) located at 1000 Oakbrook Dr." (sec. 4.10)
- "Both facilities are secure, climate controlled sites designed and built for high available production services."[24]
- "In the event of a customer disaster with large-scale (a full server or more) data loss, ITCS will work with the customer to optimize the restore time to best of our ability. We will only be able to devote resources to the extent that other customers are not affected. Restoring large file servers (multiple Terabytes) can take several days. If customers want to minimize this amount of time to restore, we can purchase additional resources for this purpose. Contact us directly, and we'll work out a scenario with costing information. In the event of a MAJOR campus outage affecting a large number of customers, ITCS management will work with customers to determine how to prioritize customer restores." (sec. 4.11)
- "Disaster Recovery planning is the responsibility of the customer unit." (sec. 5.8)

Having established the main Disaster Recovery strategies employed by HathiTrust, we may now proceed to investigate the means by which it anticipates and mitigates the most common threats facing digital repositories.

---

[23] IBM. "IBM Tivoli Storage Manager: Features and Benefits" (2009) retrieved from http://www-01.ibm.com/software/tivoli/products/storage-mgr/features.html?S_CMP=rnav on 16 June 2009.

[24] Information Technology Central Services at the University of Michigan. "Frequently Asked Questions about the TSM Backup Service" (2009) retrieved from http://www.itcs.umich.edu/tsm/questions.php on 16 June 2009.

Scenario 1: Hardware Failure or Obsolescence and Data Loss

- Review: *Risks Involving Hardware Failure or Obsolescence and Data Loss*

The following table highlights the various events which pose a risk to the hardware and data of HathiTrust. These threats may stem from flaws or malfunctions in the equipment itself or as a result of external events that include physical security breaches and natural or manmade disasters. The arrangement of these potential risks reflects the relative severity of their respective consequences.

Severity	Event
High impact	*Loss at a single point of failure* • An additional failure past tolerances when only one site is operational • Service is unavailable and cannot be restored until component is repaired/restored
Moderate Impact	*Failure of a component past redundancy tolerance* • System no longer has redundancy: additional loss or failure of components will result in loss of system. This is a particular problem if one site is already down. • Loss of db server (home of Rights db) or of both Web servers at a site will render that location inaccessible • Loss of four drives or nodes in either Isilon storage cluster will result in the loss of that instance. The cluster will be offline and unable to handle read or write requests; all traffic would have to be handled by the remaining site. • Loss of UM Arbor Lakes site would prevent performance of tape backups. • Loss of UM MACC site would deprive IU site of data redundancy • Loss of ingest servers would prevent new content from entering repository

Low Impact	*Failure of redundant system components*
	• Includes redundant components within each site as well as general redundancy between the IU and UM sites
	  ○ HT infrastructure has been designed to avoid single points of failure and to ensure data and equipment redundancy
	  ○ Service continues in an uninterrupted and transparent manner

- HathiTrust's Solutions for Hardware Failure and Data Loss

The threats faced by HathiTrust's hardware (and associated applications as well as the data stored therein) are comprised of the failure of redundant features, failure that exceeds components' tolerance for redundancy, and single points of failure. While the failure of redundant components may happen more frequently (i.e., the loss of an individual drive within the Isilon IQ cluster), such losses do not have a large impact on the repository; events which compromise single points of failure will have much greater consequences for the continuity of HathiTrust operations. At the same time, while a component may have redundancy on one level (for example, there are five servers dedicated to ingest), that component simultaneously may be considered at a higher level to be a single point of failure (i.e., because the ingest servers are housed in a single chassis, the entire unit is vulnerable to an event such as a fire). This duality highlights the need for vigilance and foresight in managing the repository's infrastructure.

Because HathiTrust relies heavily upon hardware to fulfill its mission and deliver services to its designated community of users, the selection of equipment and development of system architecture

has aimed at minimizing the dangers posed by single points of failure through the introduction of strategic redundancies. The basic means for avoiding the disastrous effects of hardware failure or data loss have been the establishment of the Indianapolis mirror site and the nightly backup of content to tape. (For more detail, please refer to the preceding section). While these strategies account for extraordinary events, HathiTrust's server replacement schedule allows the repository to anticipate the results of normal equipment use and depreciation. Steps to safeguard the long-term functionality of HathiTrust have therefore been complemented by a consideration of best practices for disaster preparedness.

- Redundant Components and Single Points of Failure in the HathiTrust Infrastructure

The following sections provide a general outline of HathiTrust's redundant components and single points of failure. Given the complexity of the repository's infrastructure, unknown or unanticipated scenarios may exist; future Disaster Recovery Planning will thus involve a periodic review of key features and vulnerabilities.

  - *Site Redundancy*: The establishment of the mirror site in Indiana provides HathiTrust with a fully redundant operation. Because both instances provide full access to content in addition to other repository functions, users will not experience a loss or degradation of service in the event that service is lost from one site. Key exceptions to HathiTrust's site redundancy are noted below.
  - *Redundant Components at Each Site*: The following components provide each site with a tolerance under which limited failures will not disrupt major HathiTrust functions and user services.
    - *Web servers*: each site has two servers so that if one fails, the other may continue to handle traffic. These also host the GeoIP database.
    - *Isilon IQ clusters*: the current configuration of 21 nodes features N+3 parity protection; this data redundancy permits the simultaneous failure of 3 drives on separate nodes or the loss of three entire nodes without service degradation.
    - *Ingest servers*: the Ann Arbor site possesses five servers so that ingest may continue (albeit at a slower rate) in the event of any failures.
    - *Large Scale Search (LSS) Solr index*: currently housed on the web servers, but will soon be maintained on five new servers in Ann Arbor.
  - *Single Points of Failure*:[25] These are components of a system which, if lost, will prevent the entire system from functioning. Even those components with wholly redundant peer devices (such as the web or ingest servers) may be considered single points of failure if they have exceeded their capacity to sustain losses (i.e., if one web server at a site has already been lost).
    - Single Points of Failure at the Component Level: Because only one of these components exists at each HathiTrust site, a loss will result in system failure.
      - *MYSQL database server*: houses the rights database, ingest tracking database, and the Collection Builder Solr index
      - *Server network switches*
      - *Outbound network switches*
    - Single Points of Failure at the System Level: While any given component may have various degrees of internal redundancy (such as multiple power supplies or

[25] Content in this section is courtesy of Cory Snavely (personal email from 13 July 2009).

multiple drives) it might still fail as a whole and thus result in the loss of a particular instance of HathiTrust. The following are components located at each site which, while possessed of internal redundancies, are still subject to complete loss (as in the event of a fire) and may thus render a site inoperable.

- *Isilon IQ storage cluster*: the entire cluster could be lost in a large-scale event. Additionally, the loss of a fourth drive or node will exceed the cluster's failure tolerance and result in a service disruption.
- *Web servers*: should one fail, the remaining server will be a single point of failure.
- *Blade server chassis*: since web, ingest, and database servers are housed in one chassis, the entire unit could potentially fail.
- *LSS index*: in the near future, the servers in Ann Arbor will be the sole instance of the Large Scale Search index.
- *Mirlyn database* and *Mirlyn2 Solr index*[26]: these are currently key components of the UM Library infrastructure; should these be unavailable, access to and use of HathiTrust will be compromised.

- Key Features of HathiTrust's Isilon IQ Clustered Storage

The Isilon IQ storage cluster stores and provides digital objects for HathiTrust's partner libraries and members of its designated community. The cluster provides a high degree of inherent redundancy, which gives both HathiTrust sites a considerable degree of tolerance in regards to the failure of various aspects of the storage units. As one example, Isilon's proprietary OneFS operating system permits the individual storage nodes—the individual servers that are the building blocks of the cluster—to function as 'coherent peers' so that any one node 'knows' everything contained on the other units in the cluster.

- "Isilon's OneFS operating system […] intelligently stripes data across all nodes in a cluster to create a single, shared pool of storage."[27]
- "Because all files are striped across multiple nodes within a cluster, no single node stores 100% of a file; if a node fails, all other nodes in the cluster can deliver 100% of the files within that cluster."[28]
- "A distributed clustered architecture by definition is highly available since each node is a coherent peer to the other. If any node or component fails, the data is still accessible through any other node, and there is no single point of failure as the file system state is maintained across the entire cluster."[29]

---

[26] *Mirlyn* is the name of the University of Michigan's current Online Public Access Catalog, which is supported by the Aleph integrated library system. *Mirlyn2* is a beta version of UM's recently implemented next generation catalog, based on the VuFind platform, which will become the main library catalog on August 3, 2009.

[27] Isilon Systems, Inc. "Isilon IQ OneFS Operating System" (2009) retrieved from http://www.isilon.com/products/OneFS.php on 17 June 2009.

[28] Isilon Systems. "Uncompromising Reliability through Clustered Storage: Delivering Highly Available Clustered Storage Systems" (2008) p. 7. "In computer data storage, **data striping** is the technique of segmenting logically sequential data, such as a single file, so that segments can be assigned to multiple physical devices. […] if one drive fails and the system crashes, the data can be restored by using the other drives in the array." (http://en.wikipedia.org/wiki/Data_striping, retrieved on 16 August, 2009).

[29] Isilon Systems. "Breaking the Bottleneck: Solving the Storage Challenges of Next Generation Data Centers" (2008) p. 8

HathiTrust's Isilon IQ clusters ensure a high degree of data redundancy with their N+3 parity protection. N+3 provides triple simultaneous failure protection so that up to three drives on separate Isilon IQ nodes, or three entire nodes, can fail at the same time and all data will still be fully available.

- "Traditional RAID-5 parity protection results in data loss if multiple components fail prior to the completion of a rebuild. FlexProtect, in contrast, automatically distributes all data and error correction information across the entire Isilon cluster and with its robust error correction techniques efficiently and reliably ensures that all data remains intact and fully accessible even in the unlikely event of simultaneous component failures."[30]
- "Each file is striped across multiple nodes within a cluster, with [three] parity stripes for each data block."[31]

The file system may also perform a Dynamic Sector Repair (DSR) at the time of any file writing. If it encounters a bad disk sector, the file system will use parity information elsewhere in the system to rebuild the necessary information and rewrite a new block elsewhere else on the drive. The bad sector will be remapped by the drive so that it is never used again and the write operation will be completed.

The Isilon "restriper" is a meta-process/infrastructure that has four primary phases to help manage and protect data in the event that components of the cluster sustain a partial failure or malfunction. The processes run as background operations and do not require system downtime.[32][33]

- FlexProtect repairs data (i.e., in the event of a drive loss) using parity.
  - "Isilon OneFS with FlexProtect can boast the industry leading Mean Time to Data Loss (MTTDL) for petabyte clusters."[34]
  - "FlexProtect introduces state-of-the-art functionality, which rebuilds failed disks in a fraction of the time, harnesses free storage space across the entire cluster to further insure against data loss, and proactively monitors and preemptively migrates data off of at-risk components."[35]
- AutoBalance "rebalances the data in a cluster according to business rules, in real time, non-disruptively."[36]
  - "As soon as the [new or repaired] node is turned on and network cables are connected, AutoBalance immediately begins to migrate content from the existing storage nodes to the newly added node across the cluster interconnect back-end switch, re-balancing all of the content across all nodes in the cluster and maximizing utilization."[37]

---

[30] Isilon Systems, Inc. "Isilon IQ OneFS Operating System" (2009) retrieved from http://www.isilon.com/products/OneFS.php on 30 June 2009.
[31] Isilon Systems. "Uncompromising Reliability through Clustered Storage: Delivering Highly Available Clustered Storage Systems" (2008) p. 7
[32] *Isilon X-Series Specifications* (product brochure)
[33] Information on the Isilon restriper comes from a personal email sent by Kip Cranford of Isilon Systems, Inc. on 1 June 2009.
[34] Isilon Systems. "Data Protection for Isilon Scale-Out NAS" (2009) p. 4
[35] Isilon Systems, Inc. "Isilon IQ OneFS Operating System" (2009) retrieved from http://www.isilon.com/products/OneFS.php on 15 June 2009.
[36] McFarland, Anne. "Isilon Accelerates Delivery of Digital Content" *The Clipper Group Navigator* (2003).
[37] Isilon Systems. "The Clustered Storage Revolution" (2008) p. 13

- Collect cleans up orphaned nodes and data blocks to prevent fragmentation of data.
- MediaScan verifies disk sectors.
  - The function of MediaScan is to scan every block in the file system looking for bad disk sectors. If it encounters a bad sector, it will perform a Dynamic Sector Repair (DSR) and use parity information elsewhere in the system to rebuild the necessary information and rewrite a new block somewhere else on the drive.
  - MediaScan periodically reviews data blocks and disk sectors that may not have been accessed, from a file level, in months or years and thereby helps to keep the drives as healthy as possible.
- As of the OneFS 5.0 release, all file system metadata can be checked by the IntegrityScan restriper phase. This process will allow HathiTrust to completely check file data and metadata via associated checksums.

