From 2009-2010, the Center for Research Libraries performed an assessment of HathiTrust to evaluate it for compliance with the Trustworthy Repository Audit and Certification (TRAC) criteria. HathiTrust was certified by CRL in March 2011. The audit report is available on the CRL website. HathiTrust's response to the Minimum Required and additional elements of TRAC is below.

Review Of Compliance With Trustworthy Repositories Audit & Certification: Criteria And Checklist (TRAC)

Version 3.0, updated 3/22/2011

The HathiTrust Digital Library (HathiTrust) is a collaborative effort of Indiana University, the University of Michigan, and the University of California with support from the charter participating libraries of the Committee on Institutional Cooperation (CIC) and participation by other libraries (the University of Michigan and Indiana are hereunder referred to as the "Repository Administrators"; the University of Michigan is the "Host Institution", having primary responsibility for the operation, administration, and legal liability of the HathiTrust). HathiTrust is funded in large part by base funding (i.e., not grant or other one-time funding sources) from the Repository Administrators, with contributions from the charter participating libraries of the CIC, and with additional funding from institutions and consortia wishing to archive or sustain digital content in the repository ("Partnering Institutions"). The current version of this document represents its third public release. Although work on the document is ongoing, unless otherwise noted, we consider work on these criteria to be in compliance with the guidelines as set out in the TRAC Checklist.

[toc title=Contents; collapsed=false; list=ul]

A: Organizational Infrastructure

A1.1 Repository has a mission statement that reflects a commitment to the long-term retention of, management of, and access to digital information.

The mission of HathiTrust is to contribute to the common good by collecting, organizing, preserving, communicating, and sharing the record of human knowledge (see HathiTrust Mission and Goals).

A1.2 Repository has an appropriate, formal succession plan, contingency plans, and/or escrow arrangements in place in case

the repository ceases to operate or the governing or funding institution substantially changes its scope.

The Repository Administrators have funded HathiTrust Digital Library for an initial five-year period beginning January 2008, with a planned process of review and renewal. A review of the model used for funding and management of HathiTrust is scheduled for the third year of the first five-year period. This and subsequent processes of review provide HathiTrust with an opportunity to develop appropriate plans (e.g., succession) if necessary.

In cooperation with the Participating Libraries and in conjunction with that three-year review, HathiTrust has planned a constitutional convention for early 2011. In that process, HathiTrust will, in collaboration with the Participating Institutions, shape the next stage of governance for operating the repository through this partnership.

Currently, long-term curation of content in HathiTrust is part of the base-funded responsibilities of the University of Michigan Library, and ongoing funding for the Library is provided to that end. The Dean of Libraries reviews this funding and purpose with the Provost annually. Should the funding or organizational imperatives of the University of Michigan Library change, the Library will develop a succession plan and will devote multi-year funding to support of the transition to another host institution.

A2.1Repository has identified and established the duties that it needs to perform and has appointed staff with adequate skills and experience to fulfill these duties.

A diagram outlining the key roles in the HathiTrust functional framework is available at http://www.hathitrust.org/functional_framework. Staff from a variety of partner institutions fulfill these duties on behalf of, and for the benefit of, their institutions and the partnership as a whole. The HathiTrust partners see the need-based allocation of resources (as the needs of the institution and its interest in the partnership require) at a local level as an efficiency of the enterprise, enabled by the trust the partners have in one another and the stake that each of the partners has in the success of the collaboration.

Partners are able to determine the skills and competencies needed for a particular task or set of tasks and bring appropriate human and technical resources to bear. It is the responsibility of individual institutions to ensure that their staff have the requisite skills and training to meet the needs of their institution. HathiTrust benefits from this broad depth and pool of expertise when, to fulfill its own interests and those of the larger partnership, institutions provide local resources to accomplish common ends. This kind of resource allocation is evidenced in everything from the development of preservation specifications and end-user applications, to advice and implementation strategies regarding matters of policy, budget, technology, and innovation.

A2.2 Repository has the appropriate number of staff to support all functions and services.

As in A2.1, the individual interests of the partners and their common goals (as set out by the Executive Committee and guided by the Strategic Advisory Board) create an incentive for partners to allocate appropriate resources to meet intended individual and collaborative goals. The commitments of the repository are outlined in its mission statement and goals, the HathiTrust functional objectives, and partner contracts.

