

Preserving the H-Net Email Lists: A Case Study in Trusted Digital Repository Assessment

Lisa M. Schmidt

Abstract

In 2007, the National Historical Publications and Records Commission (NHPRC) funded a two-year project to assess and improve the preservation of the academic email lists of the H-Net: Humanities and Social Sciences Online consortium using the Trustworthy Repositories Audit and Certification (TRAC): Criteria and Checklist. The project demonstrated that the TRAC could be applied to a repository functioning as a live access system, although the installation of a separate, server-based preservation environment is recommended. Tools used in this study, including a framework for documenting digital preservation policies, are applicable to the preservation of email list archives as well as more complex data sets.

As electronic mail, or email, has become a widespread, pervasive form of business and personal communication, email messages have become records of organizational and human activity that must be managed and sometimes preserved. Several projects exploring the long-term preservation of email have been conducted over the last decade. To date, however, there has been no formally documented research into the preservation of electronic mailing lists, a special case of email in which messages are distributed to multiple users.

Email lists exist for nearly every topic and subject area, including lists focused on academic subjects such as those of the H-Net: Humanities and Social Sciences Online consortium. In 2007, the National Historical Publications and Records Commission (NHPRC) awarded a grant to MATRIX: Center for Humane Arts, Letters and Social Sciences Online, a digital humanities research center at Michigan State University, to assess the H-Net email list archives as a trusted digital repository. MATRIX hosts the H-Net consortium on its servers.

© Lisa M. Schmidt.

The H-Net email lists provide a record of more than twenty years of academic discourse on humanities and social science research likely to be useful to future scholars and thus worthy of preservation.

Email and Preservation

Since the 1990s, when the commercialization of the Internet brought email into widespread business and personal use, it has been lauded as *the* “killer” software application. From the White House to large businesses and educational institutions to the home email account, millions of email messages are exchanged around the world every day. As records of communication as well as transactions deemed admissible as legal evidence, the individual email messages of businesses and other organizations must be managed, disposed of in a timely fashion, and, in some cases, preserved.

Email management has emerged as an area of interest and concern. Managing the Digital University Desktop, a joint University of North Carolina at Chapel Hill/Duke University project that ran from 2001 through 2005, recognized email as a “particularly problematic area” in electronic records management due to the ease of creating, transmitting, and storing large volumes of messages. The project studied how users manage email and developed email management and best practices guidelines.¹ The Joint Information Systems Committee (JISC) funded a pilot study to examine the management of email as institutional records and to develop retention and disposal policies and practices, especially as email messages are often the only records of transactions previously in paper form.² Many organizations have adopted policies and guidelines for the management of email, such as those developed by the Collaborative Electronic Records Project/Rockefeller Archive Center in 2008.³

Susan Lukesh discusses the need for the preservation of email as personal correspondence and informal scholarly communication. Digital technology now causes the “melding of the once different processes of creation, reproduction, and distribution...[and] makes preservation of email

¹ University of North Carolina at Chapel Hill, Managing the University Digital Desktop, “Project Background: Introduction and Goals,” <http://www.ils.unc.edu/digitaldesktop/Info/index.html>, accessed 6 September 2010.

² Michael Norris, “Institutional Records Management and E-mail Final Report” (November 2003), U.K. Web Archive, <http://www.webarchive.org.uk/wayback/archive/20070302174042/http://www.lboro.ac.uk/computing/irm/final-report.html>, accessed 23 August 2010.

³ Collaborative Electronic Records Project/Rockefeller Archive Center, *Records Retention and Disposition Guidelines* (November 2008), http://siarchives.si.edu/ceerp/RECORDS_RETENTION_SCHEDULE_rev3.pdf, 25 August 2010.

communication even more important as a trail of the development of ideas.”⁴ In a chapter on email curation in the *Digital Curation Manual*, Maureen Pennock also notes the need for preserving email messages of cultural, historical, and research value as well as those that have legal record status.⁵

The California Digital Library defines digital preservation as “the managed activities necessary for ensuring the long-term retention and usability of digital objects.”⁶ Digital preservation also includes strategic and organizational considerations related to the survival of digital material, as articulated in the 2002 report, *Trusted Digital Repositories: Attributes and Responsibilities*.⁷ Two major preservation strategies include migration or conversion of data into current or more accessible formats, and emulation, which uses modern hardware and software to re-create the original technology environment.⁸

Since Lukesh’s lament, researchers have taken several different approaches to the long-term preservation of email. Most major email preservation projects prefer solutions that incorporate the use of Extensible Markup Language (XML), an open standard established by the World Wide Web Consortium (W3C).⁹ XML is both human and machine readable and not tied to any specific hardware or software platform. Documents in XML can be self-describing, in that they can store information about their structure as well as content. Many file formats can be converted directly into XML, and it is a recommended archival format for text.^{10 11} XML may also be used as a “wrapper” that describes and contains files that make up a complex digital object, including the original

⁴ Susan Lukesh, “E-mail and Potential Loss to Future Archives and Scholarship or The Dog that Didn’t Bark,” *First Monday* 4, no. 9 (6 September 1999), <http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/692/602>, accessed 21 June 2010.

⁵ Maureen Pennock, “Curating E-mails: A Life-Cycle Approach to the Management and Preservation of E-mail Messages,” *Digital Curation Manual*, version 1.0 (July 2006), 10 and 14, <http://www.dcc.ac.uk/sites/default/files/documents/resource/curation-manual/chapters/curating-e-mails/curating-e-mails.pdf>, accessed 25 August 2010.

⁶ University of California, California Digital Library (CDL), “Glossary,” <http://www.cdlib.org/gateways/technology/glossary.html?field=institution&query=CDL&action=search>, accessed 6 September 2010.

⁷ Research Libraries Group (RLG), *Trusted Digital Repositories: Attributes and Responsibilities* (May 2002), <http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf>, accessed 22 December 2009.

⁸ Joint Information Systems Committee (JISC), “Digital Preservation Briefing Paper,” http://www.jisc.ac.uk/publications/briefingpapers/2006/pub_digipreservationbp.aspx, accessed 6 September 2010.

⁹ World Wide Web Consortium (W3C), “Extensible Markup Language (XML),” <http://www.w3.org/XML/>, accessed 25 August 2010.

¹⁰ National Association of State Chief Information Officers (NASCIO), “Electronic Records Management and Digital Preservation: Protecting the Knowledge Assets of the State Government Enterprise—Part III: Management Leads and Technology Follows—but Collaboration Is King!” (2007), <http://www.nascio.org/publications/documents/NASCIO-RecordsManagementPart3.pdf>, accessed 25 August 2010.

¹¹ Bronwyn Lee, Gerard Clifton, and Somaya Langley, “PREMIS Requirement Statement Project Report, Appendix 2: Recommended List of Supported Formats, Australian Partnership for Sustainable Repositories (APSR), National Library of Australia” (July 2006), 25.

file or files to be preserved, and related objects, such as versions of the file in other formats, information that describes the file (metadata), and access and preservation requirements.¹² Alternatively, XML may be implemented as a framework, or tree structure, that links the components rather than containing them.¹³

In 1999, the Collection-Based Long Term Preservation project of the San Diego Supercomputer Center (SDSC) experimented with the preservation of Usenet Newsgroup messages. One million plain text messages were tagged in XML and “ingested, archived, and dynamically rebuilt within a single day.”¹⁴ DAVID, a 2002 Belgian study,¹⁵ articulated four ways to archive email: printing and storing of hardcopies; central archiving through a mail server; electronic archiving within the mail system; or electronic archiving outside the email system, including the migration of each separate file to XML, PDF, or HTML format. It recommends the latter, with XML preferred.

A Dutch initiative, the Digital Preservation Testbed, tested various means of preserving email and published its research in 2003.¹⁶ This study assessed three digital preservation strategies and their suitability for preserving email: migration; the use of XML; and emulation. It determined XML to be the best preservation strategy for email, both for creation and conversion of messages and for use as a framework to link the components deemed necessary for long-term preservation. These components include the original email transmission file; the XML file created from the transmission file; metadata extracted from the transmission file; a log file containing information about preservation actions taken; and possibly additional files containing the body text of the message and attachments in their original formats.

The Digital Preservation Testbed project team also developed XMaiL, an open-source software module that can mark up email messages in XML for use in a preservation framework.¹⁷ Likewise, the National Archives of Australia (NAA) developed the open-source XML Electronic Normalising of Archives (XENA) tool that converts email messages into XML for long-term preservation. Another

¹² Digital Preservation Testbed, “From Digital Volatility to Digital Permanence: Preserving Email” (The Hague, April 2003), 34. Paper in possession of author.

¹³ Digital Preservation Testbed, “From Digital Volatility to Digital Permanence,” 37.

¹⁴ Reagan Moore, Chaitan Baru, Amaranth Gupta, Bertram Ludaescher, Richard Marciano, and Arcot Rajasekar, “Collection-Based Long-Term Preservation,” San Diego Supercomputer Center (June 1999), 35, <http://www.sdsc.edu/NARA/Publications/nara.pdf>, accessed 22 December 2009.

¹⁵ Filip Boudrez and Sofie Van den Eynde, “DAVID—Archiving E-mail,” Stadsarchief Stad Antwerpen (August 2002), <http://www.expertisecentrumdavid.be/davidproject/teksten/Rapporten/Report4.pdf>, accessed 25 August 2010.

¹⁶ Digital Preservation Testbed, “From Digital Volatility to Digital Permanence,” 29–37.

¹⁷ Pennock, “Curating E-mails,” 49–50. As of 25 August 2010, XMaiL no longer appeared to be available for download.

open-source tool, the Digital Preservation Recorder (DPR), records the meta-data associated with the XENA conversion process.¹⁸

Two recent email preservation research projects focused on XML-based solutions for preserving email: the NHPRC-funded Preservation of Electronic Mail Collaboration Initiative, a collaboration among the state archives of North Carolina, Kentucky, and Pennsylvania;¹⁹ and the Collaborative Electronic Records Project (CERP), a partnership between the Smithsonian Institution and Rockefeller Archives Center.²⁰ The partners worked together to develop a tool that converts all of the email messages in one email account, along with their associated metadata and any attachments, from the generic MBOX format for email into a single XML file for long-term preservation. Email messages in proprietary formats such as Microsoft Outlook, Lotus Notes, and Novell Groupwise are converted into the MBOX format before the conversion of an account into XML.

Electronic Mailing Lists

An electronic mailing list is a special case of email that allows for the widespread distribution of email to multiple subscribed users. Computer scientist Eric Thomas developed LISTSERV, the first automated electronic mailing list software application, in 1986. Before LISTSERV, human list administrators had to manage email lists manually.²¹ Email lists based on LISTSERV and other automated email list software are now available for nearly every topic of interest. The H-Net email lists and JISCMail, a U.K.-based service designed to facilitate communication in the academic community, both use LISTSERV.²² No documented research on the preservation of email lists created by LISTSERV or other automated email list management software appears to have occurred before the H-Net email list preservation project.

¹⁸ National Archives of Australia (NAA), "Tools for Digital Preservation," <http://www.naa.gov.au/records-management/preserve/e-preservation/at-NAA/software.aspx>, accessed 4 February 2011.

¹⁹ North Carolina State Archives, Preservation of Electronic Mail Collaboration Initiative, <http://www.records.ncdcr.gov/EmailPreservation/>, accessed 22 December 2009.

