

Preserving Electronic Mailing Lists as Scholarly Resources: The H-Net Archives

Lisa M. Schmidt

lisa.schmidt@matrix.msu.edu

<http://www.h-net.org/archive/>

MATRIX: The Center for Humane Arts,
Letters & Social Sciences Online

Michigan State University

August 26, 2008



H-Net: Humanities and Social Sciences Online

- International consortium of scholars and teachers
- Oldest collection of born-digital and content-moderated arts, humanities, and social science material on the Internet
- Valuable scholarly resource
 - More than 180 networks, or e-mail lists
 - More than 230 “private” lists
- More than 1 million e-mail messages
- Hosted by MATRIX

NHPRC Grant

- Conduct assessment of existing H-Net preservation policies and practices
- Apply NARA/OCLC TRAC checklist
- Develop and implement an improved long-term preservation plan
- Useful to those managing large collections of electronic records
- Research semantic clustering search techniques

Preserving E-Mail Lists as Scholarly Resources

- How H-Net Works
- Current Preservation Practices
- Preservation Improvement Plan

How H-Net Works: Backup & Security

- 2.7 TB of data, including H-Net
- Server rack kept in climate controlled, physically secured room
- Daily incremental backups, weekly full
 - Tapes cycle through system every 6 weeks
 - Swapped tapes stored in secure location
 - Tapes replaced as needed
- Monthly full, permanent tape backups
 - Tapes kept in minimally secure cabinet
 - Plans to keep log and move to offsite storage

How H-Net Works: Posting Messages

- H-Net runs on LISTSERV Software
- Users must be list subscribers to post
- Messages written in plain text
- No attachments allowed on public lists
- Editors approve and post messages
- Editors can overwrite creation metadata

How H-Net Works: Archiving of Lists

- Messages post from a few seconds up to several days after approval
- Messages kept in flat text files called “notebooks”
- Notebook includes messages posted during a weekly time period

How H-Net Works: Archiving of Lists

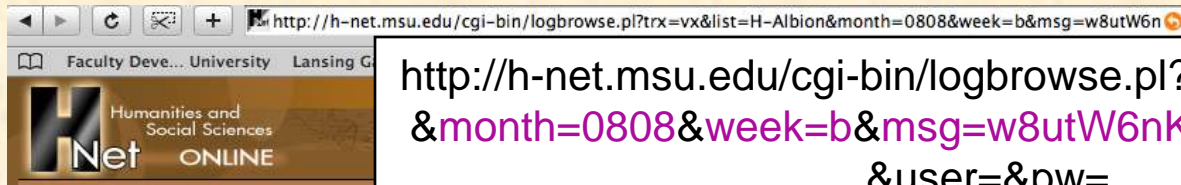
Time Period	Day of Month
a	1-7
b	8-14
c	15-21
d	22-28
e	29-31

Ex. "h-africa.log0802a"

How H-Net Works: Archiving of Lists

- Log browse cache application extracts key metadata, creates MD5 hashes
- Cache builder script writes metadata to MySQL database cache
 - Notebook filename
 - Offset (byte position) of message
 - Author name and e-mail address
 - Subject
 - Date in two formats
 - Messageid (MD5 hash)

How H-Net Works: Message Retrieval



View the [H-Albion Discussion Logs](#) by month
View the **Prior Message** in H-Albion's August 2008 logs by: [\[date\]](#) [\[author\]](#) [\[thread\]](#)
View the **Next Message** in H-Albion's August 2008 logs by: [\[date\]](#) [\[author\]](#) [\[thread\]](#)
Visit the [H-Albion home page](#).

From: Maggie <maggiros@gmail.com>
List Editor: "Kelly, Jason" <jaskelly@IUPUI.EDU>
Editor's Subject: Sumptuary Laws of Henry VIII
Author's Subject: Sumptuary Laws of Henry VIII
Date Written: Fri, 8 Aug 2008 14:18:41 -0400
Date Posted: Tue, 12 Aug 2008 14:42:24 -0400

Everyone was so helpful in tracking down the details of the 1560 Treaty of Edinburgh for me, I hope you won't mind another similar quest. I have the text of several of the sumptuary statutes of Elizabeth I (edited for general readers at <http://elizabethan.org/sumptuary/index.html>). And at the beginning of every one, the law makers reference the sumptuary regulations of the previous reigns, usually Henry and Mary, with language saying essentially "this law includes everything in those laws, plus what we're about to tell you."

So the question is, is the text of these Henrican and Marian statues available anywhere online? Could someone point me in the right direction?

