Preserving Electronic Mailing Lists: The H-Net Archive

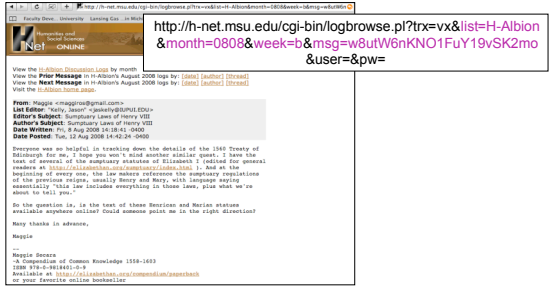
Lisa M. Schmidt
MATRIX, Michigan State University
DigCCurr, April 2009

Overview

- About H-Net
  - International consortium of scholars and teachers
  - Oldest collection of born-digital, content-moderated arts, humanities, and social science material online
  - Hosted by MATRIX, a digital humanities research center at Michigan State University
  - More than 1 million e-mail messages on more than 180 public networks and more than 230 private lists
- Matrix received NHPRC grant to assess and improve upon H-Net e-mail list preservation practices
- Primary assessment tool: Trustworthy Repository Audit & Certification: Criteria and Checklist (TRAC)

How H-Net Works

- Runs on LISTSERV software
- Public list subscribers send plaintext messages, no attachments, to list editor for approval
- Messages post from seconds to days after approval
- Seven-day concatenations of messages to a list become "notebook" files
- Key metadata extracted and MD5 hashes created for each message, written to database cache
- Messages identified through combination of notebook file name ("h-africa.log0802a") and MD5 hash



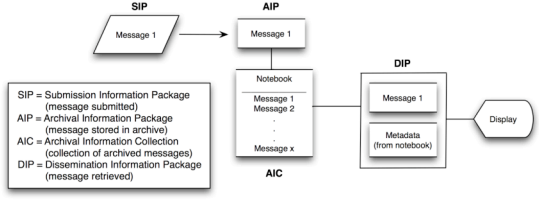
Preservation Assessment

Use of the TRAC

- Administrative and technical criteria for trusted repository
  - Third-party audit or internal assessment
  - Requires supporting documentation
- Internal assessment of H-Net
  - Ensure efficacy as preservation environment
  - Highlight areas that require improvement

Existing Preservation Practices

H-Net Mapped to the OAIS Model



Backup and Storage

- 3TB data stored on MATRIX servers, including H-Net
- Daily incremental, weekly full backups to tape stored at Computer Center, six-week cycle through system
- Monthly full "permanent" backup tapes stored in secure room at MATRIX

Format

- Messages created/stored in plaintext ASCII, UTF-8
  - Non-proprietary, archival format
  - No need for migration plan at this time
- Private list attachments proprietary, require migration

Authenticity

- Informal measures only
- No fixity established for messages or files

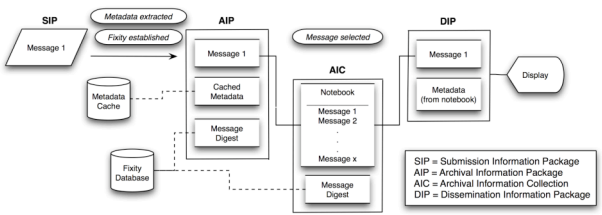
Preservation improvements

Backup and Archival Storage

- Establish offsite and reciprocal server storage
- Create more than one set of long-term backup tapes, put on retention schedule—no longer "permanent"
- Create annual archival copies to tape, H-Net data only
- Explore dark and distributed storage options

Authenticity

- Establish fixity for messages on submission and notebooks on creation using SHA-256
- Validate message hashes on notebook completion and notebook hashes weekly



Attachments (Private Lists)

- Determine most pervasive attachment formats
- Provide browser access to private lists
- Conversion tools in reserve or pointers to websites
- Establish technology watch for format changes

Other Improvements

- Preserve links in messages, redirect to archived websites
- Succession plan
- Write policies, gather supporting documentation

Futures

- Perform new TRAC assessment after improvements
- Digital preservation planning for Michigan State