²⁰ Rockefeller Archives Center and Smithsonian Institution, The Collaborative Electronic Records Project (CERP), <http://siarchives.si.edu/cerp/index.htm>, accessed 22 December 2009.

²¹ L-Soft, "History of LISTSERV," <http://www.lsoft.com/products/listserv-history.asp>, accessed 30 October 2009. Note that Thomas's creation was a paradigm-shifting improvement over the manually administered electronic mailing list for the BITNET computer network, also called LISTSERV, that had been in use since 1984.

²² JISCMail, "What Is JISCMail?," <http://www.jiscmail.ac.uk/about/whatisjiscmail.html>, accessed 26 August 2010.

H-Net and the Project to Preserve the List Archives

H-Net: Humanities and Social Sciences Online is an international consortium of scholars and teachers with a mission to “create electronic networks and resources dedicated to advancing research, teaching, learning, public outreach, and professional service within their own specialized areas of knowledge.”²³ The oldest collection of born-digital, content-moderated arts, humanities, and social science materials on the Internet, H-Net began in 1992 as a virtual service hosted at the University of Illinois at Chicago. Content includes *H-Net Reviews*, the largest online scholarly book review journal in the world; job and meeting announcements; and more than 180 free, public academic “networks,” or email lists. A twelve- to seventeen-member council governs the policies and activities of the H-Net consortium. The H-Net Council is elected from the membership and includes H-Net’s executive director.²⁴

At the heart of the H-Net consortium, the academic email lists cover a wide range of humanities and social science topics. More than 450 editors and 130,000 subscribers participate in the H-Net networks, and, on average, some 5,000 posts are made to the public lists each month. An estimated 230,000 messages were viewed during the last week of April 2009 alone. In addition to its public lists, H-Net includes more than 230 “private” lists used by editors, board members, and administrators for planning, testing, and advisory purposes. There are more than one million email messages in the H-Net list archive, and that number continues to grow.

MATRIX: Center for Humane Arts, Letters, and Social Sciences Online, a digital humanities research center at Michigan State University, has hosted H-Net since 1994. As a research organization “devoted to the application of new technologies in teaching, research, and outreach,” MATRIX “creates and maintains digital libraries of humanities and social science materials, provides training in computing and new teaching technologies, and creates forums for the exchange of ideas and expertise.” In addition to the H-Net scholarly community, MATRIX hosts the major digital library repositories African Online Digital Library (AODL), Detroit Public Television’s American Black Journal video archives, Historical Voices, and the Quilt Index.²⁵

In 2007, MATRIX received a grant from the National Historical Publications and Records Commission (NHPRC) to assess existing preservation practices for

²³ H-Net: Humanities and Social Sciences Online, “H-Net Mission Statement” (March 2000), <http://www.h-net.org/about/mission.php>, accessed 7 December 2007.

²⁴ H-Net: Humanities and Social Sciences Online, “H-Net Constitution” (November 2003), <http://www.h-net.org/about/constitution.php>, accessed 26 August 2010.

²⁵ MATRIX: Center for Humane Arts, Letters, and Social Sciences Online, <http://www2.matrix.msu.edu/about/>, accessed 11 November 2009.

the H-Net email list archives and to develop and implement an improved long-term preservation plan. The H-Net lists represent a compilation of years of academic discourse. Preserving the lists would ensure their future availability to readers of scholarly research and as a resource to provide deeper understanding of the context and evolution of the represented fields.

NHPRC and MATRIX believed that this research on current and suggested preservation practices for the born-digital H-Net email lists might be useful to archivists and others who manage email lists and other large collections of electronic records. The assessment would include the use of the Trustworthy Repositories Audit and Certification (TRAC): Criteria and Checklist, guidelines on trusted digital repositories published by the Online Computer Library Center (OCLC) and Center for Research Libraries (CRL).²⁶

TRAC as Preservation Repository Evaluation Tool

In 1996, the Task Force on Archiving of Digital Information noted the need for a process to certify digital archives as trustworthy preservation environments, laying the foundation for the development of the TRAC checklist.²⁷ The Open Archival Information System (OAIS) reference model²⁸ standard approved in 2002 provided a common conceptual framework for the long-term preservation of digital material; however, it described system responsibilities at a high level with no criteria for measuring compliance.²⁹ In May 2002, the Research Libraries Group (RLG) and OCLC published *Trusted Digital Repositories: Attributes and Responsibilities*, a seminal report that identified the need for certification of digital repositories to demonstrate that an organization meets standards for storing, migrating, and providing access to digital collections.³⁰ NARA and RLG established a Task Force on Digital Repository Certification in 2003 to develop an objective, prescriptive methodology to establish the trustworthiness of a digital repository based on the earlier report. The first version of the TRAC document was published in February 2007.

An external third-party organization may use TRAC to assess and certify a repository, or it can be used for internal self-assessment, as MATRIX did with the

²⁶ The Center for Research Libraries (CRL) and Online Computer Library Center, Inc. (OCLC), "Trustworthy Repositories Audit and Certification: Criteria and Checklist," version 1.0 (February 2007), <http://www.crl.edu/PDF/trac.pdf>, accessed 22 December 2009.

²⁷ CRL and OCLC, "Trustworthy Repositories Audit and Certification," 1.

²⁸ Consultative Committee for Space Data Systems (CCSDS), "Reference Model for an Open Archival Information System (OAIS)," *Blue Book* 1, Issue 1 (CCSDS Secretariat, January 2002), <http://public.ccsds.org/publications/archive/650x0b1.pdf>, accessed 22 December 2009.

²⁹ CRL and OCLC, "Trustworthy Repositories Audit and Certification," 1.

³⁰ RLG, *Trusted Digital Repositories*.

H-Net email list archives. The checklist consists of eighty-four audit criteria organized into three sections: “Organizational Infrastructure”; “Digital Object Management”; and “Technologies, Technical Infrastructure, and Security.” Each section is divided into subsections of audit criteria to compare to current local capabilities. This comparison is known as a “gap analysis,” the difference between the current state and the desired state of a criterion governing the trustworthiness of a repository. Once the gap has been identified, strategies may be formulated to narrow it. As improvements are made, the gap narrows with each iteration of applying the TRAC.³¹

Applicability of criteria varies by institution, and not all criteria apply to all repositories. The TRAC also includes examples of documentation and other evidence that prove support of the criteria, and an appendix lists minimum required documents that will satisfy multiple criteria.

TRAC audits have been performed on some repositories. For example, CRL reports on third-party TRAC audits of the Lots of Copies Keep Stuff Safe (LOCKSS) and Inter-University Consortium for Political and Social Research (ICPSR) repositories, published respectively in 2007 and 2006, identified strengths of the repositories as well as areas that required improvement. Although the audit acknowledged that the LOCKSS software was a “solid technology” and that the LOCKSS collaborative digital preservation service was a benefit to smaller institutions, it illuminated potential user access issues under certain circumstances.³² ICPSR rated favorably in terms of economic sustainability and its data deposit, ingest, and preservation processes, with the audit recommending the development of policies and documentation for the repository.³³ The MetaArchive Cooperative, a preservation service for digital assets of universities, libraries, museums, and other cultural heritage institutions, conducted a self-assessment in 2010 that demonstrated its conformance to the TRAC criteria.³⁴

As a self-assessment, the TRAC audit for the H-Net email list archives was similar to the MetaArchive’s audit; in contrast, CRL performed third-party audits on LOCKSS and ICPSR. Unlike MetaArchive and the archives audited by

³¹ Anne R. Kenney, Nancy Y. McGovern, et al., “Digital Preservation Management: Implementing Short-term Strategies for Long-term Problems,” tutorial and workshop, Cornell University Library, version 1 (2003), Inter-University Consortium for Political and Social Research (ICPSR), University of Michigan, <http://www.icpsr.umich.edu/dpm/>, accessed 22 December 2009.

³² Robin Dale, *LOCKSS Audit Report*, Center for Research Libraries Auditing and Certification of Digital Archives Project (November 2007), http://www.crl.edu/sites/default/files/attachments/pages/LOCKSS_Audit_Report_11-07.pdf, accessed 21 December 2009.

³³ Robin Dale, *ICPSR Audit Report*, Center for Research Libraries Auditing and Certification of Digital Archives Project (24 October 2006), http://www.crl.edu/sites/default/files/attachments/pages/ICPSR_final.pdf, accessed 21 December 2009.

³⁴ Matt Schultz, *MetaArchive Cooperative TRAC Audit Checklist* (April 2010), http://www.metaarchive.org/sites/default/files/MetaArchive_TRAC_Checklist.pdf, accessed 26 August 2010.

CRL, the H-Net email list archives only holds H-Net messages and related metadata rather than deposits from other institutions, allowing MATRIX more latitude in determining acceptable levels of compliance to TRAC criteria. The H-Net list archives also differ from these other archives in that it is a live access system rather than a separate preservation repository.

The H-Net Email List Preservation Project

MATRIX received the grant funding for the H-Net email list preservation project in early 2007, and work formally began on the project after the hiring of Lisa Schmidt as electronic records archivist that October. Schmidt consulted with the systems administrator for H-Net to learn how the system worked and how it fit into the operations and technology infrastructure at MATRIX, including backup and security planning. In February 2008, she conducted an initial TRAC evaluation of the system. Through the summer of 2008, she developed administrative and technical improvement recommendations for narrowing the preservation “gaps” identified through the TRAC analysis, again in consultation with the systems administrator. Over the next several months, Schmidt researched, created, and documented digital preservation policies while the systems administrator made the recommended technical changes to the H-Net system and processes. Schmidt then performed a second TRAC analysis in July 2009. Throughout the project, she consulted with an archival advisory board of academics and practitioners, and kept them apprised of developments.

How the H-Net Lists Work

Before conducting the first TRAC analysis, Schmidt needed to learn about how users post to and access the H-Net lists as well as back-end H-Net systems operations. The following description, based on research on H-Net administration and consultations with the systems administrator, details the H-Net list archives workflow and back-end operations at the beginning of the project.

The H-Net email lists run on LISTSERV list administration software, and many of the processes of the H-Net lists follow basic LISTSERV functionality. MATRIX has also customized some aspects of the software and developed a Web-based interface to better serve the needs of the H-Net community.

One or more experts in a given subject area act as editors and moderate the H-Net list dedicated to that subject. For example, British Isles scholars edit the H-Albion list, a discussion network for British and Irish history.³⁵ List subscribers

³⁵ H-Net: Humanities and Social Sciences Online, “H-Albion: The H-Net Discussion List for British and Irish History,” <http://www.h-net.org/~albion/>, accessed 21 December 2009.

as well as nonsubscribers may post to most lists, although some lists will reject messages from nonsubscribers. To subscribe, most lists require only a simple username and password signup; some also require that the subscriber fill out a survey form. Editors, subscribers, and other users send messages and otherwise interact with the H-Net lists through standard LISTSERV email commands or a Web-based interface developed by MATRIX. Users with administrative privileges may also create and add lists.

Messages sent to the H-Net public lists must be written in a plain text character set such as ASCII or Unicode, and attachments are not allowed to minimize virus risks.³⁶ The private administrative lists include some messages with attachments, most in common formats such as Microsoft Word (.doc), Microsoft Excel (.xls), Microsoft PowerPoint (.ppt), PDF (.pdf), and image formats such as JPEG (.jpg).

