Born-Digital Archives in Collecting Repositories:

Turning Challenges into Byte-Size Opportunities

Gretchen Gueguen, Mark A. Matienzo, Simon Wilson, and Peter Chan

Session 502, 27 August 2011
Society of American Archivists Annual Meeting
AIMS Project

• “Born-Digital Collections: An Inter-Institutional Model for Stewardship”

• Two year project to create a framework for stewardship of born-digital archival records in collecting repositories

• Funded by the Andrew W. Mellon Foundation
Grant Goals

• Processing of Hybrid Collections
• Software Development
• Community Development
  – Unconference (May 2011, Charlottesville, VA)
  – Workshop (August 2011, Chicago, IL)
• White Paper and Project Report
Framework Development

A framework for collecting and delivering the born-digital materials that are quickly beginning to constitute the collections of contemporary scholarly, literary, and political figures and organizations.
Collection Development

Gretchen Gueguuen
University of Virginia
What is Collection Development?

Actions and policies of institutions to bring in material for end users (both current and future); includes prioritizing, developing relationships with creators, assessments, negotiating agreements and preparing for accessioning.

Within the AIMS framework

Viable, practical method to capture/process born-digital material from hybrid collections requires sound work at the beginning (i.e. policies, practices, agreements with donors, etc.) to set up later work.
Elements of Collection Development

1. Prerequisites
2. Establish relationship with donor
3. Analyze Feasibility
4. Negotiate Agreements
5. Prepare for Accessioning
Prerequisites...


AIMS Digital Material Survey – Personal Digital Archives (Part II)

Note: This part of the survey is designed to be filled out by digital archivists regarding technical details of the tools used to create digital material.

1. Hardware
1.1 List the hardware configurations of each computer / mobile device. (e.g. manufacturer, model no, cpu, ram, hard drive capacity, video card, etc.)
1.2 Find out if the computers have USB ports or CD writers which could be used to copy the digital files.

2. Software
2.1 List the operating system and other system software with version no., installed in all the hardware (in 1).
2.2 Check if system date and time are set correctly. List the “Time Zone” used, if any.
2.3 With the help of the donor, list the main application software, with version no., used to create digital files.
2.4 If Microsoft Office is used, find out if the “User Name” field is set to the name of the donor. Find out similar settings for other main application software used.

3. Internet Access
3.1 Find out if the digital archivist can use the Internet access in the donor’s office using the digital archivist’s portable computer?

4. Networking
4.1 With the help of the donor, confirm if the computer is connected to file servers. Confirm if the donor save files in the file server. How much file server space is used by the donor?

5. Security
5.1 With the help of the donor, confirm if login is required to access desktop computers / mobile devices?
5.2 With the help of the donor, confirm if a digital certificate is used by the donor to login / sign digital files / encrypt digital files?
5.3 With the help of the donor, confirm if digital files are encrypted?

This work is based on the Paradigm records survey published by the Bodleian Library, Oxford University.
Enhanced Curation
Analyzing Feasibility...
Negotiate Agreements...

**THE PERFECT HANDSHAKE**

\[ PH = \sqrt{(e^2 + ve^2)(d^2) + (cg + dl)^2 + (4c5>2)(4c2>p>2))^2 + (vi + t + te)^2 + (4c2>2)(4cdu>2))^2} \]

(a) is eye contact (1=none, 5=direct); (v) is verbal greeting (1=totally inappropriate, 5=totally appropriate); (d) is Duchenne smile - smiling in eyes and mouth, plus symmetry on both sides of face, and slower offset (1=totally non-Duchenne smile (false smile), 5=total); (c) is completeness of grip (1=very incomplete, 5=full); (g) is dryness of hand (1=damp, 5=dry); (e) is strength (1=weak, 5=strong); (p) is position of hand (1=away from body, 5=other person's body); (l) is vigor (1=too low/too high, 5=mid); (i) is temperature of hands (1=too cold/too hot, 5=mid); (t) is texture of hands (1=rough, 5=smooth); (c) is control (1=low, 5=high); (d) is duration (1=brief, 5=long)

1. **RIGHT HAND**
2. **DRY PALM**
3. **STRONG GRIP WITH FINGERS UNDER RECEIVING PALM**
4. **THREE TO FOUR VIGOROUS SHAKES FOR TWO TO THREE SECONDS**
5. **EYE CONTACT AND APPROPRIATE SMILE THROUGHOUT**

All rights reserved by Chevrolet UK
Prepare for Accessioning...