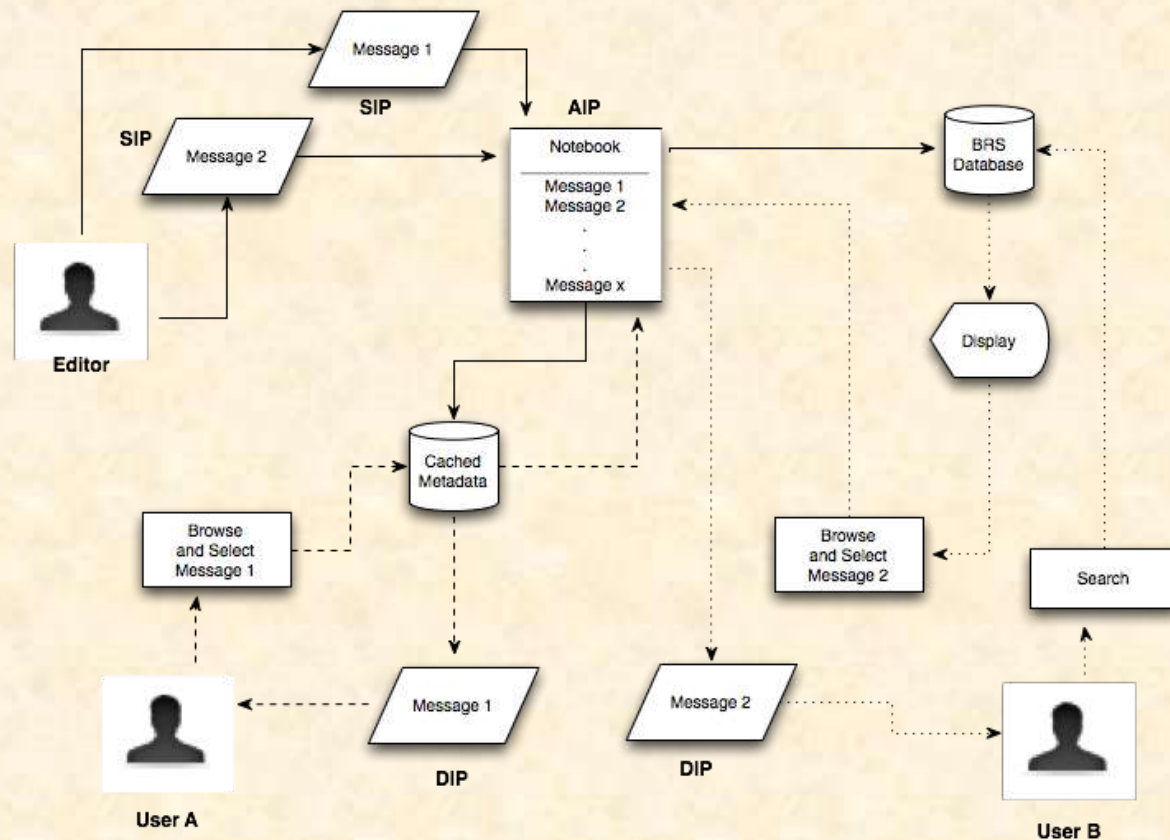
Many thanks in advance,

Maggie

--
Maggie Secara
-A Compendium of Common Knowledge 1558-1603
ISBN 978-0-9818401-0-9
Available at <http://elizabethan.org/compendium/paperback>
or your favorite online bookseller

Current Preservation Practices

Message Ingest, Storage, and Retrieval Processes



Current Preservation Practices

- Backup and storage
- Significant property: message/notebook content, stored in plain text formats
- Authenticity
 - Informal check by author and/or editor on posting
 - Broken URL on message retrieval attempt
- Notebook filename partially fulfills PDI recommendation
 - Reference, Content, Provenance Information
 - (ex., h-albion.log0808b)
 - No Fixity Information

Preservation Improvement Plan: Backup & Storage

- Media refreshment schedule
- More than one set of permanent backup tapes, or a server mirror
- Secure storage systems
- Backup log
- Participation in distributed storage system

Preservation Improvement Plan: Authenticity

Fixity: Individual Messages (SIPs)

- Shorten time window for generation of MD5 hashes
- Create database of MD5 hashes for fixity checks
- Validate message hashes on notebook completion

Fixity: Notebook Files (AIPs)

- Create SHA-2 message digests on completion of notebooks
- Calculate SHA-2 message digests for existing notebooks
- Create database of SHA-2 message digests for fixity checks
- Validate notebook hashes on weekly basis

Preservation Improvement Plan: Authenticity

Accurate Message Creation Metadata

- Build list editing web interface for editors
- Will only help with new messages

Restriction of Editors' Administration Capabilities

- Eliminate editors' ability to retrieve and change notebooks
- Restrict notebook modification rights to MATRIX postmasters

H-Net Tampering Risk?

- Low—staff with root system account privileges are trusted employees
- No action required

Preservation Improvement Plan: Attachments

Browser Access for Private Lists

- Provide constructed URLs, as with public lists
- Provide download links to attachments

Migration Strategy

- Conduct inventory of attachments on H-Net-related lists
- Provide conversion on demand
 - Option 1: Keep conversion tools in reserve
 - Option 2: Automate conversion
- Establish or leverage technology watch

Preservation Improvement Plan: Other Technical Improvements

Preservation of Links to Original Content

- Redirect URLs within messages to archived websites

Shorter Persistent URLs

- Develop naming scheme for shorter URLs
- Map shorter URLs to actual URLs

Preservation Improvement Plan: From TRAC Checklist

- Succession plan
- Periodic review or trigger event definition
- Document, document, document!
 - Technology history
 - Change management system
 - Staff roles, responsibilities, and authorizations
 - Written recovery plan

References

- H-Net Archives, Documentation, <http://www.hnet.org/archive/doc.php>
- H-Net: Humanities and Social Sciences Online, <http://www.h-net.org>
- InterPARES, <http://www.interpares.org>
- MATRIX: The Center for Humane Arts, Letters, and Social Sciences Online, <http://www.matrix.msu.edu>
- OAIS Reference Model, <http://public.ccsds.org/publications/archive/650x0b1.pdf>
- Trustworthy Repositories Audit & Certification: Criteria and Checklist, <http://www.crl.edu/PDF/trac.pdf>