When a subscriber sends a message to an H-Net list, it is forwarded to an editor for approval. Approved messages are then forwarded back to LISTSERV, which processes them for posting. While the editor may edit the message for content, sometimes working with the original author to develop a final posting, he or she often simply approves and posts the message as it was received.³⁷ The editor also has the option to ignore the message, in which case it will expire after forty-eight hours (see Figure 1). All of the public lists require confirmation from the editor that the forwarded posting came from that editor's address, providing protection against spam and spoofed addresses. In contrast to the strict protocol of the public H-Net lists, some private lists permit subscribers to post messages without going through an editor.

If an editor makes any changes to a message, it essentially becomes a new message, and the author and date written (creation date) metadata are overwritten to reflect the editor's name and the current date. To make up for this unfortunate, provenance-destroying LISTSERV "feature," the editor may manually re-enter the original message's author, date, and subject before posting. This method of preserving the original metadata is labor intensive and prone to error, however, and the editors of one-third of the lists do not perform such updates. Regardless of whether a list editor chooses to enter a message's original metadata, most authors include email signature blocks in their messages with their name and other information, such as their title and the name of their institution. Researchers can therefore still find messages written by particular authors through the H-Net lists' full-text search capabilities, even if they cannot browse through a list of messages by author names.

³⁶ H-Net: Humanities and Social Sciences Online, "H-Net By-Laws," Section 2.03, Guidelines on Posting and Subscribing (d) (vi), (March 2000, amended November 2003), <http://www.h-net.org/about/by-laws.php>, accessed 22 December 2009.

³⁷ H-Net, "H-Net By-Laws."

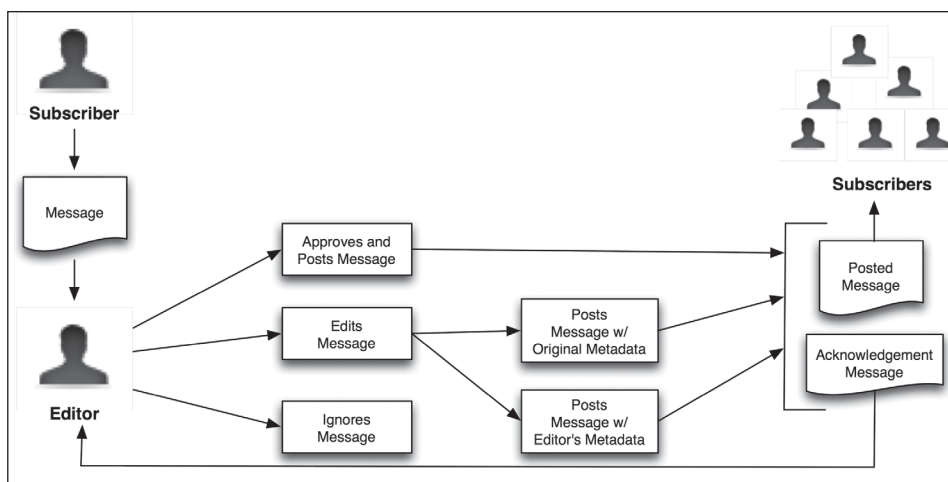


FIGURE 1. H-Net message posting process.

The message approval process also functions as a means of appraising the contents of the H-Net lists for archiving purposes. If an editor approves a message—a form of peer review—he or she considered it part of the academic discourse in that subject area and of current and future interest to researchers. Therefore, all messages approved and posted by list editors are permanently archived and considered worthy of preservation.

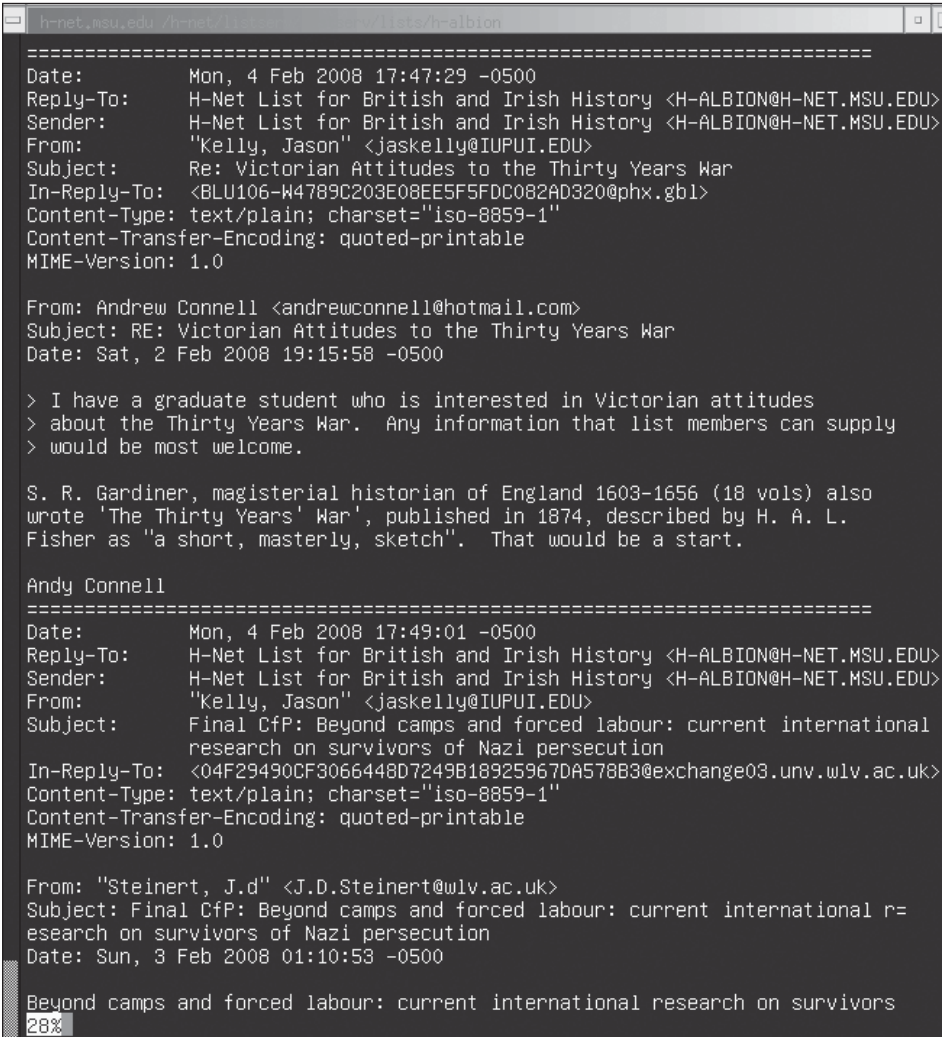
The LISTSERV software arranges approved, posted messages into flat text files known as “notebooks.” H-Net systems administration has set up the LISTSERV template so that a single notebook includes messages posted to a particular list during a seven-day time period. The messages concatenate in the notebook as they post—that is, each successive message writes to the file after the one that preceded it—until the period ends. Messages thus remain stored in their original order of posting (see Figure 2). Most of the descriptive metadata for messages are automatically generated on creation or posting, with the “subject” added by the original author.

Notebooks are named according to the time periods they cover, with days 1–7 of a given month as time period “a,” days 8–14 as time period “b,” and so on, with an extra time period (“e”) for months having twenty-nine, thirty, or thirty-one days. These periods become part of the notebook file name. For example, a notebook with the name “h-africa.log0802a” would be in the H-Africa lists and would include postings from the first seven days (a) of August 2002. At the end of a given time period, a new notebook file is started.

Every twenty-four hours, the newest messages in the current notebook file are copied to a proprietary bibliographic retrieval services (BRS) database,

where they are available for full-text search through an application created by MATRIX software programmers. This search application also assembles links to a custom browse application, enabling display of messages searched for by an H-Net user.

The browse application also reads notebook messages, extracts key meta-data, and generates MD5 hashes for each message up to seven days after the last message posts to a given notebook. MD5 is a message digest algorithm, or cryptographic hash function, used to verify the integrity, or “fixity,” of digital files. Each file has a unique MD5 hash.



```

h-net.msu.edu /h-net/.../lists/h-albion
=====
Date:      Mon, 4 Feb 2008 17:47:29 -0500
Reply-To:  H-Net List for British and Irish History <H-ALBION@H-NET.MSU.EDU>
Sender:    H-Net List for British and Irish History <H-ALBION@H-NET.MSU.EDU>
From:      "Kelly, Jason" <jaskelly@IUPUI.EDU>
Subject:    Re: Victorian Attitudes to the Thirty Years War
In-Reply-To: <BLU106-W4789C203E08EE5F5FDC082AD320@phx.gbl>
Content-Type: text/plain; charset="iso-8859-1"
Content-Transfer-Encoding: quoted-printable
MIME-Version: 1.0

From: Andrew Connell <andrewconnell@hotmail.com>
Subject: RE: Victorian Attitudes to the Thirty Years War
Date: Sat, 2 Feb 2008 19:15:58 -0500

> I have a graduate student who is interested in Victorian attitudes
> about the Thirty Years War. Any information that list members can supply
> would be most welcome.

S. R. Gardiner, magisterial historian of England 1603-1656 (18 vols) also
wrote 'The Thirty Years' War', published in 1874, described by H. A. L.
Fisher as "a short, masterly, sketch". That would be a start.

Andy Connell
=====
Date:      Mon, 4 Feb 2008 17:49:01 -0500
Reply-To:  H-Net List for British and Irish History <H-ALBION@H-NET.MSU.EDU>
Sender:    H-Net List for British and Irish History <H-ALBION@H-NET.MSU.EDU>
From:      "Kelly, Jason" <jaskelly@IUPUI.EDU>
Subject:    Final CfP: Beyond camps and forced labour: current international
            research on survivors of Nazi persecution
In-Reply-To: <04F29490CF3066448D7249B18925967DA578B3@exchange03.unv.wlv.ac.uk>
Content-Type: text/plain; charset="iso-8859-1"
Content-Transfer-Encoding: quoted-printable
MIME-Version: 1.0

From: "Steinert, J.d" <J.D.Steinert@wlv.ac.uk>
Subject: Final CfP: Beyond camps and forced labour: current international r=
            esearch on survivors of Nazi persecution
Date: Sun, 3 Feb 2008 01:10:53 -0500

Beyond camps and forced labour: current international research on survivors
28%

```

FIGURE 2. H-Net notebook file, showing concatenation of messages posted in original order.