Other instances of inherent redundancy include non-volatile RAM, a fully journaled file system, and software applications that manage client connections in the event of a node's failure.

- "OneFS is a fully-journaled file system with large amounts of battery-backed non-volatile random access memory (NVRAM) within each node, which ensures the integrity of the file system in the event of unexpected failures during any write operation."[38]
- "The Isilon SmartConnect module [… ensures] that when a node failure occurs, all in-flight reads and writes are handed off to another node in the cluster to finish its operation without any user or application interruption. […] If a node is brought down for any reason, including a failure, the virtual IP addresses on the clients will seamlessly fail over across all other nodes in the cluster. When the offline node is brought back online, SmartConnect automatically fails back and rebalances the NFS clients across the entire cluster to ensure maximum storage and performance utilization."[39]

- Hardware Support and Service

HathiTrust equipment is covered by support and service agreements with its various vendors (Sun Microsystems, Dell, CDW-G, etc.). A good example of one such agreement is found in the "Platinum" support provided by Isilon Systems and which includes:

- Extended 24x7x365 Telephone & Online Hardware and Software Support
- 24x7 Proactive Monitoring & Alerts – Email Home (for Hardware and Software)
- Return Parts to Factory for Repair and 4-hour Replacement Parts Delivery
- SupportIQ (Enhanced Serviceability Diagnostics) and System Event Tracking
- On-site Troubleshooting
- Isilon Hardware Installation
- Software Product Documentation, Release Notes, and access to Product Technical Notes
- Remote Diagnosis (Provided User Grants Access)
- Maintenance & Patch Releases

[38] Isilon Systems. "Uncompromising Reliability through Clustered Storage: Delivering Highly Available Clustered Storage Systems" (2008) p. 9
[39] Isilon Systems. "Data Protection for Isilon Scale-Out NAS" (2009) p. 6

- Minor and Major Upgrade Releases (Includes Performance Improvements, New Features, Serviceability Improvements). [40]

- Equipment Tracking

LIT Core Services (CS) maintains an inventory of servers on a wiki page accessible to its staff. Details include each server's name, location, online and retire dates, upgrades, notes on storage, and its primary service. Additional information is provided related to specifications, support contracts, and key contact information. **The CS server inventory is currently out of date**.

- Hardware Replacement Schedule
  - "HathiTrust replaces storage regularly, approximately every 3-4 years or as the usable life of storage equipment dictates" (*HT TRAC* C1.7)
  - "HathiTrust staff upgrade hardware on a regular basis (i.e., every three or four years), and to help detect more rapid growth in demands, the web server and storage infrastructures have their own performance monitoring that indicate overload conditions." (*HT TRAC* C1.10)

- Timeline for Emergency Replacement of HathiTrust Infrastructure

Should a serious event require the replacement of part (or all) of the HathiTrust technical infrastructure, the following timeline provides a general estimate of the time required to order, ship, and install new equipment. A cursory review of the time necessary for HathiTrust to recover from a major disaster at the main Ann Arbor or Indianapolis data center suggests that a large event could idle an instance of the repository for at least a month and a half. In addition to the servers and switches mentioned above, critical components include four 30A power distribution units (PDUs) per rack and four racks per data center as of this writing.

  - Submission of Purchase Orders:
    - For orders under $5,000, the M-Pathways application allows the University Library's business manager to send purchase orders directly to vendors.
    - For orders over $5,000, Procurement Services normally takes one to two business days to approve the purchase, but the process may take up to a week if questions arise or additional purchase information is needed.
  - Delivery of Equipment:
    - Products the vendor has in stock and available for immediate shipment take 1-3 days to be delivered.
    - Items that need to be configured (such as servers) usually take 1-2 weeks.
    - Isilon storage will take 3 weeks to be delivered in a worst case scenario.
  - Installation:
    - 3 days FTE for Isilon IQ cluster in addition to the time required for other servers, switches, PDUs and rack units.

[40] Isilon Systems. "SupportAdvantage Offerings" (2009) retrieved from http://www.isilon.com/support/?page=plans on 30 June 2009.

- Data Restoration: about .5 TB/hour (15 days, as of June 2009)[41]
    - While HT has about 110 TB of data in its storage, the backup tapes maintained by the TSM Group contain roughly 176 TB of information due to the data encryption used to protect the intellectual rights of the material (as of 06/2009).
    - The length of time required for a 'bare-metal restoration' will be influenced by tape mounts, network speed, restoring to the NFS shares, decryption, et cetera.
    - If the library/HT were to purchase an additional tape drive (at roughly $20,000), the process could be sped up, perhaps to about 1 TB/hour.
    - In the event of a large-scale disaster in which multiple campus units require extensive data restoration, the *TSM Backup Service SLA* states that "ITCS management will work with customers to determine how to prioritize customer restores." (sec. 4.11) This determination will reflect the University of Michigan's organizational priorities[42]:
        - *Priority 1:* Health and safety of faculty, staff, students, hospital patients, contractors, renters, and any other people on University premises.
        - *Priority 2:* Delivery of health care and hospital patient services
        - *Priority 3:* Continuation and maintenance of research specimens, animals, biomedical specimens, research archives.
        - *Priority 4:* Delivery of teaching/learning processes and services
        - *Priority 5:* Security and preservation of University facilities/equipment.
        - *Priority 6:* Maintenance of community/University partnerships.
- Fractional restores would, for the most part, run at comparable speeds unless there was a need to restore a large number of random files, in which case there would be a decrease in speed due to tape seek and mount times.
- Delays in recovery could be increased dramatically if the MACC data center or its infrastructure has sustained damage and needs repair.

- HathiTrust and Insurance Coverage at the University of Michigan

The Office of Financial Operations reviews and adds financial assets greater than $5,000 to the asset management system of the University of Michigan. The Property Control Office is then responsible for tagging financial assets with unique University of Michigan identifiers and tracking them. Risk Management Services administers the University's property insurance and will provide the reimbursement of replacement costs for items self-insured by Michigan. As of July 2009, the nature and extent of the University of Michigan's insurance coverage for HathiTrust hardware remained under review. The main contact with Risk Management Services in this matter has been Cyndi Mesa, Head of UM Library Finance.

[41] Hanover, Cameron (ITCS TSM Group Storage Engineer). Personal email on 23 June 2009.
[42] University of Michigan Administrative Information Services. "Emergency Management, Business Continuity, and Disaster Recovery Planning" (2007) retrieved from http://www.mais.umich.edu/projects/drbc_methodology.html on 6 July 2009.

Scenario 2: Network Configuration Errors

- Review: *Risks Involving Network Configuration Errors*

The following table summarizes the risks facing HathiTrust as the result of network configuration errors. Consideration is given to network connections within UM data centers as well as at UM's Hatcher Graduate Library (site of key administrative and development activities). The arrangement of these events reflects the relative severity of their respective consequences.

Severity	Event
High impact	• Loss of server network switch or outbound network switch • Loss of access to UMnet Backbone
Moderate Impact	• Extended loss of power at Hatcher Library could lead to loss of local servers and disruption of administrative and operational activities.
Low Impact	• Loss of power that threatens ability to connect to Local Area Network (LAN)/Backbone o The library remains (for now) a priority recipient of electricity from the UM power plant o Campus data centers have UPSs and redundant backup power • Failure of local/server-side connections o Should problems arise with connections to individual nodes, the clustered architecture of the Isilon system will allow read/write requests to be handled by alternate nodes. o If connections fail at one HT site, traffic can be handled by remaining site.

- HathiTrust's Solutions for Network Configuration Errors

HathiTrust's continued access to the Internet via the UMnet Backbone is essential for its continued provision of service. The repository receives network infrastructure maintenance through UM's ITCS/ITCom; with its robust disaster planning in addition to the lessons learned from the Midwest blackout of 2003, ITCom guarantees continued network access in all but the most catastrophic scenarios. In the event of a widespread power outage, HathiTrust would be able to maintain access to the UMnet Backbone since datacenters are equipped with redundant power supplies and the Hatcher Graduate Library is currently categorized as a priority recipient of power from the university. ITCS also has 17 generators which can be used to maintain power to network switches in the event of a blackout. The responsibilities and obligations of both parties are outlined in the *Customer Network Infrastructure Maintenance Service Agreement*.[43]

- Extent of ITCom Support
  - "ITCom agrees to provide the Unit Network Infrastructure Maintenance to include data switches, routers, access points, hubs, uninterruptible power supplies (UPS's), firewalls, and other identified and agreed upon components." (ITCS sec. 1.0)

[43] Please refer to Appendix G (ITCS/ITCom *Customer Network Infrastructure Maintenance Service Agreement*).

- ITCom Responsibilities
    - "Provide and maintain the necessary materials and electronic components to operate the Unit Network Infrastructure." (sec. 5.2)
    - "Provide configuration and Network Infrastructure Administration support necessary to repair and maintain the Unit Network Infrastructure hardware and software covered by this agreement." (sec. 5.3)
    - "Monitor 24 hours/day and 365 days/year (24 x 365), supported protocols to the backbone interface of the Units network up to and including the extension to the first hub or switch." (sec. 5.6)
    - "Monitor 24 hours/day and 365 days/year (24 x 365), network interfaces on uninterruptible power supplies (UPS) that support the Unit network switches. Provide notification in the event that a UPS is activated, (input power is lost or degraded and system switches to battery power), deactivated, (input power is restored), or unreachable. Provide notification to the Unit Network Administrator when batteries degrade to the point of needing replacement." (sec. 5.7)
    - "Provide maintenance on the station cabling as installed by ITCom, or an approved U-M vendor which met ITCom installation specifications." (sec. 5.8)
    - "Provide Preventative Maintenance (clean & vacuum) on each Customer Unit switch covered in this agreement yearly." (sec. 5.9)

- ITCom Services in Response to Outages or Degradation Impacting the Network
    - "A response within 30 minutes of the ITCom NOC notification or the Unit's call, to provide information to the Unit on specific steps that have been/will be taken to resolve the problem." (sec. 7.2.1)
    - "An on-site visit, if necessary, within two (2) hours of the response (i.e., the maximum on-site response time will be two and a half (2 1/2) hours). An update will be provided to the Unit Network Administrator if on site and a best guess ETR will be provided based on available facts. ITCom will continue to provide the Unit with updates every two hours during an outage." (sec. 7.2.1)
    - "If an outage is identified within the agreement service hours ITCom will resolve the outage even if the repair time extends beyond the service agreement hours." (sec. 7.2.1) (Repairs outside of the agreement hours result in additional labor expenses.)
    - Conduct monitoring via SNMP POLLING at one minute intervals. (sec. 7.2.1)

- HathiTrust Responsibilities

ITCom's responsibilities end at the first network switch and from there to its servers, HathiTrust is responsible for maintaining network connectivity and security. The repository uses Internet2 for communication and synchronization between the Ann Arbor and Indianapolis sites. Each Isilon node has dual 10 GB Infiniband ports for internal (i.e., intra-cluster) communication and dual 1 GB Ethernet for external communication.