A3.1 Repository has defined its designated community/communities and associated knowledge base(s) and has publicly accessible definitions and policies in place to dictate how its preservation requirements will be met.

The primary designated community of HathiTrust is comprised of the faculty and students or other immediate constituencies (e.g., in the case of public research libraries) of the Partnering Institutions, as well as the institutions themselves, to the extent that their roles as stewards and disseminators of knowledge bring them into engagement with HathiTrust. By extension, meeting the needs of this designated community will ensure that HathiTrust meets the needs of higher education more generally. In “collecting, organizing, preserving, communicating and sharing [these records] of human knowledge” to our primary designated community, we will also be meeting many needs of the broader public.

Primary services that the archive provides are long-term preservation of the content held (both bit-level preservation and format migration) and support for an array of basic uses of that content, including:

- Persistence of object address (OAIS “Reference”)
- Reading (dependent on the user and his/her rights)
- Searching
- Assembling materials into private and public (i.e., shareable) collections

HathiTrust’s preservation policy and strategies for meeting preservation requirements are available at http://www.hathitrust.og/preservation.

See elsewhere in this report (particularly B1.4-1.5, B1.7, B4.1-4.2, C2.1-2.2) for documentation on mechanisms the HathiTrust archive employs to ensure long-term preservation of this content.

A3.2 Repository has procedures and policies in place, and mechanisms for their review, update, and development as the repository grows and as technology and community practice evolve.

Documentation of HathiTrust policies can be found at http://www.hathitrust.org/policies, and is provided throughout this document as well. As described in the HathiTrust partnership Features and Benefits, a formal review of repository governance and sustainability will be conducted in 2011, the third year of the initial 5-year period for which the HathiTrust partners have funded the

repository. This review will be conducted by partner institutions that joined HathiTrust on or before October 31, 2010.

A3.3 Repository maintains written policies that specify the nature of any legal permissions required to preserve digital content over time, and repository can demonstrate that these permissions have been acquired when needed.

HathiTrust is devoted to an array of archival and access services in support of the record of human knowledge. As such, all objects in the archive are either in the public domain, have the necessary permissions to support the level of access afforded, or are simply archived in such a way as to ensure an enduring copy of the content. HathiTrust provides reading access only to those publications where permitted by law or by the rights holder. In cases where a rights holder has granted HathiTrust permission to provide reading access to a publication, the administrative office of the University of Michigan Library retains a record of those permissions. The conclusions of all such determinations are registered in a rights database that controls access. All other forms of access are conducted in light of US copyright law and with the guidance of the University of Michigan's Office of the General Counsel. These policies are enumerated and explained in greater detail at http://www.hathitrust.org/copyright.

A3.4 Repository is committed to formal, periodic review and assessment to ensure responsiveness to technological developments and evolving requirements.

HathiTrust is committed to a formal self-review in the third year of every 5-year period of its operation. The first such review will be conducted in 2011 as mentioned in A3.2. HathiTrust has identified TRAC requirements, among other digital repository evaluation standards, as requirements it is striving to meet and maintain. HathiTrust engaged in a review process with DRAMBORA in the fall of 2008 and is being auditedcurrently by the Center for Research Libraries for compliance with TRAC. The results of the DRAMBORA review board were published as the following:

Seamus Ross, Andrew McHugh, Perla Innocenti, Raivo Ruusalepp: Investigation of the potential application of the DRAMBORA toolkit in the context of digital libraries to support the assessment of the repository aspects of digital libraries, Glasgow: DELOS NoE, August 2008, ISBN: 2-912335-41-8

A3.5 Repository has policies and procedures to ensure that feedback from producers and users is sought and addressed over time.

HathiTrust will respond to any request sent to a publicly-listed administrative email address support account within one business day. These responses may consist of requests for

clarification or more information; at the very least, they will acknowledge receipt of the original request.

Email requests submitted through HathiTrust feedback links are tracked in a helpdesk application, and assigned to appropriate staff members on the same business day of receipt.

HathiTrust has an active program of user testing, with the results of these tests influencing interface design and software functionality (see section C2.2). The Repository Administrators are active participants in the wider digital library community, and stay current with developments in technology, standards and best practices.