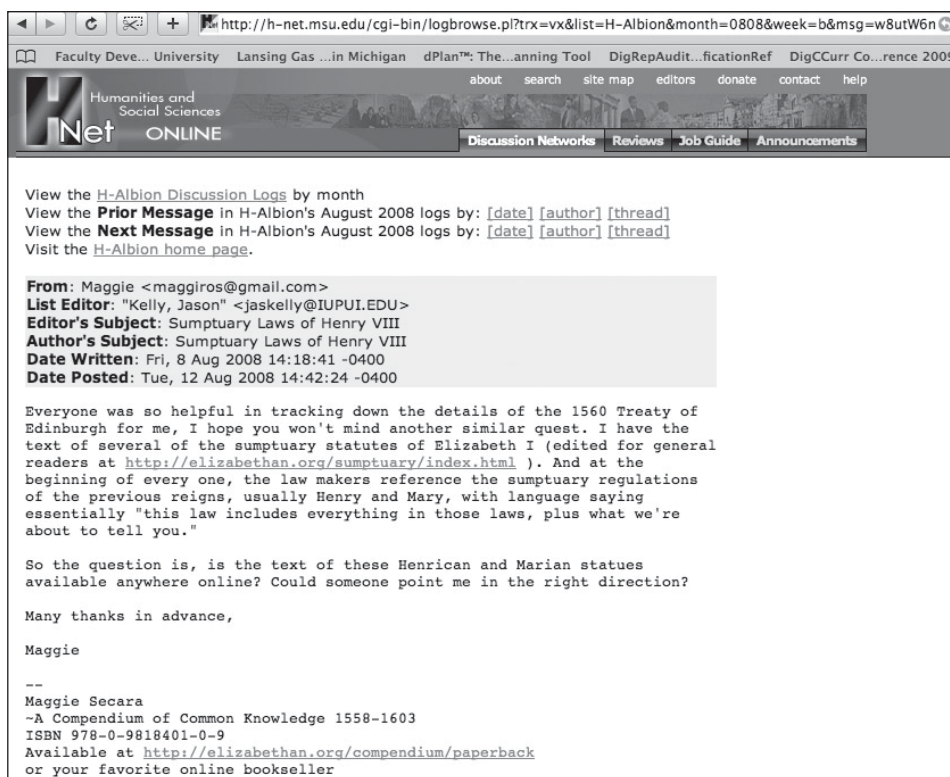


FIGURE 3. H-Net message retrieval view.

A script writes the metadata—including the MD5 hashes—to a database, enabling more efficient browsing and message retrieval. Metadata for each message include

- filename—name of notebook file where message is stored, such as h-africa.log0802a
- offset—byte position in notebook file where message is stored
- from—name and email address
- subject
- dpb—date posted
- cbd—date in a different format for sorting purposes
- messageid—MD5 hash

When a user visits the H-Net website,³⁸ browses a list, and clicks on a message to view it, the browse application pulls the message from the original notebook file, builds a Web address (URL) for the message, and transforms it into HTML for viewing in the browser. The URL is a combination of the message's filename and MD5 hash, as shown in Figure 3. This serves as a persistent

³⁸ H-Net: Humanities and Social Sciences Online, <http://www.h-net.org>, accessed 21 December 2009.

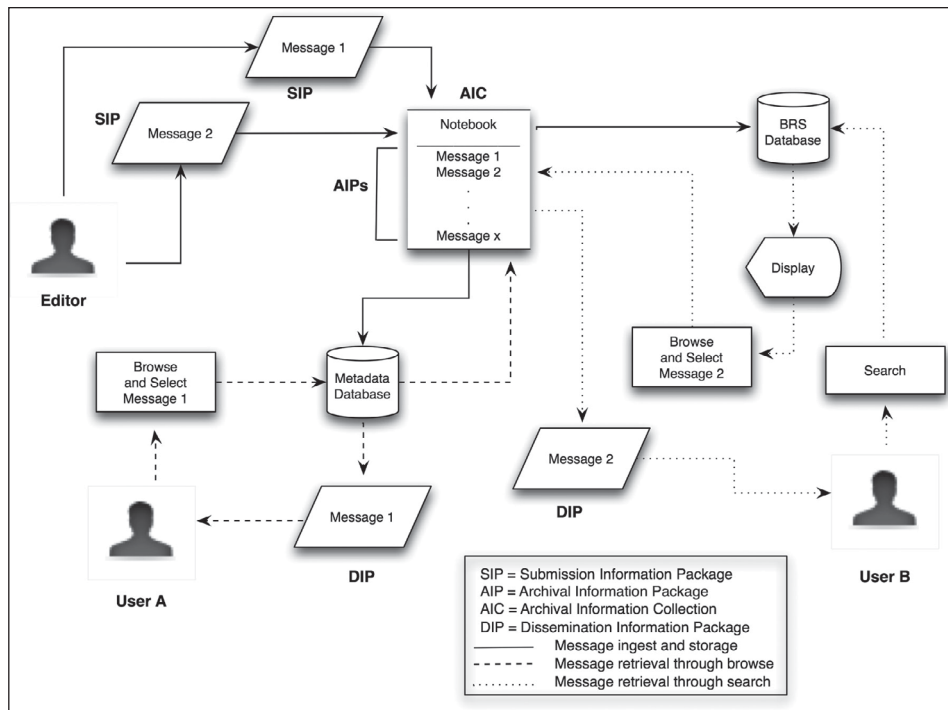


FIGURE 4. H-Net message ingest, storage, and retrieval processes mapped to the OAIS model.

identifier for the message that can be bookmarked for reference and citation purposes.

Subscribers to some of the private lists may log in and similarly access messages in a browser view. For most private lists, however, subscribers may only access logs using *LISTSERV* commands. Attachments may be embedded in private list messages.

The H-Net ingest, storage, and retrieval processes described above map to the OAIS reference model³⁹ as shown in Figure 4. Messages submitted by the editors are the Submission Information Packages (SIPs). After submission, the messages become Archival Information Packages (AIPs). They are stored in the notebooks, which may be considered Archival Information Collections (AICs). When a user browses and selects a message, the page view received is the Dissemination Information Package (DIP). A user may also receive a DIP by searching the BRS database, as it pulls selected messages from the notebooks.

³⁹ CCSDS, "Reference Model for an Open Archival Information System."

MATRIX Backup and Security

Schmidt also gathered information about the technology infrastructure of MATRIX, especially backup and security, before conducting the first TRAC assessment. As MATRIX hosts H-Net and its discussion networks, H-Net relies on MATRIX to back up the email list archive and keep it secure. MATRIX stores approximately three terabytes of data, including the H-Net software, message logs, and databases, on its servers. While no formal plan for adding storage capacity is in place, it is understood that more storage space will be acquired as needed. A rack containing the servers is kept in a climate-controlled, physically secured room on the MATRIX premises. All of the MATRIX servers run the Debian distribution of the Linux operating system.

Systems administrators perform incremental backups to Linear Tape Open (LTO) magnetic data storage tapes on a daily basis and a full backup weekly. An open-format, high-capacity tape storage technology developed by Hewlett-Packard, IBM, and Certance (now part of Quantum), LTO is widely used for digital backups.⁴⁰ At the time of the audit, MATRIX used NetVault: Backup software for all tape backups. Although proprietary, files backed up using NetVault: Backup may be read using standard Unix tools.

MATRIX backup tapes are stored at another secured location on the Michigan State University campus. The tapes cycle through the backup system approximately every six weeks. To further ensure against data loss, MATRIX systems administration performs an additional full tape backup approximately every four months. At the time of the original assessment, the MATRIX systems administrator planned to keep the tapes of the additional backup “permanently” and stored them in a cabinet in a minimally secured room on the MATRIX premises.

A good chain of communication ensures that systems administrators are apprised of any systems-related activity. MATRIX executive staff and several key technical managers with root system accounts have access to the H-Net and MATRIX systems.

First TRAC Assessment

Schmidt conducted the initial assessment of the H-Net email list archive in February 2008.⁴¹ Based on interviews with the H-Net systems administrator,

⁴⁰ SearchStorage.com Definitions, “What Is Linear Tape-Open?,” http://searchstorage.techtarget.com/sDefinition/0,,sid5_gci347603,00.html, accessed 7 September 2010.

⁴¹ Lisa M. Schmidt, “Trustworthy Repositories Audit and Certification: Criteria Checklist” (March 2008), <http://www.h-net.org/archive/documentation/TRAC%20current%20publish.pdf>, accessed 22 December 2009.

the H-Net associate director, and the MATRIX office administrator, and on informal conversations with other H-Net and MATRIX staff, the assessment revealed that the H-Net system conformed in part to the TRAC criteria. For example, TRAC criterion C1.1, subsection "System Infrastructure" in the "Technologies, Technical Infrastructure, and Security" section states: "Repository functions on well-supported operating systems and other core infrastructural software."⁴² As noted, the servers hosting H-Net run on the Debian distribution of Linux, an open-source operating system used by many different types of organizations and thousands of individuals.⁴³ MATRIX had satisfied this criterion at the time of the original assessment.

Large preservation gaps existed in other areas, however. For example, TRAC criterion A1.2, subsection "Governance and organizational viability," under the "Organizational Infrastructure" section states: "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."⁴⁴ MATRIX had not established a succession plan for the H-Net list archive. Other problems exposed in the course of the evaluation are detailed below. Note that all references to TRAC criteria may be found in the CRL-OCLC checklist document.

Lack of adequate authenticity and integrity measures

At the time of the first TRAC assessment, the H-Net message posting system relied on the author or editor informally checking a message after it posted to determine its accuracy. Also, a broken URL notification when attempting retrieval of a message indicated a problem. This lax approach to ensuring authenticity violated the guidelines of International Research on Permanent Authentic Records in Electronic Systems (InterPARES): "An authentic record is one that is what it purports to be and that is free from tampering or corruption. Determining that it is what it purports to be means confirming its identity. Determining that it is free from tampering or corruption means demonstrating that its integrity remains intact through space and time."⁴⁵ Two TRAC criteria that note the necessity of integrity checking include B4.4 "Repository actively

⁴² CRL and OCLC, "Trustworthy Repositories Audit and Certification," 43.

⁴³ Debian Project, "About Debian," <http://www.debian.org/intro/about>, accessed 29 August 2010.

⁴⁴ CRL and OCLC, "Trustworthy Repositories Audit and Certification," 11.

⁴⁵ International Research on Permanent Authentic Records in Electronic Systems (InterPARES), *The Long-Term Preservation of Authentic Electronic Records: Findings of the InterPARES Project*, "Appendix 6: How to Preserve Authentic Electronic Records" (8 October 2001), 3, http://www.interpares.org/book/interpares_book_o_app06.pdf, accessed 22 December 2009.

monitors integrity of archival objects” and C1.5 “Repository has effective mechanisms to detect bit corruption or loss.” Although the H-Net email lists employed MD5 hashes for message discovery purposes, those hashes were not used to perform the integrity checks, or checksums, that would ensure authenticity in a manner in keeping with InterPARES. Even if checksums were calculated, the lag time of up to seven days between when an editor sent a message for posting and when it was actually assigned an MD5 hash posed an obstacle to ensuring authenticity.

Lack of complete preservation description information (PDI)

TRAC criterion B2.9 states: “Repository acquires preservation metadata (i.e., PDI) for its associated Content Information.” The metadata in the database could partially fulfill the requirements for PDI—that is, reference, context, provenance, and fixity information—as recommended by the OAIS model.⁴⁶ The filename provides reference, context, and provenance information for notebook files. Filename plus an MD5 provides reference information for an individual message, additional context information for a message may be found in its subject line, and additional provenance information may be found in a message’s header. Unfortunately, as described previously, key provenance information from the message header is lost when the editor makes a change to a message and does not manually add back the creator’s information.

At the time of the original assessment, there was no fixity information to help ensure the authenticity of messages and notebook files. As noted, the MD5 hashes that could be used as checksums to establish and check fixity for messages were only used for discovery purposes.