Scope and extent determined?

Method and time determined?

Pre-acquisition appraisal performed?

Test capture if needed?

Development of new methodologies undertaken as needed/possible?

Coordination with acquisition of analog material?

Enhanced curation carried out?
Accessioning

Mark A. Matienzo, Yale University
What is Accessioning?

Archival institution takes physical and legal custody of a group of records from a donor and documents the transfer in a register or other representation of the institution’s holdings.

Within AIMS Framework

Processes which establish physical, administrative and intellectual control over transferred records; assessment and documentation of future needs; documentation of actions taken; beginning of safe storage and maintenance.
Elements of Accessioning

1. Prerequisites
2. Transfer records and gain administrative control
3. Physical control and stabilization
4. Intellectual control and documentation to support further processes
5. Maintain accessioned records
Case Study: Re-Accessioning at Yale

• Collaborative capacity building across two repositories
  – Manuscripts and Archives
  – Beinecke Rare Book and Manuscript Library

• Addressing previously received accessions of containing electronic records on media

• Still in testing phase, but working towards implementing in production
Types of Records and Media

• Wide variety of records creators
  – Literary authors
  – University faculty
  – University offices
  – Architectural firms

• Common types of media
  – Floppy disks: 5.25” and 3.5”
  – Optical media: CDROM, CD-R, DVD-R, etc.
  – Zip disks
  – USB flash drives
Goals of Re-Accessioning

- Identify, document, and register media
- Mitigate risk of media deterioration and obsolescence
- Extract basic metadata from filesystems on media and files contained on filesystems
Re-Accessioning Workflow

Start accessioning process

Retrieve media

Assign identifiers to media

Write-protect media

Record identifying characteristics of media in media log

Create image

Extract filesystem-and file-level metadata

Package images and metadata for ingest

Transfer package

Ingest transfer package

Document accessioning process

End accessioning process
Disk Imaging

• Using “forensic” (bit-level) imaging process
• Ensure data on media is not manipulated using write-protection
• Uses software to acquire images
• Includes hash-based verification process
Media Log

- Using SharePoint list
- Contains unique identifier of media
- Records physical/logical characteristics of media
- Documents success, failure, or status of various processes and additional notes
# Media Log