A3.7 Repository commits to transparency and accountability in all actions supporting the operation and management of the repository, especially those that affect the preservation of digital content over time.

HathiTrust is committed to transparency and accountability in all of its actions and endeavors. HathiTrust distributes a monthly newsletter to partners and publicly on the Web, outlining news and events, development efforts, outages, and other repository information. HathiTrust strives to place documentation relating to all aspects of the repository including technology, governance, policies and procedures, partnership, papers and presentations, and accountability (e.g. TRAC documentation) on its website. HathiTrust also strives to make as much repository content as possible freely available on the Web within the bounds of what the law permits.

As a public university, the University of Michigan (the current host institution for HathiTrust) is subject to Sunshine Laws and the Freedom of Information Act. The UM standard practice guide is available at http://spg.umich.edu/. Particularly relevant sections include 601.08-1, “Identification, Maintenance, and Preservation of Digital Records created by the University of Michigan”, and 608.12, “Institutional Data Resource Management Policy".

A3.8 Repository commits to defining, collecting, tracking, and providing, on demand, its information integrity measurements.

HathiTrust records actions that occur on objects, including digitization, ingest, fixity checks that occur on ingest, ingest validation, and other events that act on the content in any way from the time it enters the repository. This information is stored in PREMIS metadata in a METS file as part of the AIP (see Digital Object Specifications). HathiTrust performs additional integrity checks on all content on a quarterly basis. The results of these checks are logged and can be made available at any time.

A4.2 Repository has in place processes to review and adjust business plans at least annually.

A discussion of the budget and any related changes to process and operations (e.g., moving funds to meet specific needs) is included as a part of all Executive Committee meetings. No adjustments to business plans have been necessary to-date.

A4.3 Repository's financial practices and procedures are transparent, compliant with relevant accounting standards and practices, and audited by third parties in accordance with territorial legal requirements.

The budget of HathiTrust is a separately maintained "agency" budget, managed by the University of Michigan Library, and other financial components of the operation are represented in the budgets of several University of Michigan Library Information Technology (LIT) departments, including:

- Core Services: staff (e.g., system administration), hardware and maintenance.
- Digital Library Production Service: staff (e.g., most publicly available services).
- LIT administration: staff (e.g., coordination and review).

Other financial components, such as central accounting, exist elsewhere in the University of Michigan Library's budget structure. Most budget line items associated with support of the archive are identified as part of that activity, wherever practicable.

The University Library's financial procedures are subject to audits by the University of Michigan Office of University Audits. No cost elements of the archive currently rely on grants or charged-for services. All documented activities are subject to FOIA and may be reviewed with appropriate requests made to the University of Michigan's Freedom of Information and Policy Administration Coordinator.

A5.1 If repository manages, preserves, and/or provides access to digital materials on behalf of another organization, it has and maintains appropriate contracts or deposit agreements.

Deposit of content in HathiTrust is governed by agreements with partners that are kept on file. Agreements with partners that contribute content include a Digital Assets Submission Inventory (DASI) that designates the content for deposit and terms on which it is deposited. HathiTrust's partnership agreements can be found at http://www.hathitrust.org/join.

A5.3 Repository has specified all appropriate aspects of acquisition, maintenance, access, and withdrawal in written agreements with depositors and other relevant parties.

See A5.1.

A5.5 If repository ingests digital content with unclear ownership/rights, policies are in place to address liability and challenges to those rights.

The University of Michigan, which bears legal and fiscal responsibility for the repository, makes every effort to ensure that it has appropriate rights to ingest and provide access to content. In those cases where the rights are unclear (e.g., when we encounter copyright information that, relative to the work in hand, is ambiguous or contradictory), HathiTrust archives but does not provide access to the work. Where possible, the University of Michigan secures rights to use works that are in copyright. HathiTrust's policy governing procedures for responding to complaints is represented in the University of Michigan Library's Take-down Policy for addressing challenges to access rights. General policies on HathiTrust rights management can be found at http://www.hathitrust.org/copyright.

B: Digital Object Management

B1.1 Repository identifies properties it will preserve for each class of digital object.