Lack of file format migration strategies for private H-Net lists

The most significant property to be preserved is the message content, which must originate in plain text as required by public list use policy. Plain text is a recommended nonproprietary, archival format for text,⁴⁷ so no migration strategy or XML conversion was required for the preservation of public list messages. These open standards are readily available, and content stored in them may be accessed using text viewers. Attachments on the private lists are in proprietary formats, however, and there was no provision to normalize or migrate them to archival formats at the time of the assessment. Therefore, the H-Net lists did not

⁴⁶ CCSDS, “Reference Model for an Open Archival Information System,” 4–28.

⁴⁷ Lee et al., “PREMIS Requirement Statement Project Report,” 25.

fully support TRAC criterion B4.2 “Repository implements/responds to strategies for archival object (i.e., AIP) storage and migration.”

No archival copy of H-Net lists

Most of MATRIX’s existing backup and storage processes adequately met the criteria presented in Section C of the TRAC, “Technologies, Technical Infrastructure and Security.” Without a provision to create and maintain an archival copy of the H-Net lists separate from MATRIX data, however, the repository failed to fully address the archival object storage requirement cited above in criterion B4.2.

Security loopholes for possible administrative tampering

At the time of the original TRAC assessment, many MATRIX and executive staff held root system accounts that might have allowed tampering with the H-Net lists. In addition, list editors had privileges that would have allowed them to delete notebook files. The possibility of restricting privileges in both of these cases needed to be considered, in keeping with TRAC criterion C3.3 “Repository staff have delineated roles, responsibilities, and authorizations related to implementing changes within the system.”

Little to no documentation of preservation and other repository policies

The instructions for use of the TRAC checklist include noting “evidence” of support for each criterion, with a preference for written documentation.⁴⁸ At the time of the original assessment, little to no such documentation existed. An initial description of then-current practices and recommendations was referenced in the checklist, but there was no comprehensive set of written policies.

Addressing the Digital Preservation Gaps: Suggested Improvements

On completion of the TRAC assessment, Schmidt noted the preservation gaps in the H-Net email list archives. She worked with the H-Net systems administrator through the summer of 2008 to draw up a list of technical improvements and develop plans for closing the gaps and a timeline for their completion. In

⁴⁸ CRL and OCLC, “Trustworthy Repositories Audit and Certification,” 6.

fall 2008, the director of H-Net took the suggested improvements to the H-Net Council for approval. Most were approved, with exceptions discussed below. Schmidt also planned to create preservation policy documentation, as the TRAC assessment revealed gaps in that area.

Succession plan

As noted, MATRIX lacked a succession plan for the H-Net list archives at the time of the original assessment. A succession plan is necessary to ensure the continuity of operations of the archives if MATRIX can no longer fulfill that responsibility. The grant proposal to NHPRC noted this as an area that needed to be addressed, identifying the Library of Congress and OCLC as possible successor organizations, and Schmidt included a succession plan in the recommended improvements to the H-Net list archives. MATRIX and H-Net did not, however, identify or make plans with a successor organization during the course of the project.

Authenticity

To meet the authenticity and integrity requirements of InterPARES and the TRAC, MATRIX now establishes fixity both for individual messages on submission and notebook files on completion using the Secure Hash Algorithm (SHA)-256 message digest algorithm. Like MD5, SHA-256 is a cryptographic hash function used to verify the integrity of digital files. The National Institute of Standards and Technology (NIST) recommends the use of SHA-256 rather than the MD5 algorithm,⁴⁹ as data collisions that have occurred with MD5 indicate that it may not guarantee file integrity.⁵⁰

Both SHA-256 and MD5 hashes are now generated and metadata extracted within twenty-four hours of message submission. At this point in the workflow, the H-Net list archives have formally taken preservation responsibility for the submitted message, as stipulated in TRAC criterion B1.7. Narrowing the metadata extraction and hash generation gap are significant improvements over the previous timing of up to seven days, and MATRIX and H-Net consider the delay acceptable if not ideal. (To that end, the H-Net systems administrator has requested that L-Soft International, current owner and developer of LISTSERV,

⁴⁹ National Institute of Standards and Technology, "NIST's Policy on Hash Functions" (15 March 2006), <http://csrc.nist.gov/groups/ST/hash/policy.html>, accessed 22 December 2009.

⁵⁰ Department of Commerce, "Announcing Request for Candidate Algorithm Nominations for a New Cryptographic Hash Algorithm (SHA-3) Family," NIST, Docket No.: 070911510-7512-01, 72 Fed. Reg. 212 (2 November 2007), http://csrc.nist.gov/groups/ST/hash/documents/FR_Notice_Nov07.pdf, accessed 22 December 2009.

allow changes to the software that would eliminate the gap entirely by permitting the extraction of metadata and generation of hashes at the time of submission rather than up to twenty-four hours later.) The SHA-256 hashes are stored in a fixity database, separate from the metadata database but also residing on an H-Net server. This database is used to check fixity of the messages at the time of notebook file completion.

Fixity for notebook files is established at the time of notebook completion, also through the use of SHA-256. Notebooks in existence at the time of implementation of the fixity strategy were assigned SHA-256 hashes as well. All SHA-256 message digests for the notebooks are stored in the fixity database with those of the messages and validated weekly to ensure file integrity. Any errors found in message digest calculations for messages or notebook files will be logged and manually investigated. Figure 5 shows the H-Net information packages and how they relate to the new fixity measures.

Note that although MATRIX has moved to using the SHA-256 message digest algorithm to establish fixity and ensure file integrity, MD5 hashes will continue to be generated and used for individual message identification and discovery purposes.⁵¹

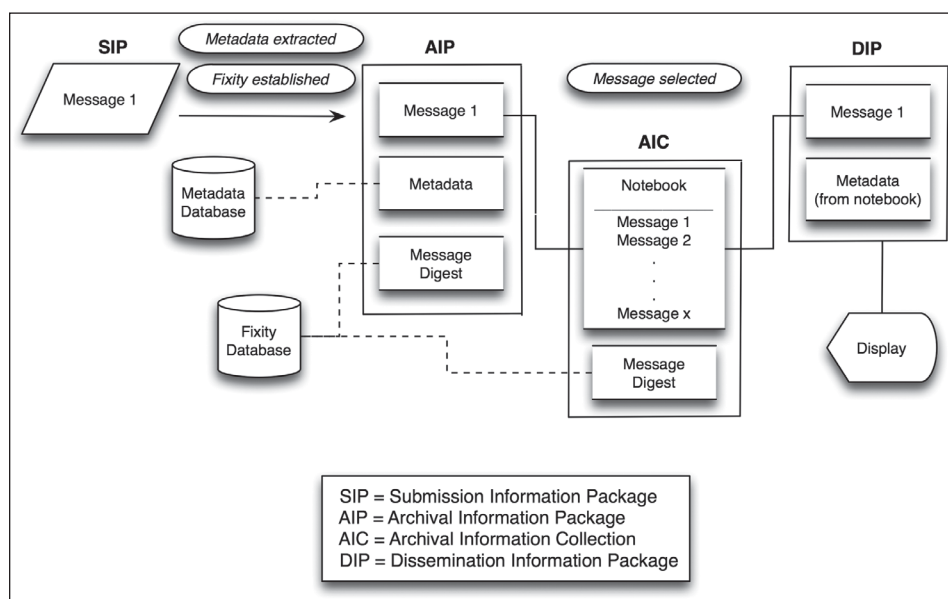


FIGURE 5. H-Net information packages.

⁵¹ The use of the strong SHA-256 cryptographic hash function to ensure the integrity of H-Net email messages may not have been necessary. Although present, the risk of MD5 data collisions is low, and H-Net email list messages are not “sensitive” data in personal, legal, or national security terms. MATRIX programming and IT staff were using SHA-256 in other software, however, and made the decision to use it with the H-Net lists as well.

More complete PDI

The SHA-256 hashes described above meet the need for message and notebook fixity information and complete the PDI requirement for the H-Net email lists. Recommendations to ensure accurate message creation metadata and provenance information when an editor makes a change to a message were not approved, however. The systems administrator proposed developing a Web-based list editing interface that would have automatically retained the original metadata. As legacy messages would not benefit from this improvement, the H-Net Council decided against using development resources to create the new interface.

Preservation and access strategy for the private lists, including attachments

The private lists must be made browsable with links to attachments. To that end, Schmidt advised H-Net administration to provide current private list subscribers with the information needed to browse list messages and to rewrite list welcome messages to include that information for new subscribers.

Attachments on the private H-Net lists actually comprise less than 0.01 percent of all H-Net messages. As there are so few attachments and most are in MS Office, PDF, JPEG, and other common formats, MATRIX has decided against implementing a formal file normalization or migration plan at this time. Most of the attachments should open in viewer software that can display the original file, later versions of the original applications, or even other applications. MATRIX will assist any users who report problems opening attachments. When the preservation practices of the H-Net email list are re-evaluated in the future, MATRIX may decide to normalize or migrate some formats or provide migration-on-demand for individual files as necessary.

Backup and archival storage

MATRIX's backup and other technological systems and processes at the time of the original assessment adequately fulfilled most of the TRAC criteria in Section C, which focuses on the repository's technical infrastructure. Regular incremental and full backups to LTO tape are performed.⁵² One outcome of the investigation of the preservation practices for the H-Net email list archive has been the strengthening of these backup processes by establishing two off-site

⁵² In 2010, MATRIX moved from performing backups using the proprietary NetVault software to the widely used Amanda open-source backup and recovery software.

backup plans. First, the Michigan State University Archives and Historical Collections will provide off-site storage for what had been known as the “permanent” MATRIX backup tapes. Now referred to as “long-term” rather than “permanent,” these tapes have been put on a two-year retention schedule. MATRIX has also entered into a reciprocal backup storage arrangement with the Inter-University Consortium for Political and Social Research (ICPSR) at the University of Michigan in Ann Arbor. This entails a synchronized daily backup of MATRIX and H-Net data (3 TB) to storage at ICPSR. MATRIX provides ICPSR with that same service for three terabytes of its data.

In keeping with the need for a separate archival copy of the H-Net email lists, MATRIX has implemented an archival storage plan. On an annual basis, the H-Net systems administrator makes two copies of the past year’s H-Net data and related databases and scripts to magnetic tape, with one copy held on the MATRIX premises and one at the MSU Archives; media refreshment is scheduled for every five years. The first set of archival copies was made to LTO magnetic tape using open-source GNU Tar archiving software. Schmidt recommended that MATRIX eventually move to storing the archival copies of the H-Net lists in a server-based digital repository rather than on the tapes. As part of a future greater-Michigan State University data management and preservation initiative, MATRIX and H-Net may also eventually participate in a university-wide distributed archival storage system.

Restriction of administrative capabilities

In the interests of eliminating the loophole that allows H-Net editors to delete or make changes to notebook files, notebook rights have been restricted to MATRIX and H-Net staff with postmaster privileges. Staff with access to root accounts will retain those privileges, however. The need to ensure 24/7 availability of MATRIX systems—including online history courses hosted by MATRIX—was deemed too important, and the likelihood of staff tampering with H-Net files is low. Although allowing this access might be considered a significant risk factor in other settings, MATRIX considers it an acceptable one.