Scenario 3: Network Security and External Attacks

- Review: *Risks Involving Network Security and External Attacks*

The following table gives a general overview of the basic threat an external attack or network security breach poses to HathiTrust; entries are arranged by severity. The list, however, is not exhaustive and no attempt has been made to publicize potential vulnerabilities.

Severity	**Events**
High impact	• Unauthorized access to HathiTrust content leads to the infringement of copyrights. • Loss of data or functionality for an extended period of time as a result of malicious activity.
Moderate Impact	• HathiTrust services are temporarily unavailable as a result of malicious activity.
Low Impact	• The delivery of HathiTrust services slows as the result of malicious activity. • A security weakness exists within the system but remains unexploited.

- HathiTrust's Solutions for Network Security

Malicious activity against HathiTrust could involve unauthorized access to a system or data, denial of service, or unauthorized changes to the system, software, or data. As an academic entity, the repository is seen as less of a target for such actions than commercial or governmental targets; despite this perceived lower risk, HathiTrust has not been lulled into a false sense of security. The repository takes seriously the potential for violations of its network and operating system security and therefore has instituted a program of periodic software updates in addition to the maintenance of an ITCom-supported firewall, authentication-required access, and other measures (such as throttling software to deter denial of service attacks). Because content is currently accepted from trusted sources (namely, Google and legacy digital collections from HathiTrust partners) the GROOVE process does not include a virus detection phase. As digital objects are ingested from a greater number of sources, additional security measures should be considered.

  - "HathiTrust staff apply security updates to the operating system and to networking devices as soon as they become available in order to minimize system vulnerability. As with new software releases, security updates are tested in a development environment before being released to production. Software packages that present a lower security risk and that have a greater potential to affect application behavior (web servers, language interpreters, etc.) are generally installed, configured and tested manually to allow for greater control in managing updates. Software updates are not applied automatically; moreover, updates that present a potential for having an impact on system behavior are applied and tested first in the development environment. If no impacts are seen, HathiTrust staff apply these updates in production after a testing period of at least one week." (*HT TRAC* C1.10)

Scenario 4: Format Obsolescence

- Review: *Risks Involving Format Obsolescence*

The following table outlines the threats posed by format obsolescence and arranges them according to their potential severity.

Severity	Events
High impact	• Applications and hardware are no longer able to read or display digital objects. • Errors in translating and reading files are not understood or acknowledged by repository users.
Moderate Impact	• Problems with the translation of file formats result in DIPs that do not faithfully reflect the original digital objects.
Low Impact	• Formats and associated applications change but retain compatibility with older versions of the file formats.

- HathiTrust's Solutions for Format Obsolescence

An awareness and acknowledgement of the dangers of format obsolescence has led HathiTrust to implement proactive policies and procedures to ensure long-term access to the repository's content. The repository only accepts specific formats that meet rigorous specifications and, through the prior experience of University of Michigan personnel, has developed protocols for the successful migration of content from one format to another. In addressing the threat of format obsolescence, the preservation of the integrity and authenticity of deposited content has been an overarching concern.

- Selection of File Formats
  - "HathiTrust is committed to preserving the intellectual content and in many cases the exact appearance and layout of materials digitized for deposit. HathiTrust stores and preserves metadata detailing the sequence of files for the digital object. HathiTrust has extensive specifications on file formats, preservation metadata, and quality control methods, included in the University of Michigan digitization specifications, dated May 1, 2007."[44] (*HT TRAC* B1.1)
  - "HathiTrust currently ingests only documented acceptable preservation formats, including TIFF ITU G4 files stored at 600dpi, JPEG or JPEG2000 files stored at several resolutions ranging from 200dpi to 400dpi, and XML files with an accompanying DTD (typically METS). HathiTrust supports these formats because of their broad acceptance as preservation formats and because the formats are documented, open and standards-based, giving HathiTrust an effective means to migrate its contents to successive preservation formats over time, as necessary. The Repository Administrators have undertaken such transformations in the past; moreover, HathiTrust offers end-user services that routinely transform digital objects stored in HathiTrust to "presentation" formats using many of the widely available software tools associated with HathiTrust's

[44] Specifications are available at http://www.lib.umich.edu/lit/dlps/dcs/UMichDigitizationSpecifications20070501.pdf

preservation formats. HathiTrust gives attention to data integrity (e.g., through checksum validation) as part of format choice and migration."[45]

- "Each format conforms to a well-documented and registered standard (e.g., ITU TIFF and JPEG2000) and, where possible, is also non-proprietary (e.g., XML)." (*HT TRAC* B4.2)

- Format Migration Policies and Activities
  - "HathiTrust is committed to migrating the formats of materials created according to [its] specifications as technology, standards, and best practices in the digital library community change." (*HT TRAC* B1.1)
  - "HathiTrust staff members conduct migrations from one storage medium to another using tools that validate checksums internally. (Digital objects are stored both online and on tape, and the online storage system conducts regular scans to detect and correct data integrity problems.) A total file count is done following a large data transfer, and regularly scheduled integrity checks follow." (*HT TRAC* C1.7)
  - "[HathiTrust] has migrated large SGML-encoded collections to XML, and Latin-1 character encodings to UTF-8 Unicode. Our success in migrating from older formats to newer formats demonstrates our commitment to our collections and our ability to keep materials in our repository viable. All migrations are documented in change logs." (*HT TRAC* B4.2)

[45] HathiTrust. "Preservation" (2009) retrieved from http://www.hathitrust.org/preservation on 16 June 2009.

Scenario 5: Core Utility and/or Building Failure

- Review: *Risks Involving Core Utility or Building Failure*

The following table summarizes the dangers a utility or building failure poses to HathiTrust and ranks events by their potential severity.

Severity	Events
High impact	• Extensive structural damage renders the MACC (or key elements of its infrastructure) unusable and necessitates the establishment of a hot site to recover and continue operations. • Additional failure past tolerance in backup cooling or power infrastructure
Moderate Impact	• Failure of backup power past redundancy tolerance (failure of 2 generators) ○ Data center coordinator may initiate load shed and shut down half of the MACC (but library racks will remain operational) • Structural damage renders facility temporarily unsafe and/or unusable.
Low Impact	• Loss of power • Loss of environmental control units within redundancy

- HathiTrust's Solutions for Utility or Building Failure

The continued delivery of HathiTrust's services depends upon the maintenance of power, environmental control, and security in its server environment at the Michigan Academic Computing Center (MACC) and other locations that host components of the repository. In this respect, HathiTrust is heavily reliant upon the infrastructure of the MACC as well as that of the Arbor Lakes Data Facility, home to one instance of the TSM Group's backup tape library. Both locations provide closely monitored and highly redundant environments that help ensure that HathiTrust's infrastructure remains secure and operable. At the same time, administrative and data management functions critical to the development and maintenance of the repository take place in the University of Michigan's Hatcher Graduate Library. The service and cooperation of Michigan's Plant Operations Division are therefore critical for the continued access to and use of this structure in the operation of HathiTrust.

- General Maintenance and Repairs in University of Michigan Facilities

Facilities and maintenance issues on the University of Michigan campus are reported to the Plant Operations Division, the Department of Public Safety (DPS), and Occupational Safety and Environmental Health (OSEH) in addition to the impacted facility's manager. Repair work is coordinated by the University Library facilities manager in conjunction with administrators and workers from Plant Operations.

- The Michigan Academic Computing Center (MACC)

The MACC hosts many of the key components of the Michigan's University Library system and as well as the technical infrastructure of HathiTrust. The University of Michigan does not own the building in which the data center is located but instead operates the MACC in conjunction with the Michigan Information Technology Center (MITC) Foundation and other partners. The *MACC Server Hosting Service*

Level Agreement[46] lists the responsibilities of the data center as well as the repository; of particular significance are the MACC's agreements to:

- "Provide a controlled physical environment to support servers [with] room average temperature of between 65 and 75 degrees and 35-50% relative humidity [and] monitored environmentals (temperature, humidity, smoke, water, electrical." (sec. 4.1)
- "Provide adequate, conditioned, 60-cycle electrical service with adequate backup electrical capacity to support circuits, service, and outlets [and also to] provide Uninterruptible Power Supply (UPS) and generator backup" (sec. 4.2)
- "Provide 7x24 telephone contact for emergencies and for emergency access to facility." (sec. 4.4)

In addition to features such as redundant electrical and environmental systems, the MACC maintains a full-time coordinator and staff who provide 24x7 responses to failures or malfunctions in the server environment. Alerts prompted by issues with the environmental systems or power are sent to the University of Michigan Network Operations Center (NOC) during non-business hours.

- Overview:
  - "The MACC's redundancy is designed to ensure the safety and security of the data housed within. It consists of:
    - A dual power path from the property line to the power distribution units
    - Diesel powered generators for electrical backup
    - Flywheels (not batteries) to provide power while the generators come on
    - State-of-the-art generators and flywheels for backup power
    - Three extra computer room air conditioners
    - Two extra dry coolers
    - Glycol loop for cooling with two parallel pathways with crossover valves at regular intervals."[47]
  - "A state-of-the-art monitoring system keeps track of 1,700 different parameters and automatically notifies staff of any irregularity."[48]
- Environmental Controls and Monitoring
  - "The MACC has 18 Computer Room Air Conditioning units (CRACs). At any given time, only 15 are necessary to maintain the required temperature and humidity. [Thus, the computer room has N5+1 redundancy in its cooling ability.] It also is equipped with a number of portable coolers to address specific cooling needs. The heat from the room is transferred to an under-floor glycol loop that releases the heat to the outdoors."[49]

---

[46] Please refer to Appendix H (*MACC Server Hosting Service Level Agreement*).
[47] Michigan Academic Computing Center. "Vital Statistics" (2009) retrieved from http://macc.umich.edu/about/vital-statistics.php on 16 June 2009.
[48] --. "Michigan Academic Computing Center" (2009) retrieved from http://macc.umich.edu/index.php on 16 June 2009.
[49] --. "Vital Statistics" (2009) retrieved from http://macc.umich.edu/about/vital-statistics.php on 16 June 2009.

        - "The layout of the facility allows the front on the computer racks to be facing the cold aisles. These aisles have perforated floor tiles through which the cool air is pumped directly to the computers located there. Heat is discharged from the backs of the computers, which creates the hot aisles. This alternating arrangement facilitates the cooling process, as the hot air produced by the computers can be siphoned off before it mingles too much with the cooler air of the facility."[50]
        - "Two separate smoke detection and fire alarm systems protect the MACC. One is for the building; the other is for the MACC itself. The two systems work together to activate alarm systems and notify the fire department and key personnel. In the event of an actual fire, the fire-suppression system pipes will not fill with water unless there is a pressure drop caused by melting of one or more of the sprinkler heads."[51]
    - Backup Power
        - "Three generators, each roughly the size of a rail car, provide backup power. Only two of the three are required to run the facility in the event of a power outage."[52]
        - "The MACC uses environmentally responsible flywheels instead of batteries for power backup while the generators come online. The combination of generators and flywheels provides the facility with a fully redundant uninterruptible power system (UPS)." [53]
        - The MACC has a contract with the UM Plant Operations Division for the delivery of diesel fuel for its generators in the event of an extended black out. [54]
        - In the event that a backup generator is disabled, the MACC coordinator will initiate load shed, in which one half of the MACC will be shut down so that the other half (and requisite environmental systems) may continue to operate. The HathiTrust and UM Library racks are among those which will retain power should this response prove necessary.[55]

- Arbor Lakes Data Facility (ALDF)

The ALDF houses the TSM Group's infrastructure and one instance of the backup tape library that forms an integral part of HathiTrust's Disaster Recovery strategy. As the home of critical components of the UMnet Backbone, the ALDF provides a safe and secure location for one set of the repository's backup tapes. In the interest of security, this report will omit further information on the exact nature of the facility's power and environmental systems.

