

MATRIX and H-Net Backup and Archival Storage: Practices and Suggested Improvements

Preservation of the H-Net E-Mail Lists Supplemental Report

Lisa M. Schmidt
MATRIX: Center for Humane Arts, Letters and Social Sciences Online
Michigan State University
December 2008

MATRIX and H-Net Backup and Archival Storage: Practices and Suggested Improvements

H-Net: Humanities and Social Sciences Online is an international consortium of scholars and teachers that has grown to include more than 180 scholarly social sciences and humanities networks hosted by MATRIX: Center for Humane Arts, Letters and Social Sciences Online at Michigan State University. In addition to its public lists, H-Net includes more than 230 “private” lists used by editors, council members, and administrators for planning, testing, and advisory purposes. MATRIX received a grant from the National Historical Publications and Records Commission (NHPRC) to advance the state of e-mail preservation by assessing and improving upon the digital preservation practices for the H-Net electronic mailing lists to ensure longevity of the content. Representing a compilation of years of academic discourse, with messages bookmarked and cited in scholarly research and publications, the H-Net lists are considered a valuable scholarly resource requiring long-term preservation.

The document “ Preservation of the H-Net E-Mail Lists: Suggested Practices,” posted on the H-Net archive project website in August 2008, provides suggestions for improvements in such areas as ensuring authenticity and preserving attachments.¹ This supplementary document provides a brief look at current backup and storage practices for MATRIX, suggests improvements to those practices, and addresses measures that should be taken to provide true archival storage for H-Net data. (For the complete original

¹ Schmidt, Lisa M., “Preservation of the H-Net E-Mail Lists: Suggested Improvements,” H-Net: Preserving and Improving Access to Specialized Electronic Mailing List Archives, August 2008, <http://www.h-net.org/archive/documentation/hnetpresimprov.pdf>.

assessment of the H-Net preservation system, refer to the “Current Practices” document posted in spring 2008.)²

Data Backup and Storage

MATRIX runs its operations, including the H-Net e-mail lists, on several servers kept in a climate controlled, physically secured room; these servers run the Debian distribution of Linux. Approximately 2.7 TB of data is stored on the servers as of December 2008. Incremental tape backups are performed daily, with a full backup performed on a weekly basis and those tapes taken to the MSU Computer Center and exchanged for the tapes stored there the previous week. Backup tapes cycle through the system approximately every six weeks and are replaced as needed, such as when a cartridge breaks.

In addition to these ongoing backups, a full “permanent” backup is performed every one-to-two months to ensure against data loss. Those tapes are kept in a cabinet in a minimally secured room, presumably in perpetuity. The MATRIX systems administrator keeps a wiki-based log of all tape backups.

Several improvements must be made to MATRIX backup and storage processes to better protect and ensure continued availability of H-Net and other data. For security purposes, the systems administrator should install a lock on the cabinet housing the permanent backup tapes. MATRIX should also create a second set of these tapes and store them offsite at a secure, climate controlled storage facility in nearby Lansing, Michigan. This will likely be accomplished through an arrangement with the Michigan

² Schmidt, Lisa M, “Preservation of the H-Net E-Mail Lists: Current Practices,” H-Net: Preserving and Improving Access to Specialized Electronic Mailing List Archives, March 2008, <http://www.h-net.org/archive/documentation/H-Net%20Current%20Practices%20Post2.pdf>.

State University Archives, which already contracts with the facility. Rather than “permanent,” these should be considered long-term backup tapes and put on a two-to-five year retention schedule.

In addition to maintaining the long-term backup tapes, MATRIX is moving forward with a reciprocal storage arrangement with the Interuniversity Consortium for Political and Social Research (ICPSR) at the University of Michigan, Ann Arbor. ICPSR will synchronize and copy over MATRIX data into “dark” storage—that is, storage that cannot be accessed by general users—and MATRIX will do the same for ICPSR. Offsite storage of data with ICPSR and the Lansing storage facility will form part of a MATRIX disaster recovery plan.

Archival Storage

Implementation of the offsite reciprocal and tape storage processes described above will ensure good backup practices for MATRIX data, including the H-Net lists. In addition, separate archival copies of the H-Net data must be maintained. On an annual basis, MATRIX will copy the H-Net records and associated metadata created during that time period onto tapes, as well as a text file containing provenance information for the archival copy. This provenance metadata will consist of information about when, where, and on what type of media the archival copy was made, as well as metadata related to actions such as media refreshment as they take place in the future. One copy of these tapes should be kept at an offsite location, such as the aforementioned storage center in Lansing, and a second copy kept in a secure location on the MATRIX premises or elsewhere, with media refreshment scheduled for every five years. MATRIX will keep a

wiki-based log, similar to that established for the backup tapes, containing descriptive and provenance metadata for each tape and the actions taken on the data.

While MATRIX is committed to maintaining and preserving the H-Net archive, the center is taking the prudent measure of identifying a possible successor in the archive's stewardship. The ideal potential partner would also provide an alternative archival storage repository for H-Net even as MATRIX continues to function as the live host, holding the records and metadata in a dark archive. MATRIX could provide the partner with a current copy of the H-Net data and associated metadata and then the new data for each successive year on an annual basis.

MATRIX should also strive to participate in a distributed archival storage system for H-Net and other data that requires archiving. Within the next two to five years, Michigan State University plans to implement the Integrated Rule-Oriented Data System (iRODS)³ or another preservation system that MATRIX could join. Another option might be participation in a Lots of Copies Keep Stuff Safe (LOCKSS)⁴- or Storage Resource Broker (SRB)⁵-based system.

³ https://www.irods.org/index.php/Introduction_to_iRODS

⁴ <http://www.lockss.org/lockss/Home>

⁵ http://www.sdsc.edu/srb/index.php/Main_Page