## Electronic Records on Media Accessioning Log

<table>
<thead>
<tr>
<th>New</th>
<th>Actions</th>
<th>Settings</th>
<th>View</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>All Items</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type</th>
<th>Media number</th>
<th>Media Format</th>
<th>Imaging Date</th>
<th>Imaging Successful?</th>
<th>Bag Created?</th>
<th>Metadata Extracted?</th>
<th>Transfer to Storage Date</th>
<th>Examiner</th>
<th>Image format</th>
<th>Imaging Software</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD-R</td>
<td>2011-M-075.0001</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Glick, Kevin</td>
<td>N/A</td>
<td>N/A</td>
<td>FAT 7.0.0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVD-R</td>
<td>2011-M-075.0002</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Glick, Kevin</td>
<td>ISO</td>
<td>ImgBurn</td>
<td>ISO9 (1.02)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVD-R</td>
<td>2011-M-075.0003</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Glick, Kevin</td>
<td>ISO</td>
<td>ImgBurn</td>
<td>ISO9 (1.02)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVD-R</td>
<td>2011-M-075.0004</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Glick, Kevin</td>
<td>ISO</td>
<td>ImgBurn</td>
<td>ISO9 (1.02)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVD-R</td>
<td>2011-M-075.0005</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Glick, Kevin</td>
<td>ISO</td>
<td>ImgBurn</td>
<td>ISO9 (1.02)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVD-R</td>
<td>2011-M-075.0006</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Glick, Kevin</td>
<td>ISO</td>
<td>ImgBurn</td>
<td>ISO9 (1.02)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD-R</td>
<td>2011-M-075.0007</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Glick, Kevin</td>
<td>ISO</td>
<td>ImgBurn</td>
<td>ISO9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD-R</td>
<td>2011-M-075.0008</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Glick, Kevin</td>
<td>ISO</td>
<td>ImgBurn</td>
<td>ISO9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD-R</td>
<td>2011-M-075.0009</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Glick, Kevin</td>
<td>ISO</td>
<td>ImgBurn</td>
<td>ISO9 (1.02)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVD-R</td>
<td>2011-M-075.0010</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Glick, Kevin</td>
<td>ISO</td>
<td>ImgBurn</td>
<td>ISO9 (1.02)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD-R</td>
<td>2011-M-075.0011</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Glick, Kevin</td>
<td>ISO</td>
<td>ImgBurn</td>
<td>ISO9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD-R</td>
<td>2011-M-075.0012</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Glick, Kevin</td>
<td>ISO</td>
<td>ImgBurn</td>
<td>ISO9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zip disk</td>
<td>2011-M-075.0013</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Glick, Kevin</td>
<td>dd (Raw)</td>
<td>FTKImager 3.0.0.1443</td>
<td>FAT 7.0.0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The Andrew W. Mellon Foundation
### Media Log

#### Electronic Records on Media Accessioning Log

<table>
<thead>
<tr>
<th>New</th>
<th>Actions</th>
<th>Settings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Type | Media number | Media Format | Imaging Date | Imaging Successful? | Bag Created | Media number | Media Format | Imaging Successful? | Bag Created | Media Density (floppies only) | Interface | Label text | Manufacturer | Serial Number (hard drives only) | Examiner | Imaging Successful? | Imaging Date | Image filename | Source File System | Image format | Imaging Software | Image Fidelity Function | Image Fidelity Value | Notes | Metadata Extracted? | Bag Created? | Transfer to Storage Date | Fiscal Year | Created at | Last modified at |
|------|--------------|--------------|--------------|---------------------|-------------|--------------|--------------|---------------------|-------------|-------------------------------|-----------|-------------|---------|-----------------|----------------------|-----------|----------------|-----------------|---------------|-------------------|--------------|-------------|---------------|
|      |              |              |              |                     |             | 2011-M-075.0001 | CD-R          | No                  | No           |                               |           |             |          |                 |                      |           |                |                 |              |                   |             |             |               |
|      |              |              |              |                     |             | 2011-M-075.0002 | DVD-R         | Yes                 | No           |                               |           |             |          |                 |                      |           |                |                 |              |                   |             |             |               |
|      |              |              |              |                     |             | 2011-M-075.0003 | DVD-R         | Yes                 | No           |                               |           |             |          |                 |                      |           |                |                 |              |                   |             |             |               |
|      |              |              |              |                     |             | 2011-M-075.0004 | DVD-R         | Yes                 | No           |                               |           |             |          |                 |                      |           |                |                 |              |                   |             |             |               |
|      |              |              |              |                     |             | 2011-M-075.0005 | DVD-R         | Yes                 | No           |                               |           |             |          |                 |                      |           |                |                 |              |                   |             |             |               |
|      |              |              |              |                     |             | 2011-M-075.0006 | DVD-R         | Yes                 | No           |                               |           |             |          |                 |                      |           |                |                 |              |                   |             |             |               |
|      |              |              |              |                     |             | 2011-M-075.0007 | DVD-R         | Yes                 | No           |                               |           |             |          |                 |                      |           |                |                 |              |                   |             |             |               |
|      |              |              |              |                     |             | 2011-M-075.0008 | CD-R          | Yes                  | No           |                               |           |             |          |                 |                      |           |                |                 |              |                   |             |             |               |
|      |              |              |              |                     |             | 2011-M-075.0009 | CD-R          | Yes                  | No           |                               |           |             |          |                 |                      |           |                |                 |              |                   |             |             |               |
|      |              |              |              |                     |             | 2011-M-075.0010 | DVD-R         | Yes                 | No           |                               |           |             |          |                 |                      |           |                |                 |              |                   |             |             |               |
|      |              |              |              |                     |             | 2011-M-075.0011 | CD-R          | Yes                 | No           |                               |           |             |          |                 |                      |           |                |                 |              |                   |             |             |               |
|      |              |              |              |                     |             | 2011-M-075.0012 | CD-R          | Yes                 | No           |                               |           |             |          |                 |                      |           |                |                 |              |                   |             |             |               |
|      |              |              |              |                     |             | 2011-M-075.0013 | Zip disk      | Yes                | No           |                               |           |             |          |                 |                      |           |                |                 |              |                   |             |             |               |