Digital preservation policy documentation

At the time of the original assessment, the H-Net email lists lacked the written digital preservation policy documentation stipulated by the TRAC. A Cornell University/ICPSR-developed digital preservation policy management methodology, based on the RLG-OCLC *Trustworthy Digital Repositories* document and the OAIS model, provided a template for a digital preservation policy framework

and a wealth of examples that aided in creating and documenting the H-Net list policies, procedures, and processes.⁵³ The framework document includes sections covering the repository's OAIS compliance, administrative responsibility, organizational viability, financial sustainability, technological and procedural suitability, system security, and procedural accountability, as described in the RLG-OCLC *Trusted Digital Repositories* report.⁵⁴ Written digital preservation policies and procedures were thus created for the H-Net lists to provide evidence of fulfillment of the TRAC criteria.⁵⁵

Second TRAC Audit and Recommendations

On completion of the technical improvements and policy documentation, Schmidt performed a new TRAC assessment of the H-Net email list archives.⁵⁶ (See appendix for the checklist, including how all criteria were addressed.) This July 2009 assessment showed that MATRIX had narrowed the gaps in the original H-Net list archive, bringing it closer to achieving the status of a trusted digital repository for the valued academic discourse contained in its discussion networks.

The H-Net email list preservation project demonstrates that the TRAC may be used to evaluate a set of LISTSERV-based email lists as a trusted digital repository. Although the archive for the H-Net email lists is a live access system rather than a separate repository, most of the requirements for preservation are covered as noted in the TRAC checklist.

Further technical improvements could be made to enhance the preservation environment, however. First, as discussed in the previous section, the H-Net systems administrator should follow up with L-Soft International about changing the LISTSERV software to permit further changes to the timing of metadata extraction and hash generation. Closing the twenty-four-hour gap would better ensure the integrity of messages and their supporting metadata on submission.

MATRIX should replace the annual making of archival copies of the H-Net email lists and databases to tape with a server-based archival repository separate from the live system. Although the TRAC can be successfully applied to the live

⁵³ Kenney et al., "Digital Preservation Management."

⁵⁴ RLG, *Trusted Digital Repositories: Attributes and Responsibilities*.

⁵⁵ Lisa M. Schmidt, "Digital Preservation Policy Framework for the H-Net Electronic Mailing Lists," H-Net Digital Preservation Policies and Procedures, H-Net: Preserving and Improving Access to Specialized Electronic Mailing List Archives (June 2009), <http://www.h-net.org/archive/framework.php>, accessed 22 December 2009.

⁵⁶ Lisa M. Schmidt, "Trustworthy Repositories Audit and Certification: Criteria Checklist" (July 2009), <http://www.h-net.org/archive/trac7-09.pdf>, accessed 22 December 2009.

access system, a repository developed for the purpose of long-term preservation would better ensure the longevity of the content. The current loophole of possible system tampering by those with root server accounts would be minimized, and the process of making archival copies could be automated and performed more often. MATRIX and H-Net may also participate in a greater-Michigan State University preservation repository or distributed archival storage system when such options are made available.

In addition, MATRIX should convert the H-Net email lists to XML for preservation. During the course of this project, CERP researchers tested their XML conversion tool, used to archive email created in proprietary formats, on H-Net list data. They reported success in preserving the H-Net notebook files using the tool and email account schema developed for CERP. As noted previously, the plain text file format of H-Net precludes the need for conversion to XML, as the data are already in an archival format. XML affords more benefits than its status as an archival format, however, including adding structure to data that can make them more easily searchable. When a search framework that leverages the information encoded in the CERP email account schema is developed, conversion of H-Net email data to the XML format will become a more attractive preservation option.

From an administrative standpoint, MATRIX and H-Net still need to follow through on a succession plan for the H-Net list archives. Ideally, MATRIX will identify, negotiate with, and make preliminary plans with a potential successor. A memorandum of understanding with the successor, subject to periodic review, would satisfy this requirement.

Conclusion

The study of the H-Net email list system as a preservation environment marked the first formal application of the TRAC to email list archives and demonstrated the successful audit of a repository functioning as a live access system. Those who manage LISTSERV-based and other email lists containing scholarly discourse, such as JISCMail, and even list managers for lists that document less formal dialogue, such as the Society of American Archivists' Archives and Archivists (A&A) discussion list, may find the results of this study useful in creating and improving the preservation practices for their list archives.

The tools used in this study—including the TRAC and the Cornell/ICPSR digital preservation framework for developing and documenting preservation policies—can be applied to more complex data sets than the relatively small, mostly homogenous H-Net lists. Other digital preservation projects at Michigan State University involving proprietary file formats and complex digital objects will leverage the use and knowledge of these tools.

Appendix: TRAC Audit of H-Net Email Lists, July 2009

Note that the documents referenced throughout this checklist are available online as follows:

Document	Link
Archival Copies of H-Net	http://www.h-net.org/archive/copies.php
Digital Asset Policies for the H-Net E-Mail Lists	http://www.h-net.org/archive/asset.php
Digital Preservation Policy Framework for the H-Net Electronic Mailing Lists	http://www.h-net.org/archive/framework.php
Digital Preservation Strategies for the H-Net E-Mail Lists	http://www.h-net.org/archive/strategies.php
H-Net Ingest, Storage, and Retrieval Processes	http://www.h-net.org/archive/message.php
Ensuring the Integrity of the H-Net E-Mail Lists	http://www.h-net.org/archive/integrity.php
Executive Director's Annual Report	http://www.h-net.org/about/report09.pdf
H-Net By-Laws	http://www.h-net.org/about/by-laws.php
H-Net Constitution	http://www.h-net.org/about/constitution.php
H-Net E-Mail List Conformance to OAIS: Information Packages	http://www.h-net.org/archive/conformance.php
H-Net Strategic Plan	http://www.h-net.org/about/strategic.php
H-Net's Policy on Copyright and Intellectual Property	http://www.h-net.org/about/intellectualproperty.php
Information Security for Digital Assets at MATRIX	http://www2.matrix.msu.edu/information-security-for-digital-assets-at-matrix/
IRS Form 990	http://www.h-net.org/about/taxforms09.pdf
Roles and Responsibilities for Digital Preservation of the H-Net E-Mail Lists	http://www.h-net.org/archive/roles.php

Section A: Organizational Infrastructure**Aspect A1: Governance & Organizational Viability**

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
A1.1. Repository has a mission statement that reflects a commitment to the long-term retention of, management of, and access to digital information.	Digital Preservation Policy Framework for the H-Net Electronic Mailing Lists, Section 2, Administrative Responsibility H-Net Strategic Plan, Content Development section, Strategies	The digital preservation program for the H-Net e-mail lists supports H-Net's mission of enhancing scholarly communication by ensuring continued access to the academic discourse contained in H-Net messages.	Good
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.	Digital Preservation Policy Framework for the H-Net Electronic Mailing Lists, Section 4, Financial Sustainability, 4.1 Institutional Commitment	Acknowledgment that chosen succession partner must ensure ongoing preservation support. No formal succession plan. Directors at MATRIX and H-Net have committed to moving forward with this.	Incomplete

Aspect A2: Organizational Structure & Staffing

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
A3.1. Repository has defined its designated community(ies) and associated knowledge base(s) and has publicly accessible definitions and policies in place to dictate how its preservation service requirements will be met.	Digital Preservation Policy Framework for the H-Net Electronic Mailing Lists, Section 2, Administrative Responsibility, 2.1 Purpose; Digital Asset Policies for the H-Net E-Mail Lists; H-Net Ingest, Storage, and Retrieval Processes; Digital Preservation Strategies for the H-Net E-Mail Lists H-Net By-Laws (Sections 2.03, 2.04) and Constitution (Article VIII)	Policy Framework and supporting documents will not be publicly accessible until approved by H-Net Director and Council.	Good; but better when approved
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.	Digital Preservation Policy Framework for the H-Net Electronic Mailing Lists, Section 7, Procedural Accountability, 7.2 Digital Preservation Policy Framework Administration	Digital Preservation Policy Framework will be reviewed every two years and updated as necessary.	Good

Aspect A3: Procedural Accountability & Policy Framework

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
A3.1. Repository has defined its designated community(ies) and associated knowledge base(s) and has publicly accessible definitions and policies in place to dictate how its preservation service requirements will be met.	Digital Preservation Policy Framework for the H-Net Electronic Mailing Lists, Section 2, Administrative Responsibility, 2.1 Purpose; Digital Asset Policies for the H-Net E-Mail Lists; H-Net Ingest, Storage, and Retrieval Processes; Digital Preservation Strategies for the H-Net E-Mail Lists H-Net By-Laws (Sections 2.03, 2.04) and Constitution (Article VIII)	Policy Framework and supporting documents will not be publicly accessible until approved by H-Net Director and Council.	Good; but better when approved
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.	Digital Preservation Policy Framework for the H-Net Electronic Mailing Lists, Section 7, Procedural Accountability, 7.2 Digital Preservation Policy Framework Administration	Digital Preservation Policy Framework will be reviewed every two years and updated as necessary.	Good
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.	Digital Preservation Policy Framework for the H-Net Electronic Mailing Lists, Section 3, Organizational Viability, 3.5 Access and Use H-Net's Policy on Copyright and Intellectual Property; H-Net Constitution Article VIII, Section 7; H-Net By-Laws Section 2.04	Authors of messages retain copyright, but sending a message to an H-Net list constitutes granting permission to H-Net for distribution and (implicitly) preservation.	Good
A3.4. Repository is committed to formal, periodic review and assessment to ensure responsiveness to technological developments and evolving requirements.	Digital Preservation Policy Framework for the H-Net Electronic Mailing Lists, Section 7, Procedural Accountability, 7.1 Audit and Transparency	Assessment will be run every two years.	Good
A3.5. Repository has policies and procedures to ensure that feedback from producers and users is sought and addressed over time.	H-Net Constitution, Article VIII: H-Net Networks, Section 5	"In managing their networks, editors shall consult regularly with their editorial boards and their subscribers."	Adequate

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
A3.6. Repository has a documented history of the changes to its operations, procedures, software, and hardware that, where appropriate, is linked to relevant preservation strategies and describes potential effects on preserving digital content.	On MATRIX wiki; Technological Infrastructure for the H-Net E-Mail Lists (internal); backup logs; provenance metadata for archival copies on tape; systems info Information Security for Digital Assets at MATRIX; Archival Copies of H-Net Some comments in source code	Information on technological infrastructure will be maintained internally, backup logs, and provenance metadata will be maintained internally. System admin has started to write documentation for internally developed log browse and log search applications. Note that little tech history was documented before 2008.	Good
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.	Digital Preservation Policy Framework for the H-Net Electronic Mailing Lists, and supporting documents Notification of availability of policy/procedures documents	Digital preservation policy and procedures documents, which include information on all operations of the H-Net archive, will be made publicly available on approval of the H-Net Director and Council.	Good, when policy docs approved for publication
A3.8 Repository commits to defining, collecting, tracking, and providing, on demand, its information integrity measurements.	Digital Preservation Policy Framework for the H-Net Electronic Mailing Lists, Section 5, Technological and Procedural Suitability; Section 6, System Security Ensuring the Integrity of the H-Net E-Mail Lists; Information Security for Digital Assets at MATRIX; Archival Copies of H-Net; H-Net Message Ingest, Storage, and Retrieval Processes	The H-Net E-Mail List preservation system uses cryptographic hash functions to ensure message integrity. Comprehensive policy documentation describes integrity, security, and archival and workflow processes.	Good
A3.9 Repository commits to a regular schedule of self-assessment and certification and, if certified, commits to notifying certifying bodies of operational changes that will change or nullify its certification status.	Digital Preservation Policy Framework for the H-Net Electronic Mailing Lists, Section 7, Procedural Accountability, 7.1 Audit and Transparency	Assessment will be run every two years. Repository not seeking certification status, so no certifying bodies to notify. Internal self-assessment.	Good