HathiTrust is committed to preserving the intellectual content and in many cases the exact appearance and layout of materials digitized for deposit. In the case of print materials, HathiTrust stores and preserves metadata detailing the sequence of files for the digital object. HathiTrust relies on the extensive specifications on file formats, preservation metadata, and quality control methods included in the University of Michigan digitization specifications. HathiTrust is committed to bit-level preservation and format migration of materials created according to these specifications as technology, standards, and best practices in the digital library community change. More information can be found in HathiTrust's preservation policy.

As additional types of materials (such as audio) begin to be ingested into HathiTrust, significant aspects that should be preserved for each type will be identified.

B1.2 Repository clearly specifies the information that needs to be associated with digital material at the time of its deposit (i.e., SIP).

HathiTrust has a set of requirements for digital objects deposited in the repository that clearly identify the digital object, define the files associated with any given digital object (e.g., the page images that comprise a book), define the relationships between files (e.g., the association of text files with page images for any given page), and identify any technical or administrative metadata. HathiTrust has defined a METS profile for these metadata (see also B2.1). Digital objects must be accompanied by METS metadata that conforms to this profile; in cases where no METS metadata are provided, these metadata are created by HathiTrust in the process of ingest and validation. Examples of METS files and a description of HathiTrust's PREMIS implementation are available at http://www.hathitrust.org/digital_object_specifications. Specifications for bibliographic metadata and a description of the steps involved in ingest can be found at http://www.hathitrust.org/ingest. See also section B2.3.

B1.3 Repository has mechanisms to authenticate the source of all materials.

HathiTrust consists of digital content created by Partnering Institutions, their agents and external organizations. In all cases, a bibliographic record with associated digital object identifier is available in HathiTrust's administrative systems for each object ingested into HathiTrust. Additionally, for materials received from partner institutions, HathiTrust receives and retains a Digital Assets Submission Inventory.

B1.4 Repository's ingest process verifies each submitted object for completeness and correctness as specified in B1.2

The HathiTrust ingest process conducts the following tests on each submitted item:

- Metadata: internal tests to ensure MARC21 conformance and completeness
- OCR text: tested for well-formedness using JHOVE;
- Image files: tested for well-formedness using JHOVE;
- Metadata in image files: internal tests for consistency with conventions (see the HathiTrust Deposit Form and Guidelines);
- Digital signatures (MD5 checksums) for all OCR text and image files: checksum verification (see B1.1);
- Additionally: a one-to-one correspondence is ensured between OCR text and image files.

Note: these processes do not detect problems originating in capture (e.g., missing pages), nor do they detect readability or other subjective problems.

Detailed information about the requirements for ingested objects is available at http://www.hathitrust.org/ingest.

B1.5 Repository obtains sufficient physical control over the digital objects to preserve them

HathiTrust collects the following:

1. Partnership agreements and submission forms:
    a. HathiTrust documents its acceptance of archival responsibility for materials deposited by Partnering Institutions or external organizations in partnership and/or submission inventory agreements signed by appropriate authorities from that organization or institution. Signed partnership agreements (available at Eligibility and Agreements) and submission forms are filed with University of Michigan Library administration.
2. Workflow documents:
    a. The Michigan Digitization Project workflow diagram documents ingest procedures for content from internal digitization processes, vendor-supplied data, and data from both the Google digitization effort and local digitization activities.
    b. The HathiTrust ingest workflow diagram and notes document the steps of digital objects from the time they are ingested until the time they are made accessible to end-users.

3. Records of preservation events: As items are transformed after receipt, ingested, or when preservation operations are conducted, a record of these events and the date and time of their occurrence is maintained. The records use the PREMIS markup conventions and are embedded in the METS document corresponding to the item. Examples of HathiTrust's use of PREMIS are included in HathiTrust's Digital Object Specifications.

B1.6 Repository provides producer/depositor with appropriate responses at predefined points during the ingest processes

At each of the milestones listed below, logs with the following information (including error reports) are created and provided on request to the depositing organization:

1. Metadata. Bibliographic and item records are loaded into HathiTrust prior to ingest of content. At their request, partners receive reports of records that load successfully and records that failed to load;
2. Digital objects. Validation reports for all ingested content are available at http://www.hathitrust.org/ingest_logs. Additionally, when ingest of a given object is complete, its bibliographic item record is updated (see B1.7) and the object is added to an inventory of all repository content, located at http://www.hathitrust.org/hathifiles.
3. Rights metadata. At the time the bibliographic item record is updated, an automatic rights determination is performed on each object using the object's bibliographic data. The results of this determination are included in the repository inventory at http://www.hathitrust.org/hathifiles. Access to content is not provided without this rights determination step.