---

[50] Ibid.
[51] Ibid.
[52] --. "Michigan Academic Computing Center" (2009) retrieved from http://macc.umich.edu/index.php on 16 June 2009.
[53] Ibid.
[54] Gobeyn, Rene (MACC Data Center Coordinator). Personal interview on 23 June 2009.
[55] Ibid.

Scenario 6: Software Failure or Obsolescence

- Review: *Risks Involving Software Failure or Obsolescence*

The following table details various risks inherent to software failure or obsolescence and ranks them according to their severity.

Severity	Events
High impact	• Software bug escapes detection in development environment and results in crash of application.
Moderate Impact	• Software bug escapes detection in development environment and prevents full access to digital objects. • Improper version of software is introduced to system (could have a greater or lesser impact depending on results of error and repository's ability to detect it).
Low Impact	• Software bug escapes detection in development environment and prevents full use of system capabilities (i.e., rotation of images or additional functionality)

- HathiTrust's Solutions for Software Issues

The development and use of HathiTrust's tools and resources depends on highly functional software applications. Repository policies have therefore been crafted to ensure that these applications are thoroughly tested and regularly updated to minimize the threat of service outages as a result of software failure or obsolescence. HathiTrust furthermore employs open source applications that are well-supported and enjoy wide spread use and development within the digital library community.

  - "Changes in software releases of all components of the system (from ingest to access) are developed and tested in an isolated "development" environment to prepare for release to production. When ready for release, developers record the changes made and increment version numbers of system components as appropriate using a version control system. New versions of software are released using automated mechanisms (in order to prevent manual errors). Major changes and upgrades in hardware architecture are recorded in monthly reports of unit activity, and thus are traceable to that level of detail." (*HT TRAC* C1.8).
  - "Additionally, subsets of production data are available in the development environment to allow developers to ensure proper system behavior before releasing changes to production." (*HT TRAC* C1.9)
  - "In order to design, build and modify software for the designated end-user community, HathiTrust conducts an active usability program and seeks input from the Strategic Advisory Board of HathiTrust. Similarly, with regard to software development in support of the archiving needs of the Participating Libraries, HathiTrust focuses on the development of highly functional ingest and validation mechanisms. HathiTrust also seeks and responds to guidance from the Strategic Advisory Board with regard to archiving services." (*HT TRAC* C2.2)

Scenario 7: Operator Error

- Review: *Risks Involving Operator Error*

The following table summarizes risks to HathiTrust posed by operator error; events are ranked according to their potential severity.

Severity	Events
High impact	• Operator error results in the irreparable loss of data or damage to equipment. • Operator error results in loss of key repository functions (ingest, storage, dissemination, etc.) for an extended period of time.
Moderate Impact	• Operator error remains undetected and causes persistent problems in the system but has no long term consequences.
Low Impact	• Operator error is detected by normal procedures or via an activity log and can be readily corrected.

- HathiTrust's Solutions for Operator Error

In any human enterprise, occasional operator error is unavoidable; HathiTrust strives to ensure that any such events are detected and resolved in a timely fashion.[56] To help avoid occurrences and mitigate their potential impact, HathiTrust has automated many procedures and also relies upon application assertions, which can notify administrators when processes are not operating correctly. Even if an error is introduced to the file system and then backed up, the TSM client saves up to seven versions of a file for up to six months so that an earlier version can be retrieved.

- Ingest: The Google Return (Object-Oriented) Validation Environment (GROOVE) process is entirely automated to avoid the introduction of operator error to the process; steps include:
    - Identification of material for ingest
    - Decryption and unzipping of files Format verification and validation with JHOVE
    - Lun Barcode and MD5 checksum validation
    - Creation of HathiTrust METS documents
    - Establishment of HathiTrust handles (persistent URLs)
    - Extension of the pairtree file directory (as new material enters the system)

- Archival Storage: Files stored within the repository are not accessed directly or manipulated by staff so that neither the zipped image and OCR files nor the METS document may be accidently altered or deleted.

- Dissemination: The page-turner application references the stored image and then creates a .png (for TIFFs) or .jpg (for JPEG2000s) file for display to the viewer.

- Data Management: "New versions of software are released using automated mechanisms (in order to prevent manual errors)." (*HT TRAC* C1.8)

[56] Please refer to Appendix B (HathiTrust Outages from March 2008 through April 2009).

Scenario 8: Physical Security Breach

- Review: *Risks Involving a Physical Security Breach*

Maintaining the physical security of the HathiTrust infrastructure is yet another crucial element in the repository's efforts to manage risks and thereby lessen the chance that a disaster-type event occurs. Risks involve the damage and destruction of equipment and could even extend to unauthorized system access. Multiple levels of security exist at both the Michigan Academic Computing Center (MACC) and the Arbor Lakes Data Facility (ALDF) to protect HathiTrust from the acts of vandalism, destruction or malicious tampering. Details on the potential impacts of a physical security breach are covered in "Scenario 1: Hardware Failure" and "Scenario 3: Network Security."

- HathiTrust's Solutions for Physical Security
  - "Each of [the HathiTrust] storage or tape instances is physically secure (e.g., in a locked cage in a machine room) and only accessible to specified personnel."[57]

- Security at the MACC

The *MACC Server Hosting SLA* states the data center staff will:

  - "Provide services necessary to maintain a safe, secure, and orderly environment for all tenants of the MACC." (sec. 4.7)
  - "Provide access control via HiD card and biometric readers for those listed on the Tenant Staff Authorized for Access list." (sec. 4.5)

The MACC Web site and the *Michigan Academic Computing Center Operating Agreement*[58] provide additional details concerning the resources and procedures that help protect HathiTrust's equipment at the MACC. The MACC Data Center Coordinator personally oversees the enforcement of security protocols and conducts regular audits of security logs and, when necessary, reviews surveillance video footage.

  - Security Systems
    - "State-of-the-art security devices such as iris scanners, cameras, closed circuit television and on-call staff keep the data and machines housed in the MACC safe."[59]
    - "Access to the data center will be by two-factor authentication (access card and iris scan) or escorted, supervised access. Access to the building will be by access card." (*MACC OA*, sec. 5.3.1)
    - "Cameras throughout the corridor, security trap, and facility will be monitored and maintained by the Data Center Coordinator." (sec. 5.2.1)
  - Security Procedures

[57] HathiTrust. "Technology" (2009) retrieved from http://www.hathitrust.org/technology on 15 June 2009.
[58] Please refer to Appendix I (*Michigan Academic Computing Center Operating Agreement*).
[59] Michigan Academic Computing Center. "Vital Statistics" (2009) retrieved from http://macc.umich.edu/about/vital-statistics.php on 17 June 2009.

- "The Operations Advisory Committee will establish procedures for granting access cards to the facility to those whose jobs require hands-on access to systems. All requests for access cards will be vetted and approved by the Operations Advisory Committee at their next meeting." (sec. 5.3.2)
- "Everyone on the access list for the data center will be required to attend a training session before working in the data center and sign an access agreement stating policies they must observe while in the data center." (sec. 5.3.8)

- Security at the ALDF

As noted in the *TSM Backup Service SLA*, the University of Michigan's ITCS "is responsible for physical security" at the ALDF. (sec. 4.9) While this document will not detail specific features of the ALDF's operation, multiple levels of security and oversight are employed.

Scenario 9: Natural or Manmade Disaster

- Review: *Risks Involving a Natural or Manmade Disaster*

The following table details the risks to HathiTrust posed by a natural or manmade disaster; events are ranked by order of their severity. Due to possible overlap between this scenario and Scenario 1 (Hardware Failure), readers are encouraged to consult that earlier section.

Severity	Events
High impact	• Widespread damage to a data center and/or its infrastructure that forces an instance of the repository to find a new hot site with sufficient power supply, environmental controls, and security. • Damage to work areas force staff to relocate to a new center of operations. • Extensive loss or damage to hardware requires large-scale replacement. • With the extended loss of one site, HathiTrust loses redundancy (and possibly some functionality: i.e. the ability to ingest new material in Ann Arbor) and thus a central component of its disaster recovery and backup plans. • An act of violence or terrorism occurs at or near HathiTrust facilities.
Moderate Impact	• An event results in an extended outage at one site that exceeds the recovery time objective. • Hardware sustains some damage and site is able to continue operation in a reduced capacity. • An actual or threatened act of violence or terrorism forces the temporary evacuation or quarantine of HathiTrust facilities.
Low Impact	• Local conditions result in a temporary outage at a HathiTrust site.

- HathiTrust's Solutions for Natural or Manmade Catastrophic Events

The University of Michigan Ann Arbor Campus Emergency Procedures (revised January 2008) has set procedures to address building evacuations (in the event of fire), tornadoes, severe weather, flooding, chemical/biological/radioactive spills, as well as bomb threats, civil disturbances, and acts of violence or terrorism.[60] In all cases, staff will follow the directions of Public Safety and not re-enter buildings or resume work "until advised to do so by DPS or OSEH or someone from on-site incident command."

In the event of a severe natural or manmade disaster, the repair and restoration of the physical locations of HathiTrust infrastructure would need to be coordinated between the repository and the appropriate facility managers. Such activity would rely upon the disaster recovery plans in place at the MITC Building (home of the MACC) and University of Michigan (which includes the Hatcher Graduate Library and the ALDF). It must be noted that an event which causes significant damage to an important structure or to a building's infrastructure could result in the loss of an instance of the repository for an extended period of time. In such a case, HathiTrust would need to set up an alternate hot site until structural restoration is complete (or a new facility has been found).

[60] Please see Appendix C (Washtenaw County Hazard Ranking List).

- Basic Disaster Recovery Strategies

In the immediate aftermath of a large-scale manmade or natural disaster, the repository's immediate recovery will be enabled by its basic system architecture:

  - "the initiative's technology concentrates on creating a minimum of two synchronized versions of high-availability clustered storage with wide geographic separation (the first two instances of storage are located in Ann Arbor, MI and Indianapolis, IN), as well as an encrypted tape backup (written to and stored in a separate facility outside of Ann Arbor)."[61]

The establishment of the mirror site in Indianapolis and the retention of multiple backup tapes at two locations in Ann Arbor ensure that a serious event at either location will not impede the continued functioning of the repository at the other. Consideration must be given as to how data at the Indianapolis site will be backed up and how key repository functions (such as ingest) will proceed if the Ann Arbor instance is off-line for an extended period of time. Likewise, a long-term outage at the IU location would require HathiTrust to establish a third site for data backup (i.e., a location where additional copies of backup tapes could be stored).

[61] HathiTrust. "Technology" retrieved from http://www.hathitrust.org/technology on 15 June 2009.

Scenario 10: Media Failure or Obsolescence

- Review: *Risks Involving Media Failure or Obsolescence*

The following table summarizes risks to HathiTrust posed by the failure of the media used for its data backups. While the risks from this are limited (both copies of the tape backups would have to be impacted for data to be unavailable), the issue should nonetheless be addressed with regular test restorations and/or inspections of the media.