Aspect A4: Financial Sustainability

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
A4.1. Repository has short- and long-term business planning processes in place to sustain the repository over time.	Digital Preservation Policy Framework for the H-Net Electronic Mailing Lists, Section 4, Financial Sustainability, 4.1 Institutional Commitment H-Net Strategic Plan Administration, Funding & Structure; Executive Director's Annual Report; IRS Form 990		Good
A4.2. Repository has in place processes to review and adjust business plans at least annually.	Executive Director's Annual Report		Good
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.	IRS Form 990 Budget spreadsheets (internal)		Good
A4.4. Repository has ongoing commitment to analyze and report on risk, benefit, investment, and expenditure (including assets, licenses, and liabilities).	H-Net Strategic Plan, Administration, Funding & Structure; Executive Director's Annual Report		Good
A4.5. Repository commits to monitoring for and bridging gaps in funding.	H-Net Strategic Plan, Administration, Funding & Structure; Executive Director's Annual Report; IRS Form 990		Good

Aspect A5: Contracts, Licenses, & Liabilities

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
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.	Digital Asset Policies for the H-Net E-Mail Lists H-Net By-Laws, Section 2.03	Terms of "deposit" are spelled out in by-laws, so no need for individual contracts with message posters.	Good
A5.2 Repository contracts or deposit agreements must specify and transfer all necessary preservation rights, and those rights transferred must be documented.	Digital Preservation Policy Framework for the H-Net Electronic Mailing Lists, Section 3, Organizational Viability, 3.5 Access and Use; Digital Asset Policies for the H-Net E-Mail Lists H-Net's Policy on Copyright and Intellectual Property; H-Net Constitution, Article VIII, Section 7; H-Net By-Laws, Section 2.04	Authors of messages retain copyright, but sending a message to an H-Net list constitutes granting permission to H-Net for distribution and (implicitly) preservation.	Good
A5.3 Repository has specified all appropriate aspects of acquisition, maintenance, access, and withdrawal in written agreements with depositors and other relevant parties.	Digital Preservation Policy Framework for the H-Net Electronic Mailing Lists, Section 3, Organizational Viability, 3.4 Selection and Acquisition, 3.5 Access and Use; Digital Asset Policies for the H-Net E-Mail Lists H-Net By-Laws, Sections 2.02 and 2.03		Good
A5.4 Repository tracks and manages intellectual property rights and restrictions on use of repository content as required by deposit agreement, contract, or license.	Digital Preservation Policy Framework for the H-Net Electronic Mailing Lists, Section 3, Organizational Viability, 3.5 Access and Use; Digital Asset Policies for the H-Net E-Mail Lists H-Net's Policy on Copyright and Intellectual Property; H-Net Constitution, Article VIII, Section 7; H-Net By-Laws, Section 2.04		Good
A5.5 If repository ingests digital content with unclear ownership/rights, policies are in place to address liability and challenges to those rights.	Digital Asset Policies for the H-Net E-Mail Lists H-Net's Policy on Copyright and Intellectual Property; H-Net Constitution, Article VIII, Section 7; H-Net By-Laws, Section 2.04		Good

Section B: Digital Asset Management

Aspect B1: Ingest: Acquisition of Content

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
B1.1. Repository identifies properties it will preserve for digital objects.	Digital Preservation Policy Framework for the H-Net Electronic Mailing Lists, Section 5, Technological and Procedural Suitability; Digital Preservation Strategies for the H-Net E-Mail Lists	Significant property is message content. Bit-level preservation Messages: Already in text formats (ASCII, UTF-8) Attachments: < 0.01% of all messages; not enough to merit more preservation attention at this time.	Good
B1.2. Repository clearly specifies the information that needs to be associated with digital material at the time of its deposit (i.e., SIP).	H-Net E-Mail List Conformance to OAIS: Information Packages H-Net By-Laws, Section 2.03 (d) (iv)	Metadata in e-mail header	Good
B1.3. Repository has mechanisms to authenticate the source of all materials.	Digital Preservation Policy Framework for the H-Net Electronic Mailing Lists, Section 3, Organizational Viability, 3.4 Selection and Acquisition; H-Net Message Ingest, Storage, and Retrieval Processes; Digital Asset Policies for the H-Net E-Mail Lists H-Net By-Laws, Section 2.03 (d) (i)	Messages go through list editors before being posted. Most lists require subscriptions before a user may post.	Good
B1.4. Repository's ingest process verifies each submitted object (i.e., SIP) for completeness and correctness as specified in B1.2.	H-Net By-Laws, Section 2.03	LISTSERV software validates message in terms of e-mail standards before it can be delivered. Once posted, a message is subject to vetting by both author and editor.	Good
B1.5. Repository obtains sufficient physical control over the digital objects to preserve them (Ingest: content acquisition).	H-Net Message Ingest, Storage, and Retrieval Processes; Digital Asset Policies for the H-Net E-Mail Lists H-Net's Policy on Copyright and Intellectual Property; H-Net Constitution, Article VIII, Section 7; H-Net By-Laws, Sections 2.03 and 2.04	Repository has physical control over the messages.	Good
B1.6. Repository provides producer/depositor with appropriate responses at predefined points during the ingest processes.	H-Net Message Ingest, Storage, and Retrieval Processes	Editor receives acknowledgment message on submission. Acknowledgment option may be turned off, in which case no acknowledgment will be sent.	Good

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
B1.7. Repository can demonstrate when preservation responsibility is formally accepted for the contents of the submitted data objects (i.e., SIPs).	H-Net Message Ingest, Storage, and Retrieval Processes	Within 24 hours of submission, repository creates SHA-256 hash, extracts key metadata for metadata cache, and posts message to notebook file where it is available for discovery and access.	Good
B1.8. Repository has contemporaneous records of actions and administration processes that are relevant to preservation.	H-Net E-Mail List Conformance to OAIS: Information Packages; H-Net Message Ingest, Storage, and Retrieval Processes	Notebook file naming process and Preservation Description Information (PDI) for H-Net messages and notebooks are described in these documents.	Good

Aspect B1: Ingest: Acquisition of Content

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
B2.1. Repository has an identifiable, written definition for each AIP or class of information preserved by the repository.	H-Net E-Mail List Conformance to OAIS: Information Packages	Includes definition of the AIC, a notebook file containing all messages posted in a seven-day period.	Good
B2.2. Repository has a definition of each AIP (or class) that is adequate to fit long-term preservation needs.	H-Net E-Mail List Conformance to OAIS: Information Packages	Includes definition of the AIC, a notebook file containing all messages posted in a seven-day period.	Good
B2.3. Repository has a description of how AIPs are constructed from SIPs.	H-Net E-Mail List Conformance to OAIS: Information Packages	Includes description of how AICs are constructed with AIPs.	Good
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.	H-Net E-Mail List Conformance to OAIS: Information Packages	Not really applicable. Once a message becomes a SIP, it's accepted. All AIPs become part of an AIC.	Good
B2.5. Repository has and uses a naming convention that generates visible, persistent, unique identifiers for all archived objects (i.e., AIPs).	H-Net Message Ingest, Storage, and Retrieval Processes; H-Net E-Mail List Conformance to OAIS: Information Packages	AIP: Each message has a unique identifier: a combination of the name of the notebook file in which it is stored and its unique MD5 hash. AIC: Notebooks are uniquely named by list, month, year, and seven-day time period.	Good
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).	NA	No unique identifiers associated with SIPs before ingest.	NA

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
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).	No use of international format registries at this time	Messages and notebook files are created and preserved as text, a well-documented format. Most of the attachments are in currently available formats, such as PDF and Microsoft Office formats, and there are too few of them to merit more than bit-level preservation at this time.	Good
B2.8 Repository records/ registers Representation Information (including formats) ingested.	Digital Preservation Policy Framework for the H-Net Electronic Mailing Lists, Section 5, Technological and Procedural Suitability; Digital Preservation Strategies for the H-Net E-Mail Lists	Representation Information (format information) recorded in documentation rather than repository itself. Messages and notebook files are created and preserved as text, a well-documented format. Most of the attachments are in currently available formats, such as PDF and Microsoft Office formats, and there are too few of them to merit more than bit-level preservation at this time.	Good
B2.9 Repository acquires preservation metadata (i.e., PDI) for its associated Content Information.	H-Net E-Mail List Conformance to OAIS: Information Packages	Document contains descriptions of PDI for the H-Net E-Mail List archive.	Good
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.	H-Net E-Mail List Conformance to OAIS: Information Packages	Content Information and PDI appropriately understandable as is.	Good
B2.11 Repository verifies each AIP for completeness and correctness at the point it is generated.	Ensuring the Integrity of the H-Net E-Mail Lists	On submission, a SHA-256 hash is created for a message as it is posted to a notebook file and becomes an AIP. All hashes are validated before the notebook file closes and becomes an AIC. The AIC receives its own hash at that point, and those are checked on a weekly basis.	Good
B2.12 Repository provides an independent mechanism for audit of the integrity of the repository collection/content	H-Net Message Ingest, Storage, and Retrieval Processes; H-Net E-Mail List Conformance to OAIS: Information Packages; Ensuring the Integrity of the H-Net E-Mail Lists	Previous criteria satisfied, so this may not even be necessary.	Good

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
B2.13 Repository has contemporaneous records of actions and administration processes that are relevant to preservation (AIP creation).	H-Net Message Ingest, Storage, and Retrieval Processes; Ensuring the Integrity of the H-Net E-Mail Lists	Metadata extracted and stored in cache at time of ingest; SHA-256 hashes created and stored in fixity database at time of ingest (messages) and notebook creation (notebook files).	Good

Aspect B3: Preservation Planning

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
B3.1. Repository has documented preservation strategies.	Digital Preservation Policy Framework for the H-Net Electronic Mailing Lists, Section 5, Technological and Procedural Suitability; Digital Preservation Strategies for the H-Net E-Mail Lists; Archival Copies of H-Net; Ensuring the Integrity of the H-Net E-Mail Lists		Good
B3.2. Repository has mechanisms in place for monitoring and notification when Representation Information (including formats) approaches obsolescence or is no longer viable.	Digital Preservation Policy Framework for the H-Net Electronic Mailing Lists, Section 5, Technological and Procedural Suitability; Digital Preservation Strategies for the H-Net E-Mail Lists	Not applicable at this time, as the messages and notebook files are in text formats. There are too few attachments to merit more than a commitment to bit-level preservation at this time.	Good
B3.3 Repository has mechanisms to change its preservation plans as a result of its monitoring activities.	NA		NA
B3.4. Repository can provide evidence of the effectiveness of its preservation planning.	Digital Preservation Policy Framework for the H-Net Electronic Mailing Lists, Section 5, Technological and Procedural Suitability; Digital Preservation Strategies for the H-Net E-Mail Lists; Archival Copies of H-Net; Ensuring the Integrity of the H-Net E-Mail Lists	Not included in documentation at this time. Messages have been accessible and usable for years, however. New preservation measures must be put into use for a period of time before their effectiveness can be truly measured.	Good enough for now