B1.7 Repository can demonstrate when preservation responsibility is formally accepted for the contents of the submitted data objects (i.e., SIPS)

After the ingest process is complete for a given item, the administrative database is updated with a special marker indicating that the item has been digitized, is stored in the repository, and is being preserved. The presence of this marker is an indication of formal acceptance of preservation responsibility.

B1.8 Repository has contemporaneous records of actions and administration processes that are relevant to preservation (Ingest: content acquisition)

The METS metadata for each digital object is updated to include information about the actions and processes performed on the object during and after ingest, including preservation actions with associated dates. These actions and processes are documented in conformance with the PREMIS Data Dictionary. Versions 1.0 and 2.0 of PREMIS are currently in use. A description of

HathiTrust's PREMIS implementation and examples of objects using both PREMIS 1.0 and 2. 0 are available at http://www.hathitrust.org/digital_object_specifications.

B2.1Repository has an identifiable, written definition for each AIP or class of information preserved by the repository.

A METS profile for Google-digitized content in HathiTrust is available at http://www.hathitrust.org/documents/hathitrust-mets-profile.xml. This document defines the AIP for Google-digitized content. We have not yet defined a profile for content from other digitization sources, but an example of an Internet Archive METS package and other digital object specification information can be found at http://www.hathitrust.org/digital_object_specifications. A detailed framework including content submission policies, guidelines, and specifications is available at http://www.hathitrust.org/ingest.

B2.3 Repository has a description of how AIPs are constructed from SIPs.

HathiTrust documents the modifications performed on SIPs in the creation of AIPs. Examples of this documentation are below.

- Creation of HathiTrust METS from Google SIPs
- Creation of Internet Archive METS from Internet Archive SIPs
- Creation of HathiTrust METS from Internet Archive METS

See our Deposit Guidelines and Digital Object Specifications for more information about HathiTrust's use of METS, and the paper "Building A Future By Preserving Our Past: The Preservation Infrastructure of HathiTrust Digital Library" for a full description of HathiTrust repository process and content packages.

Transformations may be performed on objects according to the policies and specifications outlined in the Deposit Guidelines, and the HathiTrust Deposit Form Sections III.B - Preservation description information requirements and III.C - Validation Requirements.

B2.4 Repository can demonstrate that all submitted objects (i.e., SIPs) are either accepted as whole or part of an eventual archival object (i.e., AIP), or otherwise disposed of in a recorded fashion.

HathiTrust records SIPs that are received and logs of which were ingested. Error logs are kept recording AIPs that were not ingested and reasons for failure. See http://www.hathitrust.org/ingest_logs.

B2.5 Repository has and uses a naming convention that generates visible, persistent, unique identifiers for all archived objects (i.e., AIPs).

Where possible (where identifiers have properties of uniqueness and integrity) the original identifier of an object is retained as a component of the HathiTrust identifier (a namepace prefix is pre-pended to all repository content for management purposes). If original identifiers are not unique or become non-unique when passed through repository processes (e.g., lowercasing), unique identifiers are selected using alternate available metadata, or new identifiers are created. Please see the Deposit Guidelines for more information.

HathiTrust uses the Handle service (http://handle.net) to assign permanent URLs to all content in the repository.

B2.6 If unique identifiers are associated with SIPs before ingest, the repository preserves the identifiers in a way that maintains a persistent association with the resultant archived object (e.g., AIP).

If identifiers of content received by HathiTrust conform to HathiTrust specifications, they are retained in the names of the files that compose the submitted digital object, the digital object's associated directory name, and the digital object ID in the HathiTrust METS file. The digital object ID is also saved in the bibliographic database record for each object. If identifiers do not conform to specifications, they are still maintained in the bibliographic database.

B2.7 Repository demonstrates that it has access to necessary tools and resources to establish authoritative semantic or technical context of the digital objects it contains (i.e., access to appropriate international Representation Information and format registries).