Severity	Events
High impact	• Physical degradation (i.e. in tape binder, substrate, or magnetic content) affects both copies of older backup tapes.
Moderate Impact	• Because backup tapes are not regularly tested or audited, the physical substrate of tapes may degrade over time.
Low Impact	• Bad tape is detected during a tape backup.

- HathiTrust's Solutions for Media Failure

Given the nature of HathiTrust's storage system, this scenario is only a concern in regards to the digital magnetic tapes used by the TSM Group for backups.

  - Two tape copies of all backup data are made and these are stored in separate climate-controlled conditions in tape libraries at the MACC and the ALDF.
  - Content is transferred to new tape during data defragmentation (which occurs when existing tapes are 80% full),
  - If a degraded or otherwise 'bad' section of tape is detected during a backup procedure that tape is immediately marked as "read only."
    - Data is thenceforth written to a different tape; existing data on the bad tape will be copied to properly functioning media.
    - If data cannot be reclaimed from bad tape, the TSM Group would contact HathiTrust so that the backup of content can be properly completed.

- Remaining Vulnerabilities

There is some reason for concern in this area because the TSM Group does not have a regular program to monitor its media for physical degradation or impairment after data defragmentation. While the tapes are reported to be highly dependable, problems such as "sticky shed" (the hydrolysis of the tape's binder) could become an issue with older tapes. A regular program of tape validation or test restorations would provide an opportunity to check on the physical condition and data integrity of the tapes. Likewise, the creation of a schedule for the replacement of older tapes could avoid future problems with media degradation.

Conclusions and Action Items

- Conclusions

As this report demonstrates, a variety of risk management strategies in addition to design elements, operating procedures, and service and support contracts endow HathiTrust with the ability to preserve its digital content and continue essential repository functions in the event of a range of disasters. The establishment of the Indianapolis mirror site, the performance of nightly tape backups, and the redundant power and environmental systems of the MACC reflect professional best practices and will enable HathiTrust to weather a wide range of foreseeable events. As it is, disasters often result from the unknown and the unexpected; while the aforementioned strategies are crucial components of a Disaster Recovery Plan, they must be supplemented with additional policies and procedures to ensure that, come what may, HathiTrust will be able to carry on as both an organization and a dedicated service provider.

In the effort to secure HathiTrust's long-term continuity, the present document stands merely as a preliminary step in the establishment of a legitimate Disaster Recovery Plan. The data on HathiTrust's policies, procedures, and contracts consolidated herein should facilitate the data collection requisite to the initial phases of the planning process, but the core activities of formulating technical and administrative response strategies and delegating roles and responsibilities remain to be undertaken. The following section outlines recommendations and action items derived from research into the repository as well as from discussions with Cory Snavely and other HathiTrust staff members. Items have been separated into an approximate timeline of activity ranging from Short Term through Long Term and the arrangement within each category represents a suggested (but by no means definitive) order of accomplishment. For a more detailed explanation of action items related explicitly to Disaster Recovery Planning, please refer to the overview of the planning process in Appendix E or consult Appendix D for a list of more comprehensive guides and resources.

(NB: ***** = Denotes an ongoing activity.)

- Short Term Action Items (0-6 months)
    a. Resolve the nature and extent of the insurance coverage for HathiTrust equipment.
    b. Arrange with TSM Group administrators to periodically perform a volume audit of backup tapes to ensure data integrity.
    c. Institute periodic test restores with TSM Group to ensure that the process will run smoothly in the event of a disaster.
    d. Discuss the creation of a long-term replacement schedule for backup tapes with the TSM Group to avoid the possibility of media degradation.
    e. Improve control over system components
        i. Update the hardware inventory to include all important system components; document models, serial numbers, UM ID's, associated software and version number, date of purchase, original cost, as well as vendor contact information and product support contracts.*****

ii. Establish a software inventory to document necessary applications in the event of hardware loss; should include purpose, acquisition date, cost, license number, and version number.*****
iii. Create a map identifying where components are in the MACC and within individual racks*****
iv. Review and assess points of failure as well as the adequacy of redundant components.*****

f. Establish phone trees
   i. Include key contacts for different types of disaster
   ii. Prioritize phone trees to target individuals who
      1. Make decisions
      2. Have vital information
      3. Can offer assistance in resolving situations
   iii. Distribute information and explain protocols to all relevant staff*****
   iv. Develop a regular maintenance/update schedule (once every 4-6 months)*****
g. Thoroughly document and make available (as needed) important institutional knowledge so that HathiTrust may continue to function in the event of the extended absence or loss of key staff.*****
h. Identify disaster preparedness and disaster recovery measures in place at Indianapolis.

- <u>Intermediate Term (6-12 months)</u>
  a. Form a Disaster Recovery Planning Committee to research and develop plans and to oversee their implementation.
  b. Communicate and coordinate planning activities between Ann Arbor and Indianapolis.*****
     i. Consider the formation of sub-committees for localized research and development of plans and an executive committee to oversee the implementation and management of plans.
  c. Draft a Disaster Recovery Planning policy statement to define the mandate, responsibilities, and objectives for the plan.
  d. Initiate the data collection and analysis phase of the planning process.
     i. Identify core repository functions and associated hardware and infrastructure elements.
     ii. Determine the potential impact from the loss of those functions
     iii. Define the levels of functionality required for partial as well as full recovery. Establish what level is needed for HT to fulfill its mission and the needs of its users.
     iv. Define HathiTrust's Recovery Time Objective (RTO: the maximum allowable outage period for services) and Recovery Point Objective (RPO: the point in time to which data stores must be returned following a disaster).
     v. Determine the availability of resources in the event of a disaster and establish the repository's prioritization with major service providers and vendors (i.e., TSM Group, ITCom, Isilon, etc.).

e. Address risks uncovered in the data collection phase and institute preventative controls as needed to anticipate and mitigate those risks.*****
f. Develop recovery strategies to bring core functions back online as soon as possible within a set cost range.
    i. Establish a logical progression in the restoration of services and associated components.
    ii. Identify the resources required for these efforts.
    iii. Consider alternative solutions, including partial (vs. full) recovery
g. Communicate planning goals and efforts to key contacts from service providers and vendors to better coordinate recovery efforts.*****
h. Initiate the production of core Disaster Recovery documents (see Appendix E for more information). The following list is not exhaustive; data collection and analysis will help determine if all or other plans (i.e., a web continuity plan) are needed.
    i. Business Continuity Plan: details HathiTrust's core functions and the priorities for re-establishing each in the event of a disruption.
    ii. Continuity of Operations Plan: focuses on restoring an organization's (usually a headquarters element) essential functions at an alternate site and performing those functions for up to 30 days before returning to normal operations.
    iii. IT Contingency Plan: addresses explicitly the disaster planning for computers, servers, and elements of the technical infrastructure that support key applications and functions.
    iv. Crisis Communications Plan: establishes procedures for internal and external communications during and after an emergency.
    v. Cyber-Incident Response Plan: defines the procedures for responding to cyber attacks against the HathiTrust IT system.
    vi. Occupant Emergency Plan: defines response procedures for staff in the event of a situation that poses a potential threat to the health and safety of HathiTrust personnel or their environment. (This requirement is addressed by University of Michigan *Building Emergency Action Plans*.)
    vii. Disaster Recovery Plan: brings together guidance and procedures from the other plans to enable the restoration of core information systems, applications, and services. This plan defines roles and responsibilities within Disaster Response Teams.
    viii. Disaster Recovery Training Plan: establishes the situations and procedures to be covered by HathiTrust's Disaster Recovery training.

- Long Term (12+ months)
    a. Complete and implement Disaster Recovery Plans.
        i. Distribute physical copies of the plans as needed and include at least one copy in an off-site location.
        ii. Integrate elements of response strategies into system architecture to facilitate their deployment in the event of a disaster.*****

b. Disaster Recovery Committee should monitor changes in best practices and technology, update plans, and oversee organizational readiness.**
    i. Initiate staff training so that individuals are familiar with Disaster Recovery procedures and communication protocols.**
    ii. Institute regular tests of disaster preparedness with simulated disasters involving different components of HathiTrust operations.**
    iii. Establish a schedule for maintenance and revisions to the Disaster Recovery documents.**
    iv. Coordinate Disaster Recovery Plan implementation, training, and review with Indianapolis.**
c. Store an additional copy of backup tapes at a third site to increase exposure and limit the chance that a widespread event in Ann Arbor could impact both local copies.
d. Explore the possibility of establishing a third site for HathiTrust's digital objects to increase exposure and address concerns over the relative geographical proximity of Indianapolis and Ann Arbor.
e. Determine the feasibility of moving operations to a "hot" site in Ann Arbor should a disaster render the MACC unusable.
    i. Identify suitable sites and consider making preliminary arrangements.
    ii. Identify and price out equipment/infrastructure necessary to continue operations.
f. Plan for integration of new system components should the sudden collapse of Isilon leave HathiTrust without service/support.
g. Consider an increase to system security measures as content becomes accepted from a wider range of sources and as HathiTrust becomes a higher-profile organization.

APPENDIX A: Contact Information for Important HathiTrust Resources

Indiana University Mirror Site

- Andrew Poland (Staff, Information Technology Services)
  - ajpoland@iupui.edu
  - (317) 274-0746
- Troy Dean Williams (Vice President for Information Technology, IU at Bloomington)
  - trowill@indiana.edu
  - (812) 856-5323

University of Michigan

Michigan Academic Computing Center (MACC): Houses much of the technical infrastructure of the University Library's digital resources.

- Rene Gobeyn (MACC Data Center Coordinator)
  - rgobeyn@umich.edu
  - (734) 936-2654
- ITCom UMNOC (Network Operations Center)
  - TROUBLE@UMICH.EDU
  - (734) 647-8888

ITCS-ITCom: Responsible for maintaining network connections to the UMnet Backbone and Internet; ITCom provides maintenance and support services for hardware and software.

- Mike Brower (Senior Project Manager, UM Libraries)
  - mbrower@umich.edu
  - (734) 936-9736
- Krystal Hall (Disaster Recovery Planner, ITCS/ITCom Operations)
  - kahall@umich.edu
  - (734) 647-3214
- ITCom UMNOC (Network Operations Center)
  - TROUBLE@UMICH.EDU
  - (734) 647-8888

Tivoli Storage Manager Group: Responsible for nightly automated tape backups of storage servers.

- Andrew Inman (Service Manager)
  - ainman@umich.edu
  - (734) 615-6286
- Cameron Hanover (Storage Engineer)
  - chanover@umich.edu
  - (734) 764-7019
- General Support: tsmadmin@umich.edu
- Emergency contact: adsm@beepage.itd.umich.edu
  - Message will go to on-call staff's pager in real time
- Notification of TSM and related outages via UMOD group fln@umich.edu

Arbor Lakes Data Facility: Houses one instance of the TSM backup tape library.

- ITCom UMNOC (Network Operations Center)

    - trouble@umich.edu
    - (734) 615-4209
- Ken Pritchard (ALDF facility manager)
    - kenprit@umich.edu
    - (734) 615-2812

Procurement Services: Approves departmental purchases over $5,000; buyers also work as intermediaries with vendors.

- Steve Worden (UM Hardware Purchasing Specialist)
    - sfworden@umich.edu
    - (734) 645-8972
- Shelly Eauclaire (Senior Buyer, Purchasing Services)
    - seauclai@umich.edu
    - (734) 615-8767
- Ian Pepper (UM Dell Computers Contract Administrator)
    - ipepper@umich.edu
    - (734) 647-4981
- Jeff Rabbitt (Alternate Dell Contract Administrator)
    - rabbit@umich.edu
    - (734) 644-9232

Property Control: Responsible for tracking and tagging the university's assets.

