

**H-Net:  
Preserving and Improving Access to Specialized Electronic  
Mailing List Archives**

**Interim Narrative Progress Report, August 1, 2008–January 31, 2009**

**Project Activities Undertaken**

Of the project activities scheduled, MATRIX undertook the following during the reporting period:

***H-Net Preservation Workplan***

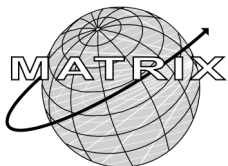
- Draft workplan for implementing H-Net preservation plan during next 12 months, to include revisions based on the TRAC
- Consult with Archival Advisory Board to review and revise H-Net preservation plan
- Present and discuss H-Net preservation plan with H-Net governing council
- Post workplan for implementing preservation plan on project website
- Begin to implement workplan for H-Net preservation
- Write supplementary document describing Information Packages in the H-Net preservation system, currently and after implementation of the H-Net preservation workplan
- Develop and document a backup and archival storage strategy
- Post supplementary Information Packages and backup/archival storage documents to project website
- Evaluate with Archival Advisory Board the progress on workplan for H-Net preservation and make any recommended revisions

***Presentations***

- Deliver presentation reporting on preliminary project results at Research Forum of society of American Archivists (SAA) annual meeting
- Deliver brief project overview at the Committee on Institutional Cooperation/University Archivists Group (CIC/UAG) meeting
- Deliver presentation reporting on project results at American Library Association (ALA) Midwinter Meeting, digital preservation interest group meeting

***Other Activities***

- Write and submit article on project to SAA Campus Case Studies
- Begin negotiations with OCLC and/or Library of Congress about a specific succession plan
- Semantic Augmented Consensus Clustering (SACC) research on H-Net records



The Center for Humane Arts,  
Letters, and Social Sciences Online

COLLEGE OF  
**SOCIAL SCIENCE**  
MATRIX

Michigan State University  
409 Natural Science Building  
East Lansing, Michigan  
48824-1120

517/355-9300  
FAX: 517/355-8363  
<http://www.matrix.msu.edu>

## **H-Net Preservation Workplan**

***Draft workplan for implementing H-Net preservation plan during next 12 months, to include revisions based on the TRAC.*** During the reporting period, archivist Lisa Schmidt drafted a document that outlined recommended improvements to the H-Net preservation based on the analysis of current practices produced in the previous reporting period. This included a workplan schedule for technological improvements to be implemented by systems administrator Dennis Boone. Recommendations included:

- Better ensuring authenticity by (1) establishing and periodically checking fixity of messages and notebook files; (2) developing an editorial interface that would allow editors to add metadata to messages rather than overwriting original message creation metadata; (3) and restricting editors' LISTSERV administration capabilities to prevent possible message and/or notebook deletions.
- Preserving attachments on the private H-Net lists by (1) making the lists browsable; and (2) providing a migration strategy for them. For the latter, an inventory must be taken and conversion on demand provided.
- Preserving links to content within messages, by redirecting broken URLs to websites archived by the Internet Archive.
- Mapping URLs of messages to shorter persistent URLs, for easier use as citations.

In addition, the plan stated that Ms. Schmidt will create and document preservation policies, as recommended by the Trustworthy Repositories Audit & Certification: Criteria Checklist (TRAC). Backup and storage recommendations were not included in this workplan.

***Consult with Archival Advisory Board to review and revise H-Net preservation plan.*** The plan was presented to the board, and Dr. Patricia Galloway in particular provided very useful feedback.

***Present and discuss H-Net preservation plan with H-Net governing council.*** Dr. Peter Knupfer, director of H-Net, presented the plan to the H-Net Council. All of the recommendations were approved, with the exception of the development of the editorial interface.

***Post workplan for implementing preservation plan on project website.*** The document "Preservation of the H-Net E-Mail Lists: Suggested Improvements" and a schedule of the recommended technology improvements were posted to the project website in August 2008.

***Begin to implement workplan for H-Net preservation.*** Mr. Boone has made good headway on implementing the recommended technological improvements. The recommended additions of fixity functionality in the system are near completion.

***Write supplementary document describing Information Packages in the H-Net preservation system, currently and after implementation of the H-Net preservation workplan.*** On the recommendations of board members, Ms. Schmidt reevaluated H-Net in terms of the Information Packages (IPs) described in the Open Archival Information System (OAIS) model. The information in this new document replaces references to the IPs in the H-Net preservation system presented in earlier documents.

***Develop and document a backup and archival storage strategy.*** Ms. Schmidt prepared this supplement to the original workplan. Plans for improving the current backup and storage strategy for H-Net include:

- Creating a second set of long-term backup tapes for MATRIX (including H-Net) and storing them at an offsite, climate-controlled facility in Lansing through an arrangement with the Michigan State University Archives.
- Entering into a reciprocal storage arrangement with the Interuniversity Consortium for Political and Social Research (ICPSR) at the University of Michigan, Ann Arbor, in which MATRIX's data is synchronized and copied into dark storage at ICPSR and vice versa. Members of MATRIX's IT and programming staff are currently implementing the storage arrangement.
- Creating archival copies of the H-Net lists to tape on an annual basis. Ideally, future plans will include storing H-Net archives at an alternative storage repository and/or participating in a distributed archival storage system. Mr. Boone has created archival copies to tape (two copies) that include all H-Net notebook files through the end of 2008 as well as scripts for the browse and search functions and some documentation.