HathiTrust has selected a limited set of primary content formats for deposit in the repository. These include ITU G4 TIFF, JPEG2000, and JPEG image files, accompanied by derivative Unicode OCR text. It has selected these formats because they are widely held, widely supported standards for digital content, and limited their number in order to avoid the overhead of preserving multiple formats and managing access systems, software, tools, for those formats. The inclusion of new formats in the repository is done in close consideration with the tools and software necessary to validate, manage, and deliver them. If one or all of the format standards currently accepted should become deprecated or obsolete, there is expected, because of their wide adoption, to be sufficient warning to ensure that repository content and associated tools and software (for validation and rendering purposes, etc.) can be migrated safely to the new standards.

B2.8 Repository records/registers Representation Information (including formats) ingested.

See item B2.7 above. We have selected broadly accepted standards for repository content to mitigate the need for additional management resources.

B2.10 Repository has a documented process for testing understandability of the information content and bringing the information content up to the agreed level of understandability

In addition to rigorous validation and ensuring the bit-level integrity of items (see, for example, C1.7), the basic level of understandability for digital objects in HathiTrust is constituted minimally by the automated checks on the completeness and correct sequence of page images for any given volume. In addition, partners may perform manual review of completeness, legibility, order and general accuracy on digital content before it is contributed to HathiTrust (see HathiTrust Quality). In all cases, HathiTrust acts on reports from users about the content that we have online by responding to the user and making an attempt to correct errors.

HathiTrust provides several methods (e.g., e-mail links and report forms in the user interface) for people to report problems. Contact via any of these methods results in a response within one business day. Additionally, HathiTrust maintains an active program of usability testing in order to continuously improve access to materials in the repository (see C2.2).

B3.2 Repository has mechanisms in place for monitoring and notification when Representation Information (including formats) approaches obsolescence or is no longer viable.

See response in B2.7.

B4.1 Repository employs documented preservation strategies

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 TEI or METS). HathiTrust supports these formats because of their broad acceptance as preservation formats and because the formats are documented, open and standards-based, thus giving HathiTrust a means by which it can effectively migrate its contents to successive preservation formats over time, as necessary. The Repository Administrators have successfully performed such transformations in the past (see B4.2). In addition to reliance on standards and migration, HathiTrust employs the following preservation strategies:

- Bit-stream copying
- Refreshing
- Analog backup
- Replication
- Encapsulation
- Normalization

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 preservation formats. The ability to routinely access content provides another check on content integrity over time.

B4.2 Repository implements/responds to strategies for archival object (i.e., AIP) storage and migration.

HathiTrust stores AIPs in the repository file system. Each AIP is stored in a separate directory that includes all files and metadata associated with the digital object. Each AIP contains technical and administrative metadata (e.g., md5 checksums). Each object is described in an associated METS document, also stored in that directory (see B1.2 and B1.8). Maintaining uniformity of AIPs throughout the repository is an important strategy for storing and providing access to content in the repository, and enabling migration of content should the need arise.

Objects stored in the repository are in a restricted set of formats (see B4.1). 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). The University of Michigan, which is actively managing the repository, has migrated large SGML-encoded collections to XML, and Latin-1 character encodings to UTF-8 Unicode. This successful migration from older 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.

HathiTrust is using a one-time fee on new content submissions from partner institutions to build a capital fund towards large scale needs, including format migration or the need to move to a new storage platform.

B6.2 Repository has implemented a policy for recording all access actions (includes requests, orders etc.) that meet the requirements of the repository and information producers/depositors

Access to digital objects is currently made possible through two web-based systems. The PageTurner application allows users to view page images and search text of items in the repository or just search text, depending on the copyright status of the item. The Data API allows objects in the repository, with their associated metadata, to be retrieved by third-party systems. Actions using both mechanisms are logged. Access statistics for the PageTurner and its associated components are provided to partner institutions via Google Analytics. As specified in the HathiTrust Partnering Institution Contract, a formal request for withdrawal will consist of (a) deletion of the specified digital object(s), (b) the creation of a tombstone record, and (c) a time-and-materials basis for providing a copy of contents to the requesting institution.

B6.4 Repository has documented and implemented access policies (authorization rules, authentication requirements consistent with deposit agreements for stored objects).