- Mary Ellen Lyon (Business Operation Manager)
    - melyon@umich.edu
    - (734) 647-3351 (t,th)
    - (734) 763-1197 (m,w,f)

Office of Financial Analysis:

- David Storey (Inventory Coordinator): Delivers UM property tags to equipment at the MACC.
    - dstorey@umich.edu
    - (734) 647-4264

Risk Management Services: Provides insurance coverage of university assets.

- Kathleen Rychlinski (Assistant Director, Risk Management Services)
    - kmrychli@umich.edu
    - (734) 763-1587

Non-University Contact Information

Isilon Systems

- Jim Ramberg (Regional Territory Manager)
    - jim.ramberg@isilon.com
    - Desk: (847) 330-6399
    - Cell: (630) 561-2463

Sun Microsystems

- Christine Sluman (Service Sales Rep—Education)
    - Christine.Sluman@Sun.COM
    - (303) 557-3660, ext.60519

    - (303) 949-1567 (Cell)
- Larry Zimmerman (Michigan Account Manager-Sales)
    - larry.zimmerman@sun.com
    - (248) 880-3756

CDW-G

- University of Michigan Account Team
    - hansenandadam@cdwg.com
- Hansen Chennikkra (Account Manager)
    - hansche@cdwg.com
    - (866) 339-3639
- Adam Sullivan (Account Manager)
    - adamsul@cdwg.com
    - (866) 339-4118

Dell Computers

- Brian Ullestad (Higher Education Account Manager)
    - Brian_Ullestad@Dell.com
    - 1-800-274-7799 ext. 7249522

APPENDIX B: HathiTrust Outages from March 2008 through April 2009[62]

- April 2009: HathiTrust experienced reduced performance from 11:00pm EDT on Thursday, April 23 to 8:22am EDT on Friday, April 24 due to a database problem at one of the sites and from 5:30pm to 9:00pm EDT on Thursday, April 30 due to unintended consequences from a networking configuration change.
- March 2009: HathiTrust was unavailable on Tuesday, March 3 from 7:00-8:00am EST and on Thursday, March 5 from 7:00-7:45am EST for operating system and database software upgrades.
- February 2009: On Sunday, February 22 at 8:40am EST, a power surge resulting from electrical system maintenance caused HathiTrust database and web servers to go offline. Staff learned of the problem at approximately 6:00pm EST, and service was restored by 6:30pm EST.
- January 2009: A brief outage is scheduled in January for a storage system software upgrade.
- December 2008: On Friday, December 19 at 7:30am EST, HathiTrust was down briefly to apply security updates to a database server. Service was restored at 7:40am EST.
- November 2008: On Tuesday, November 4 at 7:30am EST, HathiTrust was down briefly to apply security updates to a database server. Service was restored at 7:45am EST
- October 2008: No outages reported.
- September 2008: On Thursday, September 18 at approximately 9:30am EDT, HathiTrust became inaccessible due to a software problem on a storage system; the problem was related to our work with data synchronization. Support was contacted and the problem was resolved at 10:45am EDT
- August 2008: On Tuesday, August 26 at approximately 9:00am EDT, a database server was brought down to move to Indianapolis. Prior to shutting this server down, we did not update a manual failover configuration, causing volumes to be inaccessible to some users. The problem was resolved at 11:15am EDT.
- July 2008: Service was unavailable on Thursday July 31 from 7:00-7:30am EDT for a storage system software upgrade.
- June 2008: No outages reported.
- May 2008: No outages reported.
- April 2008: No outages reported.
- March 2008: No outages reported.

[62] HathiTrust. "Updates" from http://www.hathitrust.org/updates retrieved on 16 June 2009.

APPENDIX C: Washtenaw County Hazard Ranking List

The following list ranks a variety of natural and manmade events within Washtenaw County, Michigan, based upon their frequency of occurrence and the extent of their potential impact (on the county's population).

Rank	*Hazard*	*Frequency*	*Population Impacted*
1	Convective weather (severe winds, lightning, tornados, hailstorms)	Once or more/yr.	250,000
2	Hazardous materials incidents: transportation	Once or more/yr.	2,000
3	Hazardous materials incidents: fixed site	Once or more/yr.	10,000
4	Severe winter weather hazards (ice / sleet / snow storms)	Once or more/yr.	250,000
5	Infrastructure failures	Once every 5 yrs.	30,000
6	Transportation accidents: air and land	Once or more/yr.	100
7	Extreme temperatures	Once every 5 yrs.	10,000
8	Flood hazards: riverine / urban flooding	Once every 10 yrs.	2,000
9	Nuclear attack	Has not occurred	250,000
10	Petroleum and natural gas pipeline accidents	Once every 10 yrs.	1,000
11	Fire hazards: wildfires	Once or more/yr.	0

Source: *Washtenaw County Hazard Mitigation Plan* (available online at http://www.ewashtenaw.org/government/departments/planning_environment/planning/planning/hazard_html)

APPENDIX D: Annotated Guide to Disaster Recovery Planning References

The topic of disaster recovery planning for the print and analog resources of libraries has been widely dealt with in professional literature, but comparatively little information exists concerning the development and implementation of plans for the digital content of cultural institutions. The following bibliography details resources which provide guidance, examples, and explanations of the objectives and strategies for digital Disaster Recovery Plans. It consists primarily of material compiled by Lance Stuchell (ICPSR Intern) and Nancy McGovern (ICPSR Digital Preservation Officer) and is included here with their permission.

University of Michigan Resources

- University of Michigan Administrative Information Services (MAIS): Emergency Management, Business Continuity, and Disaster Recovery Planning.
    - http://www.mais.umich.edu/projects/drbc_methodology.html
    - This site broadly outlines the need for and functions of Emergency Management, Business Continuity, and Disaster Recovery Planning at UM. It also contains templates designed to help units plan, test, and audit disaster and continuity programs.

- Provost and Executive Vice President for Academic Affairs: *Standard Practice Guide: Institutional Data Resource Management Policy*
    - http://spg.umich.edu/
    - This policy defines institutional data resources as University assets and makes recommendations on identifying, preserving, and providing access to these assets. The digital resources of the library may be identified as such, based upon their use by departments across the university.

- ICPSR Disaster Planning Resources:
    - Digital Preservation Officer Nancy McGovern is part of a Disaster Recovery initiative at ICPSR and over the past several years her team (including Lance Stuchell) has produced a variety of documents and templates to help other institutions work the through the planning process.
    - Documents are available upon request and should be posted in the near future (as of July 2009) to the ICPSR Web site (http://icpsr.umich.edu/).

- Disaster Recovery Experts:
    - Rene Gobeyn (MACC Data Center Coordinator)
        - Managed and coordinated Disaster Recovery for U.S. military data centers
        - rgobeyn@umich.edu
    - Krystal Hall (Disaster Recovery Planner, ITCS/ITCom Operations)
        - Helped develop current ITCS Disaster Recovery plans
        - kahall@umich.edu

External Resources

- General Guide to Disaster Planning
  - *Contingency Planning Guide for Information Technology Systems: Recommendations of the National Institute of Standards and Technology*, NIST Special Publication 800-34, June 2002.
    - http://csrc.nist.gov/publications/nistpubs/800-34/sp800-34.pdf
    - An indispensable resource which was used heavily by ICPSR in its Disaster Recovery planning. It covers everything from initial data collection and policy formation to the structure of disaster response teams and the articulation of recovery strategies.

- Examples and Tools for the Documentation Outlined by NIST Guide**:**
  - Full Disaster Recovery Plan:
    - *United States Department of Agriculture Disaster Recovery and Business Resumption Plans*
    - *http://www.ocio.usda.gov/directives/doc/DM3570-001.htm*
  - Business Continuity Plan (BCP):
    - *MAIS: Emergency Management, Business Continuity, and Disaster Recovery Planning*
    - http://www.mais.umich.edu/projects/drbc_templates.html
    - This site provides several resources that deal with continuity planning.
  - Continuity of Operations Programs (COOP):
    - *FEMA: Continuity of Operations (COOP) Programs*
      - http://www.fema.gov/government/coop/index.shtm
      - Contains a lot of useful information on government policy, templates, and training resources to assist in the creation of a COOP.
    - *Ready.gov: Continuity of Operations Planning*
      - http://www.ready.gov/business/plan/planning.html
      - Guidelines for composing a business COOP, including what outside actors should be involved in the planning process.
    - *The Florida Department of Health: Continuity of Operations Plan for Information Technology*
      - http://www.naphit.org/global/library/basement_docs/FL_DisasterRecovery_template.doc
      - Lengthy (40 pages) and detailed COOP template written for an IT environment.
    - *Florida Atlantic University Libraries: Continuity of Operations Plan*
      - http://www.staff.library.fau.edu/policies/coop-2007.pdf
      - A detailed working COOP, which includes reactions to specific disaster scenarios.
  - IT Contingency Plan:

        - See the USDA Disaster Recovery Plan for an example of an IT Contingency Plan.
    - Cyber Incident Response Plan:
        - *Multi-State Information Sharing and Analysis Center Cyber Incident Response Guide*
            - http://www.msisac.org/localgov/documents/FINALIncidentResponseGuide.pdf
            - The guide provides a step-by-step process for responding to incidents and developing an incident response team. It may also serve a template in order to draft a Cyber-Incident Response Policy and Plan.
    - Crisis Communication Plan:
        - *Ready.gov: Write a Crisis Communication Plan*
            - http://www.ready.gov/business/talk/crisisplan.html
            - This site provides guidelines for composing a business disaster communication plan and includes suggestions for the plan's Web presence.
        - *NC State University: Crisis Communication Plan*
            - http://www.ncsu.edu/emergency-information/crisisplan.php
            - This is the policy and plan for the University as a whole. While much of this policy deals with communication at a high level, useful sections detail vital contacts within the organization (including who to contact first), and how to manage external communications.
        - Other thorough university policies and plans include the LSU: Crisis Communication Plan and the Missouri S&T: Crisis Communication Plan.
        - Heritage Microfilm Flood Update Email
            - This email was sent in response to the June 2008 flooding that occurred in the Midwest.
            - It updates clients on the outage of NewspaperArchive.com which resulted from a flood-induced widespread power failure. It is an excellent example of an external crisis communication to users.
    - Disaster Recovery Plans (DRP):
        - *The University of Iowa: IT Services Disaster Recovery Plan*
            - http://cio.uiowa.edu/ITplanning/Plans/ITSdisasterPrep.shtml
            - This policy details the data collection and assessment which informs the UI plan and also includes emergency procedures, response strategies, and a crisis communication plan.
        - *University of Arkansas: Computing Services Disaster Recovery Plan*
            - http://www.uark.edu/staff/drp/
            - A complete and thorough plan that outlines the initiation of emergency and recovery procedures, and addresses how the plan will be maintained.
        - *Adams State College (CO): Information Technology Disaster Recovery Plan*
            - http://www.adams.edu/administration/computing/dr-plan100206.pdf

        - This plan has a thorough section on risk assessment.
    - DigitalPreservationEurope *Repository Planning Checklist and Guidance*
        - http://www.digitalpreservationeurope.eu/platter.pdf
        - Designed for use with the Planning Tool for Trusted Electronic Repositories (PLATTER), this document outlines considerations for a Disaster Recovery Strategic Objective Plan (SOP) and places them in context with other repository plans.
- Occupant Emergency Plan (OEP):
    - This requirement is addressed by University of Michigan *Building Emergency Action Plans (EAP).*
        - http://www.umich.edu/~oseh/guideep.pdf
- Disaster Recovery Training Guides:
    - dPlan.org
        - Provides useful information on training and an online form that would be useful in assigning trainers and monitoring the training process.
    - *CalPreservation.org: Disaster Plan Exercise*
        - http://calpreservation.org/disasters/exercise.html
        - Provides roles and teaching points for a role-play training exercise that focuses on a disaster in a library.