***Post supplementary Information Packages and backup/archival storage documents to project website.*** Both of these documents were posted to the project website in January 2009.

***Evaluate with Archival Advisory Board the progress on workplan for H-Net preservation and make any recommended revisions.*** The board has been apprised of all developments on the project, and feedback has been encouraged.

### **Presentations**

***Deliver presentation reporting on preliminary project results at Research Forum of Society of American Archivists (SAA) annual meeting.*** On August 26, 2008, Ms. Schmidt presented "Preserving Electronic Mailing Lists as Scholarly Resources: The H-Net Archives" at the SAA Research Forum in San Francisco. This 10-minute presentation briefly described the H-Net system and recommended preservation improvements; it is now available on the project website at <http://www.h-net.org/archive/presentations.php>. Approximately 100 people attended the Research Forum.

***Deliver brief project overview at the Committee on Institutional Cooperation/University Archivists Group (CIC/UAG) forum.*** On September 26, 2008, Ms. Schmidt delivered a brief presentation on the project to the CIC-UAG forum, themed “Managing University Digital Assets and Resources,” was held at Michigan State University in East Lansing. The audience of more than 70 attendees included individuals working in such functional areas IT, legal, registrar, business and finance, archives, and records management at CIC member universities.

***Deliver presentation reporting on project results at American Library Association (ALA) Midwinter Meeting, digital preservation interest group meeting.*** On January 25, 2009, Ms. Schmidt presented “Preserving the H-Net E-Mail Lists: A Case Study in Trusted Digital Repository Assessment” at the digital preservation interest group meeting of the ALA Midwinter Meeting in Denver. Approximately 80 people were in attendance. In addition to describing how H-Net works and the planned preservation improvement strategies, a large part of this 30-minute presentation focused on the TRAC and how it was used to assess the H-Net preservation system; the audience was especially interested in use of the TRAC. The presentation has been posted to the H-Net archive project website, <http://www.h-net.org/archive/presentations.php>. The Library of Congress referenced the presentation in a press release on its website ([http://www.digitalpreservation.gov/news/2009/20090204news\\_article\\_ALAMidwin.html](http://www.digitalpreservation.gov/news/2009/20090204news_article_ALAMidwin.html)) and in its February 2009 digital preservation newsletter (<http://www.digitalpreservation.gov/news/newsletter/200902.pdf>).

#### **Other Activities**

***Write and submit article on project to SAA Campus Case Studies.*** Ms. Schmidt wrote and submitted a Campus Case Study on the project, published in February 2009 on the SAA website (<http://www.archivists.org/publications/epubs/CampusCaseStudies/casestudies/Cas e11-Schmidt.pdf>).

***Begin negotiations with OCLC and/or Library of Congress about a specific succession plan.*** Dr. Mark Kornbluh, director of MATRIX, and Dr. Knupfer have discussed the possibility of engaging in negotiations with OCLC or the CIC’s HathiTrust on a succession plan for H-Net.

***Semantic Augmented Consensus Clustering (SACC) research on H-Net records.*** Bill Punch, Associated Faculty at MATRIX and Associate Professor of Computer Science and Engineering at MSU, reports that results of the SACC research continue to be mixed. A dedicated computer science graduate student has tested multiple variations on ways to determine and utilize sense in the clustering process. Results on these experiments are compared to standard clustering approaches, such as LSI and k-means; different results have been observed depending on the dataset tested. Consequently, new approaches to measuring the characteristics of each dataset have been taken to see if differences can be discerned that affect the usefulness of semantic content. Tools to better measure the information gain/change from co-clustering methods are being developed to determine how much this adds to the clustering process.

In December 2008, an NSF grant “How Sense Makes Sense: Evaluation and Integration of Semantic Ontologies in Text Mining” (Tan, Punch, NSF IIS small, \$479,332) to further support his research was submitted. Research will continue on two major fronts: Better ways to incorporate sense in clustering, and ways to get better measures on datasets to determine the importance and/or usefulness of sense measures.

### **Additional Project Developments**

In addition to activities and events noted on the project workplan, Ms Schmidt managed the following developments during the reporting period:

- Participation in a Digital Preservation Management Symposium put on by the Midwest Archives Conference (MAC) in Lawrence, Kansas, November 2008. Presented by Nancy McGovern, Digital Preservation Officer at ICPSR, this symposium was an excellent review and source of guidance on developing policies for the H-Net archive.
- Participation in an invitation-only E-Mail Preservation Symposium sponsored by the Collaborative Electronic Records Project (CERP), a collaboration between the Smithsonian Institution and the Rockefeller Archives Center, in Washington, DC, November 2008.
- Submission and acceptance of proposals to present at the Digital Curation Curriculum (DigCCurr) Symposium in April 2009 (poster), the MAC meeting in April/May 2009 (panel), and the SAA conference in August 2009 (panel).
- Meeting with Nancy McGovern for in-depth discussion of several preservation issues related to the project.
- Invitation to speak to a practical engagement workshop class at the University of Michigan’s School of Information on March 19, 2009.



Mark Kornbluh  
Principal Investigator



Lisa M. Schmidt  
Project Manager