Access to content by users, part of HathiTrust's preservation strategy, is governed by copyright law. HathiTrust provides reading access only to those publications where permitted by law or the rights holder. In cases where a rights holder has granted the University of Michigan Library

permission to provide reading access to a publication, the administrative office of the University of Michigan Library retains a record of those permissions. A permissions agreement is available online. Similarly, when partner institutions or organizations deposit materials in the archive, a signed Digital Assets Submission Inventory is filed with the University of Michigan Library administration.

Access policies are exercised using IP address detection, user authentication, and geography detection in conjunction with the determined copyright status of each item stored in the rights database (see PageTurner for a detailed description of access implementation). All other forms of access (e.g., computational research and access for users with print disabilities) are conducted in light of US copyright law and with the guidance of the University of Michigan's Office of the General Counsel.

A description of the access policies, an overview of implementation, and a full description of the rights database are available at http://www.hathitrust.org/copyright.

C: Technologies, Technical Infrastructure and Security

C1.2 Repository ensures that it has adequate hardware and software support for backup functionality sufficient for the repository's services and for the data held, e.g., metadata associated with access controls, repository main content.

Backup and restore functionality in HathiTrust is provided at the system level and consists of a) file system backup and b) database backup. Backup services are currently provided by Tivoli Storage Manager (TSM). All content in HathiTrust is backed up on a nightly basis.

C1.3 Repository manages the number and location of copies of all digital objects.

There are several tools we use to keep track of objects in the repository. These include our bibliographic database, bibliographic Solr index, rights database, and the file system itself. Location information is only available through the file system.

C1.4 Repository has mechanisms in place to ensure any/multiple copies of digital objects are synchronized.

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 SyncIQ software logs errors in synchronization, as well as deleted files. It also does a checksum validation on synchronized files. The University of Michigan does an additional check

on Isilon's synchronization. If a problem is found, staff at Michigan use rsync to verify the contents of the repository at Michigan and Indiana and correct any errors.

C1.6 Repository reports to its administration all incidents of data corruption or loss, and steps taken to repair/replace corrupt or lost data

Isilon software automatically detects and repairs corrupt portions of nodes that it detects in the storage cluster. The system is configured to send a report by email to system administrators should this occur. HathiTrust has implemented additional auditing processes, run on a quarterly basis, to validate the integrity of content using the MD5 checksums that accompany digital files on ingest.

C1.7 Repository has defined processes for storage media and/or hardware change (e.g., refreshing, migration).

HathiTrust replaces storage regularly, approximately every 3-4 years or as the usable life of storage equipment dictates. HathiTrust staff members conduct migrations from one storage medium to another (when required) using tools that validate checksums as files are transferred. (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.

In the future we plan to use storage virtualization to manage transitions to new media for us, with checksum validation on transferred files as described above.

C1.8 Repository has a documented change management process that identifies changes to critical processes that potentially affect the repository's ability to comply with its mandatory responsibilities.

HathiTrust has experienced rapid development since its inception in 2008. Changes that have been effected so far have been guided by the University of Michigan's experience in managing change in its earlier digitized collections, resulting in a strong push for consistency across repository content, conventions, policies, and procedures. The Repository is encapsulated in different systems with different operating dependencies and versions (e.g. Perl and different Perl modules, MySQL, Apache, XSL, Kakadu, etc.), but a general principle regarding change in the repository is that the closer to the content a system or standard is, the more stable it is (the less frequently changes are made). For example, while numerous adjustments have been made over time to systems such as user interfaces and ingest mechanisms (which are tracked as described in section C1.9), very little has been done to alter content metadata specifications or formats.

As HathiTrust begins to ingest content with greater variation, such as content from the Internet Archive and local digitization efforts, policies identifying critical processes and guiding the way

change is managed across the repository will need to be created (such documentation is already in process). This is an area of current (and future) work.

C1.9 Repository has a process for testing the effect of critical changes to the system.

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.

Additionally, subsets of production data are available in the development environment to allow developers to ensure proper system behavior before releasing changes to production.

C1.10 Repository has a process to react to the availability of new software security updates based on a risk-benefit assessment.

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.

For repository hardware, we receive security updates through the Linux Redhat network. We are able to look at a system interface, which contains a history of updates that have been done for each server.