- Policy Planning Tools**:**
    - *Association of Public Treasurers of the United States and Canada: Disaster Policy Certification Guidelines*
        - www.aptusc.org/includes/getpdf.php?f=**Disaster_Policy**.pdf
        - This planning document and template for disaster management policies provides outlines and example language on several facets of a strong policy, including the possible loss of a building, the replacement of computer resources, and testing and training for the disaster plan. It also outlines the need to identify possible threats to assets.

- Examples of Disaster Planning Policies**:**
    - *Arkansas Secretary of State: Disaster Planning Policy*
        - http://www.sos.arkansas.gov/elections/elections_pdfs/register/oct_reg/016.14.01-020.pdf
        - This policy outlines areas of responsibility between departments and units, and includes training, communication, and recovery plan updates.
    - *Washington State Department of Information Services: Disaster Recovery and Business Resumption Planning Policy*
        - http://isb.**wa**.gov/policies/portfolio/500p.doc
        - This document illustrates policy formation for an IT Disaster Recovery Plan. It provides guidelines for Disaster Recovery Planning as well as maintenance, testing, and training involved with the recovery plan.

- *Florida State University: Information Technology Disaster Recovery and Data Backup Policy*
  - http://oti.fsu.edu/oti_pdf/Information%20Technology%20Disaster%20Recovery%20and%20Data%20Backup%20Policy.pdf
  - This document includes policy for data backup as well as Disaster Recovery. Part of the policy includes a definition of Best Practice Disaster Recovery Procedures, as well as an outline of the university's own IT recovery planning and implementation procedures.

- Example of a Relevant Disaster Planning Program:
  - *OCLC Digital Archive Preservation Policy and Supporting Documentation*
    - http://www.oclc.org/support/documentation/digitalarchive/preservationpolicy.pdf
    - This document has a clear articulation of OCLC's disaster policy, along with an outline of disaster prevention and recovery procedures and a time-frame for the restoration of services in the event of a disaster.
    - The policy includes a good definition of a disaster prevention and recovery plan: "A set of responses based on sound principles and endorsed by senior management, which can be activated by trained staff with the goal of preventing or reducing the severity of the impact of disasters and incidents."
    - OCLC embeds its disaster plan within its overall preservation policy, stating: "The goal of disaster prevention is to safeguard the data (content and metadata) in the Digital Archive and to safeguard the Digital Archive's software and systems. For disaster prevention and recovery, all data (content and metadata) is considered of equal value."

- Designing a Disaster Planning Program:
  - *Michigan State University: Step by Step Guide to Disaster Recovery Planning*
    - http://www.drp.msu.edu/Documentation/StepbyStepGuide.htm
    - This program breaks down the disaster planning process into steps, and provides information relevant to individual units within a university setting. The MSU Disaster Recovery Planning Home page (http://www.drp.msu.edu/) also offers a variety of resources.
  - *Minnesota State Archives: Disaster Preparedness*
    - http://www.mnhs.org/preserve/records/docs_pdfs/disaster_000.pdf
    - This document is a detailed guide to the disaster planning process. While mostly dealing with paper records, the document clearly identifies different roles and responsibilities for members of the planning and recovery team.
  - *Cisco Systems: Disaster Recovery Best Practices White Paper*
    - http://www.cisco.com/warp/public/63/disrec.pdf

        - The paper outlines Disaster Recovery using the framework of the above resources, but tailors it to an IT point of view. It has useful information on how to prepare and recover both hardware and software assets.
    - *AT&T: Key Elements to an Effective Business Continuity Plan*
        - http://www.business.att.com/content/article/Key_to_Effective_BC_Plan.pdf
        - A short paper that summarizes business continuity planning in the private sector.

- General Information
    - Federal Emergency Management Administration: *Emergency Management Guide for Business & Industry*
        - http://www.fema.gov/business/guide/index.shtm
        - A practical guide with step-by-step advice on creating a Disaster Recovery program. Includes information on the formation on a planning committee, organizational analysis, and details on specific hazards.
    - *Special Libraries Association Information Portal: Disaster Planning and Recovery*
        - http://www.sla.org/content/resources/infoportals/disaster.cfm
        - An exhaustive list of resources, this page includes articles on digital disaster recovery strategies as well as information on planning, examples of plans, and links to a wide range of resources in the public and private sector.

Written Resources:

- Wellheiser, Johanna and Jude Scott. *An Ounce of Prevention: Integrated Disaster Planning for Archives, Libraries, and Record Centres*. Lanham, MD: Scarecrow Press, 2002.
    - http://mirlyn.lib.umich.edu/F/?func=direct&doc_number=004233950&local_base=AA_PUB
- Cox. Richard J. *Flowers After the Funeral: Reflections on the Post-9/11 Digital Age*. Lanham, MD: Scarecrow Press, 2003.
    - http://mirlyn.lib.umich.edu/F/?func=direct&doc_number=004341258&local_base=AA_PUB
- Matthews, Graham and John Feather, eds. *Disaster Management for Libraries and Archives*. Burlington, VT: Ashgate, 2003.
    - http://mirlyn.lib.umich.edu/F/?func=direct&doc_number=004354795&local_base=AA_PUB

APPENDIX E: Overview of the Disaster Recovery Planning Process

Various resources agree that there is no one way to go about initiating a Disaster Recovery program or drafting a DR plan. An organization must proceed according to its functions and resources as well as the needs of its designated community of users. The following discussion draws heavily upon the *ICPSR Disaster Planning Policy Framework* (written by Nancy McGovern and Lance Stuchell) and the *Contingency Planning Guide for Information Technology Systems* published by NIST (2002). As such, it represents a consolidation and simplification of information presented in more depth elsewhere. A list of planning resources (with link information to full texts) is available in Appendix D.

- **Basic Precepts of Disaster Recovery Planning**
  1) Disaster Recovery Planning is a continuous activity that involves monitoring internal conditions as well as evolutions in technology and threats; responding to new developments that arise; revising plans so that they remain relevant and effective; training staff according to plans; and testing organizational readiness.
     a. There is no single document which contains "the plan"; rather, a Disaster Recovery Plan consists of a suite of documents that require a regular schedule of testing and revision to be effective.
     b. There is no point at which a Disaster Recovery Plan is "finished."
  2) Disaster Recovery Planning needs to be an organization wide activity
     a. Disaster recovery must be one of the basic functions of HathiTrust.
     b. An effective plan needs full administrative support.
     c. Policies and procedures must complement and conform to disaster response plans established by the university, city, and Department of Homeland Security.
  3) Disaster recovery cannot be limited to the hardware and software components or data collections of HathiTrust; planning must also account for the impact of human emergencies on the repository's operations.

- **Essential Steps in Disaster Recovery Planning**
  1) Establish a Disaster Recovery Planning Committee.
     a. This group will research and develop the plan and help with its implementation as well as monitor the training, testing, and revising of plans to ensure organizational compliance and readiness.
     b. The committee should involve individuals representing the various mission critical units within the library (from administration to Core Services to the Digital Preservation Librarian) who will participate in the development of policy and recovery planning.
     c. It is essential that the committee involve individuals with the authority to support and enforce recommendations.
     d. The committee's activities should initiate the formation of a Disaster Response Program.
  2) Draft a Disaster Recovery Planning Policy Statement

a. Enables the organization—and others—to understand the scope and nature of the Disaster Recovery Plan.
b. Establishes the organizational framework and responsibilities for the planning process.
c. Key policy elements (as detailed in the NIST report):
    i. Roles and responsibilities within the organization in regards to planning
    ii. Mandate for Disaster Recovery as well as any statutory or regulatory requirements
    iii. Scope as applies to the type(s) of platform(s) and organizational functions subject to Disaster Recovery Planning
    iv. Resource requirements for the Disaster Recovery program
    v. Training requirements
    vi. Exercise and testing schedules (at least one major annual test)
    vii. Plan maintenance schedule (elements should be reviewed annually)
    viii. Frequency of backups and storage of backup media.

3) Conduct Data Collection and Analysis (i.e. "Business Impact Analysis")
    a. Determine critical functions and identify specific system resources required to perform them. Minimum requirements for functionality should be established.
    b. Determine risks and vulnerabilities facing the repository's systems and infrastructure.
    c. Identify and coordinate with internal and external points of contact to determine how they depend on or support the repository and its functions; consider how one failure might cascade into others.
        i. Identify resources that are crucial to HathiTrust (I.e., Mirlyn)
        ii. Determine the allowable outage/disruption time for these resources
    d. Develop recovery priorities; balance the cost of inoperability against the cost of recovery
        i. Determine HathiTrust's position within the priorities of the university as well as with its major service providers and vendors (i.e., TSM Group, ITCom, Isilon, etc.) to better understand how that prioritization will impact recovery efforts.
        ii. Establish the most crucial functions which must be restored first.
        iii. Determine HathiTrust's Recovery Time Objective (RTO, i.e., the maximum allowable outage period) and Recovery Point Objective (RPO, i.e., the point in time to which data files must be restored after a disaster).
        iv. Review potential resources (financial, personnel, etc.) within HathiTrust as well as those available via contracts, service providers, and product support. This step should involve the clarification of HathiTrust's position within the university's as well as key service providers' and vendors' priorities.
4) Address risks uncovered in the data collection phase and institute preventative controls as needed to anticipate and mitigate those risks.

5) Develop recovery strategies that respond to the potential impacts and maximum allowable outage times established in the data collection phase. Efforts should focus on solutions that are cost-effective and technically viable.
    a. Strategies should be designed to bring core functions back online as soon as possible within an established cost range.
    b. Recovery efforts must be prioritized according to the nature of core functions as well as logical order of procedures.
    c. Alternative solutions should be considered based upon cost, availability of resources, outage times, levels of functionality (partial vs. full), and ability to integrate methods with existing infrastructure.
    d. Determine the practicality of partial (vs. full) recovery in order to bring services back on line in a timely and cost-effective manner.
    e. Recovery strategies and resources should be incorporated (as possible) into the repository's system architecture so that in the event of a disaster, the response may proceed in an efficient and straightforward manner.
6) Formalize and record collected data and recovery strategies in Disaster Recovery Documents. In the process of producing this wide range of documents, an organization is forced to consider and document policies and procedures related to a variety of key administrative and technical issues. The decision of which plans to include (and which to exclude) must be determined based upon a review of HathiTrust's needs and objectives. Additional documents (a Web continuity plan, for instance) may be necessary based upon data collection and analysis.
    a. Business Continuity Plan
        i. Business continuity is the ability of a business to continue its operations with minimal disruption or downtime in the event of natural or manmade disasters.
        ii. Such planning allows an organization to ensure its survival by considering potential business interruptions and establishing appropriate, cost-effective responses.
        iii. The Business Continuity Plan details HathiTrust's core functions and the priorities for re-establishing each in the event of a disruption. It should address key administrative and support functions as well as those which directly involve the repository's designated community.
        iv. The plan should thoroughly document the nature of key functions, interdependences, the impact of their loss, and alternative means to ensure their continuation in the event of a disaster. MAIS offers a useful Business Continuity planning template at http://www.mais.umich.edu/projects/drbc_templates.html.
    b. Continuity of Operations Plan (COOP)
        i. The COOP focuses on restoring an organization's (usually a headquarters element) essential functions at an alternate site and performing those functions for up to 30 days before returning to normal operations.