Aspect B4: Archival Storage & Preservation/Maintenance of AIPs

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
B4.1. Repository employs documented preservation strategies.	Digital Preservation Policy Framework for the H-Net Electronic Mailing Lists, Section 5, Technological and Procedural Suitability; Digital Preservation Strategies for the H-Net E-Mail Lists; Archival Copies of H-Net; Ensuring the Integrity of the H-Net E-Mail Lists		Good
B4.2. Repository implements/responds to strategies for archival object (i.e., AIP) storage and migration.	Digital Preservation Policy Framework for the H-Net Electronic Mailing Lists, Section 5, Technological and Procedural Suitability; Digital Preservation Strategies for the H-Net E-Mail Lists		Good
B4.3 Repository preserves the Content Information of archival objects (i.e., AIPs).	Digital Asset Policies for the H-Net E-Mail Lists; H-Net Message Ingest, Storage, and Retrieval Processes; H-Net E-Mail List Conformance to OAIS: Information Packages H-Net By-Laws, Section 2.02	Policy is not to remove messages once they are posted. Only on rare occasions is this allowed.	Good
B4.4 Repository actively monitors integrity of archival objects (i.e., AIPs).	Digital Preservation Policy Framework for the H-Net Electronic Mailing Lists, Section 5, Technological and Procedural Suitability; Digital Preservation Strategies for the H-Net E-Mail Lists; Ensuring the Integrity of the H-Net E-Mail Lists Logs of fixity checks	Fixity checks performed on messages before a notebook closes. Fixity checks performed on notebooks weekly.	Good
B4.5 Repository has contemporaneous records of actions and administration processes that are relevant to preservation (Archival Storage).	H-Net Message Ingest, Storage, and Retrieval Processes; Ensuring the Integrity of the H-Net E-Mail Lists Logs of fixity checks	Logs kept of most recent hash validations.	Good

Aspect B5: Information Management

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
B5.1 Repository articulates minimum metadata requirements to enable the designated community to discover and identify material of interest.	H-Net Message Ingest, Storage, and Retrieval Processes; H-Net E-Mail List Conformance to OAIS: Information Packages	Descriptive metadata found in browser view of list archive.	Good

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
B5.2 Repository captures or creates minimum descriptive metadata and ensures that it is associated with the archived object (i.e., AIP).	H-Net Message Ingest, Storage, and Retrieval Processes; H-Net E-Mail List Conformance to OAIS: Information Packages	Metadata captured from SIP; metadata extracted from SIP, stored in metadata cache, and associated with message for more efficient discovery; MD5 hash created for message.	Good
B5.3 Repository can demonstrate that referential integrity is created between all archived objects (i.e., AIPs) and associated descriptive information.	H-Net Message Ingest, Storage, and Retrieval Processes; H-Net E-Mail List Conformance to OAIS: Information Packages	Referential integrity between unique instance of a message (notebook file name + MD5 hash) and descriptive metadata stored in cache	Good
B5.4 Repository can demonstrate that referential integrity is maintained between all archived objects (i.e., AIPs) and associated descriptive information.	H-Net Message Ingest, Storage, and Retrieval Processes; H-Net E-Mail List Conformance to OAIS: Information Packages	Referential integrity between unique instance of a message (notebook file name + MD5 hash) and descriptive metadata stored in cache. Message could not be retrieved if this integrity was lost.	Good

Aspect B6: Access Management

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
B6.1 Repository documents and communicates to its designated community what access and delivery options are available.	Digital Asset Policies for the H-Net E-Mail Lists; H-Net Message Ingest, Storage, and Retrieval Processes	Documents not available to the public pending approval from H-Net Director and Council. H-Net website needs a "how to use" page.	Incomplete
B6.2 Repository has implemented a policy for recording all access actions (includes requests, orders, etc.) that meets the requirements of the repository and information producers/depositors.	Digital Asset Policies for the H-Net E-Mail Lists; H-Net Message Ingest, Storage, and Retrieval Processes	Access requests are kept in a log file on the Apache web server for approximately one year.	Good
B6.3 Repository ensures that agreements applicable to access conditions are adhered to.	Digital Asset Policies for the H-Net E-Mail Lists	Anyone can access messages on most of the public lists. Two public lists (H-Bahai and H-Grad) require subscription to access. Private lists require subscriptions to view messages.	Good
B6.4 Repository has documented and implemented access policies (authorization rules, authentication requirements) consistent with deposit agreements for stored objects.	Digital Asset Policies for the H-Net E-Mail Lists	Most public lists are available to all online, whether or not they are subscribers. Private lists are only available to subscribers.	Good

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
B6.5 Repository access management system fully implements access policy.	Authentication information (IDs and passwords) for private lists stored in protected directories on the web server.	Private lists are only available to subscribers. They must log in to access archived messages online, and the list would have to recognize them as subscribers for them to access messages via commands.	Good
B6.6 Repository logs all access management failures, and staff review inappropriate "access denial" incidents.	Error log on web server	Any access attempt failures would be logged as errors.	Good
B6.7 Repository can demonstrate that the process that generates the requested digital object(s) (i.e., DIP) is completed in relation to the request.	Digital Asset Policies for the H-Net E-Mail Lists; H-Net Message Ingest, Storage, and Retrieval Processes	If message requested appears in browser, the process is a success.	Good
B6.8 Repository can demonstrate that the process that generates the requested digital object(s) (i.e., DIP) is correct in relation to the request.	Digital Asset Policies for the H-Net E-Mail Lists; H-Net Message Ingest, Storage, and Retrieval Processes	If message requested appears in browser, the process is a success. Very, very occasionally, an error message is encountered, indicating that the request was not a success.	Good
B6.9 Repository demonstrates that all access requests result in a response of acceptance or rejection.	Digital Asset Policies for the H-Net E-Mail Lists; H-Net Message Ingest, Storage, and Retrieval Processes	All access requests result in some response: either the requested message or an error message.	Good
B6.10 Repository enables the dissemination of authentic copies of the original or objects traceable to originals.	H-Net Message Ingest, Storage, and Retrieval Processes; H-Net E-Mail List Conformance to OAIS: Information Packages	On selection of a message through the H-Net browser interface, a URL is constructed that includes the name of the notebook file containing the message and its MD5 hash—a unique identifier that ties the message to the metadata that was extracted and stored in the cache at time of ingest.	Good

Section C: Technologies, Technical Infrastructure, & Security

Aspect C1: System Infrastructure

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
CI.1 Repository functions on well-supported operating systems and other core infrastructural software.	Technological Infrastructure for the H-Net E-Mail Lists (internal); Information Security for Digital Assets at MATRIX	MATRIX servers run on Debian, a popular and well-established distribution of Linux. Also well-supported open source software: Apache webserver, Postfix mail transfer agent, MySQL database. Proprietary, but well supported: L-Soft LISTSERV e-mail list software. System admin has relationship with developers at L-Soft.	Good
CI.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.	Technological Infrastructure for the H-Net E-Mail Lists (internal); Information Security for Digital Assets at MATRIX	Dual-core server for backup, runs NetVault software tape backup software to Quantum tape library. Daily incremental backups, weekly full backups to tape stored in building across campus. Additional off-site backups: full tape backups every four months, off-site storage; reciprocal server backup to ICPSR in Ann Arbor daily.	Good
CI.3 Repository manages the number and location of copies of all digital objects.	Information Security for Digital Assets at MATRIX; Archival Copies of H-Net Backup and archival copy logs on wiki	Four backup copies: incremental, weekly full, long-term, reciprocal storage Two archival copies to tape (annual)	Good
CI.4 Repository has mechanisms in place to ensure any/multiple copies of digital objects are synchronized.	Information Security for Digital Assets at MATRIX	Backups are regularly scheduled. Enough redundant systems are in place to ensure security of data.	Good
CI.5 Repository has effective mechanisms to detect bit corruption or loss.	Information Security for Digital Assets at MATRIX; Archival Copies of H-Net; Ensuring the Integrity of the H-Net E-Mail Lists	SHA-256 hashes created and checked regularly.	Good
CI.6 Repository reports to its administration all incidents of data corruption or loss, and steps taken to repair/replace corrupt or lost data.	Information Security for Digital Assets at MATRIX; Archival Copies of H-Net; Ensuring the Integrity of the H-Net E-Mail Lists	System reports validation errors, which must be manually investigated and corrected.	Good

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
C1.7 Repository has defined processes for storage media and/or hardware change (e.g., refreshing, migration).	Information Security for Digital Assets at MATRIX; Archival Copies of H-Net	Incremental/weekly backup tapes replaced as needed. Long-term backups are on a three-year retention schedule. Archival copies are refreshed to new tapes every five years. No documented process for hardware system refreshment/migration. Hardware updated every 3-4 years, per agreed upon principles of technology lifecycles and service contracts.	Adequate
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.		No documented change management system.	Incomplete
C1.9 Repository has a process for testing the effect of critical changes to the system.		Informal testing of changes. Not documented.	Incomplete
C1.10 Repository has a process to react to the availability of new software security updates based on a risk-benefit assessment.	Updates recorded in file on net monitor server.	Automated security patches and updates applied monthly and as needed. Debian updates very reliable.	Good

Aspect C2: Appropriate Technologies

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
C2.1 Repository has hardware technologies appropriate to the services it provides to its designated communities and has procedures in place to receive and monitor notifications, and evaluate when hardware technology changes are needed.	Technological Infrastructure for the H-Net E-Mail Lists (internal) MATRIX wiki	Supporting hardware described in "Technological Infrastructure" document. New systems installed with current hardware and software, and older systems get software updates as needed. Monitoring process for technology changes informal, with information gathered from reading online and print sources, discussions with peers, etc. Changes made only with consensus of MATRIX/H-Net technical staff.	Good

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
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.	Technological Infrastructure for the H-Net E-Mail Lists (internal)	Software used to support the H-Net E-Mail Lists is described in "Technological Infrastructure" document. New systems are installed with current hardware and software, and older systems get software updates when convenient or if new features are needed. Monitoring process for technology changes is informal, with information gathered from reading online and print sources, discussions with peers, etc. Changes made only with consensus of technical staff.	Good

Aspect C3: Security

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
C3.1 Repository maintains a systematic analysis of such factors as data, systems, personnel, physical plant, and security needs.	MATRIX wiki	Decisions made per consensus among technical staff.	Adequate
C3.2 Repository has implemented controls to adequately address each of the defined security needs.	Information Security for Digital Assets at MATRIX	The system's biggest security threat is loss of data. Redundant backup systems guard against that threat.	Good
C3.3 Repository staff have delineated roles, responsibilities, and authorizations related to implementing changes within the system.	Authorizations documented in system code		Good
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).	Information Security for Digital Assets at MATRIX	Redundant backup plans take disaster recovery into account. Two offsite backups. Restore plans currently in heads of system administration staff, who plan to document disaster recovery procedures in the MATRIX wiki. The wiki will be backed up along with everything else, and will also be regularly saved to flash drives or another USB-based removable media for easy access in case of disaster.	Incomplete