C2.1 Repository has hardware technologies appropriate to the services it provides to its designated community(ies) and has procedures in place to receive and monitor notifications, and evaluate when hardware technology changes are needed.

The primary designated user community of HathiTrust is comprised of the faculty and students or other immediate constituencies (e.g., in the case of public research libraries) of the Partnering Institutions, as well as the departments and groups at these institutions involved in fulfilling and the archiving needs of those libraries. The hardware is consequently selected to ensure a minimum of outages and sufficient robustness to support a large number of

simultaneous users. Hardware is also selected to ensure easy expansion of storage and adaptability of other hardware needed for ingest of content.

HathiTrust staff upgrade hardware on a regular basis (every three to four years). To help detect to more rapid growth in demands, the web server and storage infrastructures have their own performance monitoring that indicate overload conditions.

C2.2. Repository has software technologies appropriate to the services it provides to its designated community(ies) and has procedures in place to receive and monitor notifications, and evaluate when software technology changes are needed.

HathiTrust has an active usability program that is involved in all aspects of design, creation, and modification of software systems and services for its designated user community (see item A3.1). With regard to software development in support of the archiving needs of the Partnering Libraries, HathiTrust focuses on the development of highly functional ingest and validation mechanisms. HathiTrust seeks and responds to guidance from the Strategic Advisory Board with regard to archiving services.

Usability activities for HathiTrust applications and interfaces are coordinated by a multi-institutional Usability Working group.The charge and membership of the group can be found at http://www.hathitrust.org/wg_usability_charge. The Usability Working group works closely with HathiTrust working groups and teams to design and test applications, recommend changes or enhancements, and coordinate user research. The group has played a crucial role in efforts such as the HathiTrust WorldCat Local Catalog, the HathiTrust information website, and improvements to the HathiTrust PageTurner. Ongoing reports about the Usability Working group's activities can be found in HathiTrust's monthly newsletter (http://www.hathitrust.org/updates).

Prior to June 2010, the University of Michigan Digital Library Production Service was responsibile for usability and user research activities surrounding HathiTrust interfaces.

C3.1 Repository maintains a systematic analysis of such factors as data, systems, personnel, physical plant, and security needs.

HathiTrust adheres to the information technology security policies of the University of Michigan Library. The University Library participates in a distributed organizational model where units across the University (of which it is one) have prime responsibility for planning and managing security within their units, coordinated by campus Information Technology Security Services (ITS). Documentation on the University of Michigan's security are linked below:

- MLibrary Information Security Plan
- MLibrary Information Security Plan Supplement

C3.2 Repository has implemented controls to adequately address each of the defined security needs.

HathiTrust follows widely accepted practices regarding system security including firewalls, patching, and limited administrative access. The University of Michigan Library's security projects and activities are listed in MLibrary Information Security Plan and Supplement (see C3.1 above).

C3.3 Repository staff have delineated roles, responsibilities, and authorizations related to implementing changes within the system

Authorization for repository content and servers is administered at four different levels:

1. Ability to authorize access to repository file system
    a. System Administrators
2. Able to read and write (make changes in) the repository (highest level of authorization):
    a. System administrators
    b. Core Developers (University of Michigan staff developers and other staff requiring full repository access)
3. Able to view content (no write permissions):
    a. Broader set of developers who have read access to repository content (e.g., developer for copyright review management system)
    b. Access for this level of authorization is controlled by system administrators
4. Advanced access to repository access system (e.g., those who are able to view copyrighted content):
    a. Usability and user interface specialists
    b. Content managers and system testers
    c. Copyright reviewers
    d. Some technical services staff
    e. Access for this level of authorization is controlled by core developers

C3.4 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).

HathiTrust consists of highly redundant storage located in two locations separated by several hundred miles (Ann Arbor, MI and Indianapolis, IN), and is backed up to tape at a third location several miles from the Ann Arbor data center.

A report on HathiTrust disaster preparedness was completed by an IMLS-funded intern in the summer of 2009. It is available on the HathiTrust website: "HathiTrust is a Solution: The Foundations of a Disaster Recovery Plan for the Shared Digital Repository".

HathiTrust is acting on the recommendations in this report and will make significant progress toward the development of a disaster recovery plan in 2011.