ii. This plan may include the Business Continuity Plan and Disaster Recovery Plan as appendices.

c. IT Contingency Plan
    i. The IT Contingency Plan addresses disaster planning for computers, servers, and elements of the technical infrastructure that support key applications and functions.
    ii. It should account for the following:
        1. Document hardware and software
        2. Develop an emergency contact list
        3. Back up and store all data files off-site
        4. Proactively monitor equipment and data
        5. Install and update antivirus software on both computers and servers
        6. Develop recovery scenarios
        7. Communicate and monitor the plan
    iii. The plan allows HathiTrust to formalize and document procedures and policies already in place and details the repository's adherence to these goals.

d. Crisis Communications Plan
    i. Communication is a vitally important aspect of Disaster Recovery Planning and an organization's actual response in a disaster.
    ii. The Crisis Communications Plan establishes procedures for internal and external communications during and after an emergency.
    iii. The different phases of crisis communication encompass the initial notification of an event, damage assessment, and plan activation as well as status reports (as needed) and the eventual completion of recovery efforts.
    iv. Activation of the communications plan must be the responsibility of a specific individual.
    v. The Disaster Response Team coordinates with the Crisis Communication Team to ensure that information provided about an emergency is clear, concise, and consistent.

e. Cyber-Incident Response Plan
    i. This plan defines the procedures for responding to cyber attacks against the HathiTrust IT system.
    ii. It provides a formal framework for the identification, mitigation, and recovery from malicious computer incidents, such as unauthorized access to a system or data, denial of service, or unauthorized changes to system hardware, software, or data.

f. Occupant Emergency Plan

i. The Occupant Emergency Plan defines response procedures for library staff in the event of a situation that poses a potential threat to the health and safety of personnel, the environment, or HathiTrust property.

ii. HathiTrust may utilize the framework provided by UM Building Emergency Action Plans for this element.

g. Disaster Recovery Plan

i. The primary focus of the Disaster Recovery Plan is the restoration of core information systems, applications, and services.

ii. The plan brings together guidance and procedures from the other plans (i.e., Business Continuity Plan, IT Contingency Plan, Crisis Communications Plan, etc.) pertaining to emergencies that result in interruptions of service that exceed acceptable down times, as defined in the BCP.

iii. The plan should detail established recovery strategies for specific disaster situations as well as the teams involved in their execution.

iv. Personnel should be chosen to staff disaster response teams based on their skills and knowledge. Ideally, teams would be staffed with the personnel responsible for the same or similar operation under normal conditions. It's also important that team members should be familiar with the goals and procedures of other teams to facilitate inter-team coordination. Each team is led by a team leader (with a suitable alternate) who directs overall team operations and acts as the team's representative to management and liaisons with other team leaders. Disaster Response cannot be individual-specific or overly reliant on specific people. Teams must assign each role at least one alternate in the event that core people are unavailable at the time of a disaster.

v. NIST suggests that a capable strategy will require some or all of the following functional groups. For HathiTrust, many of these are already in place in the form of University of Michigan units and service providers.

1. An authoritative role for overall decision-making responsibility
2. Senior Management Official
3. Management Team
4. Damage Assessment Team
5. Operating System Administration Team
6. Systems Software Team
7. Server Recovery Team (e.g., client server, Web server)
8. LAN/WAN Recovery Team
9. Database Recovery Team
10. Network Operations Recovery Team
11. Application Recovery Team(s)

12. Telecommunications Team
13. Hardware Salvage Team
14. Alternate Site Recovery Coordination Team
15. Original Site Restoration/Salvage Coordination Team
16. Test Team
17. Administrative Support Team
18. Transportation and Relocation Team
19. Media Relations Team
20. Legal Affairs Team
21. Physical/Personnel Security Team
22. Procurement Team (equipment and supplies)

h. Disaster Recovery Training Plan
   i. This plan will establish the situations and procedures to be covered by HathiTrust's Disaster Recovery training.
   ii. The contents of the plan should reflect the range of responsibilities held between administrators, department heads, and staff within HathiTrust.
   iii. The plan should accommodate Disaster Recovery Planning Committee members as well as those of the Disaster Response Team. For the latter, it should identify key roles and responsibilities in recovery efforts.
   iv. The plan should allow in-house training to be supplemented by external opportunities.
   v. A regularly scheduled emergency drills should also be included to test the readiness of staff and the appropriateness of response procedures.

7) Implement elements developed in planning process. Procedures and policies related to communication, technological solutions, etc. must be incorporated into HathiTrust's overall design and operation so that Disaster Recovery becomes a critical organizational function.
8) Institute regular program of training and testing to be sure that staff understand and accept policies and procedures and to ensure that HathiTrust is prepared for a disaster.
9) Conduct regular review and maintenance of Disaster Recovery documents to respond to changes in personnel, organizational structure or functions, and evolutions in technology and/or threats.

- **Main Phases in a Disaster Response:**

1) Notification/Activation: This phase covers the initial actions once a situation has been detected or is threatened. It includes damage assessment and the implementation of an appropriate response strategy.
   a. Proper diagnosis and communication (both internal and external) of a disaster is essential.

b. The nature of individual events will determine who needs to be involved (i.e., facilities management, core services, etc.).

2) Recovery: This phase focuses on the return to a pre-established level of functionality (plans should detail partial as well as full recoveries).
    a. Response teams implement recovery strategies and adhere to procedures and protocols outlined in Disaster Recovery Documents
3) Reconstitution: After recovery efforts are complete, normal operations must be restored. This may involve the reconstruction of facilities and/or infrastructure as well as the testing of restored elements to ensure their full functionality.

APPENDIX F: ***TSM Backup Service Standard Service Level Agreement*** **(2008)**
(Right click to open the Adobe Document Object located below)

TSM Backup Service

Standard
Service Level Agreement

University Of Michigan
Information Technology Central Services

7/1/2008

1.0 Overview

Service Level Agreements (SLAs) are between Information Technology Central Services (ITCS), which is the server provider, and the customers of that service. The TSM Backup Service is run by ITCS and is intended to support members of the University of Michigan community.

The service provides data backup and redundancy for computers over the network. It's scaled for larger scale users (minimum of 1 TeraByte). It supports a wide variety of platforms as well as some specialty applications such as databases and custom mail servers. The service is geared towards facilitating both the academic and research missions of the University. It provides greater data redundancy for server administrators and reduces their staff effort.

The primary attractions of the TSM Backup Service are easy client installs and on-demand user-based restores. The service optionally includes data compression, data encryptions, and data replication. Data is backed up to one of two locations, either the Arbor Lakes Data Facility (ALDF) or the Michigan Academic Computing Center (MACC), and is then replicated to the other for disaster recovery.

This document covers service details, expectations, roles, and responsibilities. The purpose of this document is to facilitate a partnership between ITCS and the customer as well as providing a framework for problem resolution and communication.

2.0 Pricing

The default customer rate for the service is $3,000.00 per TeraByte per year ($250 per month per TeraByte). Usage is rounded up to the largest TeraByte. For example, if you're using 1.2 TeraBytes, you'll be billed for 2 TeraBytes. The usage is measured and billed on a monthly basis. Multiple machines within the same unit or department can be aggregated together for a total, as long as they are all billed to the same Chartcom.

Special on-site consulting will be provided at an hourly rate. Customers with special needs (like additional off-site storage of tapes) or with more than 30

<u>APPENDIX G</u>: ITCS/ITCom *Customer Network Infrastructure Maintenance Standard SA* (2006)

(Right click to open the Adobe Document Object located below)

Customer Network Infrastructure Maintenance

Standard Service Agreement (SA)

University of Michigan
Information Technology Central Services (ITCS)
Information Technology Communications (ITCom)

July 1, 2008

1.0 Overview

This Service Agreement (SA) is between IT Communications (ITCS/ITCom) and a University of Michigan Unit (henceforth referred to as Unit) on the Ann Arbor Campus. Under this SA, ITCom agrees to provide the Unit, Network Infrastructure Maintenance to include data switches, routers, access points, hubs, Uninterruptible Power Supplies (UPS's), firewalls, and other identified and agreed upon components at the service rates and for the duration specified.

This SA also covers performance, reliability and other topics pertinent to maintenance; in particular, it lists the key responsibilities of ITCom and the Unit.

2.0 Purpose

The purpose of this SA is to establish a cooperative partnership between the Unit and ITCom by clarifying roles, setting rates and expectations, and providing mechanisms for resolving problems.

3.0 Service Rates

APPENDIX H: ***MACC Server Hosting Service Level Agreement*** **(Draft, 2009)**

(Right click to open the Adobe Document Object located below)

DRAFT **MACC Server Hosting Service Level Agreement** *DRAFT*

FY09: July 1, 2008 – June 30, 2009

1.0 Overview

This document and attachments constitute a Service Level Agreement (SLA) between the Michigan Academic Computing Center (MACC) and the Tenant listed in Appendix B. Under this SLA, the MACC agrees to provide hardware hosting at the data center per the rates specified in Appendix A.

2.0 Purpose

This agreement is intended to clarify roles, set rates and terms of the business relationship, and lists the key responsibilities of MACC and the Tenant.

3.0 Terms of Agreement

3.1 The terms of this SLA are listed in Appendix A.

3.2 The agreement should be renewed 60 days before the end of the term.

3.3 Modifications

- Proposed changes or modifications to this agreement must be submitted in writing to the other party to this agreement, and will not be considered effective until signed by both parties.
- The Tenant will be notified of rate changes approved by the policy committee as soon as possible prior to the start of the fiscal year in which they will be effective.
- Increases or decreases in the number of 4KW or 6KW units covered by this agreement must be done according to the procedures described below, as well as by modifying Appendix A.

3.3 Disasters

APPENDIX I: *Michigan Academic Computing Center Operating Agreement* (2006)

(Right click to open the Adobe Document Object located below)

Michigan Academic Computing Center Operating Agreement

Version of 22 December 2006

1. Description of Premises and Terms of Agreement

   1.1. The University of Michigan (U-M) and the Michigan Information Technology Center Foundation (MITC) are entering into a joint agreement commencing on the signing of this document and terminating on 30 January 2020. The purpose of this agreement is to operate a data center in the MITC building at 1000 Oakbrook Dr., Ann Arbor, MI.

   1.2. The premises are to be constructed and occupied as described in the document "University of Michigan Data Center" dated 30 January 2006, as modified by this and subsequent agreements and by approved construction change orders.

2. Governance

   2.1. The Michigan Academic Computing Center (MACC) will be operated by the Office of the Provost of the University of Michigan on behalf of U-M and MITC. In order to gather complete input from all occupants of the data center, the Provost, in concert with the MITC Board President, will form two advisory bodies, the Executive Advisory Committee and the Operations Advisory Committee. The committees' membership will be established within one month of the signing of this agreement.

      2.1.1. Executive Advisory Committee

         2.1.1.1. The Executive Advisory Committee will be charged with providing overall strategic directions and priorities, services and service levels to be provided in the facility, policies and procedures for requesting and allocating space in the facility, adjudicating disputes over shared space, funding and recharge models, and enhancement of the data center.

         2.1.1.2. The Executive Advisory Committee will consist of representatives from U-M schools, colleges, and major units as selected by the Office of the Provost; the CEO of Internet2 and the CEO of Merit Network representing MITC; one representative from the Office of the Provost; and one representative from MAV Development. The Data Center Coordinator will serve as staff to the Executive Advisory Committee. In cases where decisions cannot be reached by consensus, the Provost's Office representative, as the major tenant, will have final decision making authority.

1