**Notes:**

- mam54 04/28/2011: Could not extract metadata using fiwalk; log file from imaging process says that the block structure is Mode 2/Format 1
- Created at 4/27/2011 9:35 AM by Glidc Kevin
- Last modified at 4/29/2011 4:26 PM by Metienko, Mark
Metadata Extraction

• Can be repurposed for descriptive, administrative, and technical metadata
• Uses command-line tools (Sleuthkit, fiwalk)
• Outputs XML document
Packaging and Transfer

• Using BagIt packages/Bagger application
• Packages contain disk images, extracted metadata, imaging logs, and high-level accession information
• Transfer to storage is verified by comparison against manifest
Arrangement & Description

Simon Wilson
Hull University Archives
Purpose of Arrangement & Description

The general objectives for Arrangement & Description are:
- to preserve context
- to establish intellectual control of the material
- to provide a means of discovery

SAA definition, emphasis on *minimizing the amount of handling*

Within the AIMS framework
Processes which establish intellectual control of the material including implementation of policies and agreements with donors etc. to enable subsequent discovery and access
Elements of Arrangement and Description

1. Prerequisites

2. Plan for processing
   - gather supporting information; files captured from media (accessioning); convert files (for viewing); appraisal strategy; assess arrangement options; consider preservation issues

3. Processing
   - implement arrangement strategy; add descriptive metadata and wider context (eg Collection Level Description); copyright & other legal considerations

4. Prepare for Discovery & Access
   - remove restricted access to b-d material during processing
Case Study - Stephen Gallagher

Background:

2005: 42 boxes paper archives

2010: born-digital material:
14,320 files (13.6GB) transferred to us via external hard drive and a box of Amstrad disks

Create **integrated catalogue** to accommodate paper, born-digital and future accruals
Case Study - Stephen Gallagher

Approach:
- current work higher priority in filing system
- considered each work a distinct ‘project’

- structure reflect his way of working & the archival principles of control that creator, archivist & user can all understand

Series level was most logical solution
- all related files placed in the series
- reasonable return for our effort
Case Study - Stephen Gallagher

300 files created using FinalDraft screenwriter software
- view file (as created) to identify appropriate format for long term preservation

Other issues:
- copyright/third-party content
- commercial implications: access via repository = publication?
- re-purposing of work from one (unsuccessful) project to another
Challenges faced

Each collection is unique, approach will vary:
- integrate born-digital material with existing material/arrangement?
- one-off collection (eg project) or likely to be subsequent accruals?
- collection type; differs for personal papers & organisational records
- same personnel work on paper and born-digital components?
- can we appraise without knowing the contents?
  similar to paper material that is in a different language?
Challenges faced

Volume of material:
- depositor perception that 'storage is cheap' - does this mean we shouldn’t appraise the material we receive?
- wide range of file types encountered
- not practical to describe each and every file
- risk management - if you don’t check every file for sensitive information

- we need to automate as much of the processing as possible
Digital archivists identified a gap in current tools – used experiences to define the requirements for a new tool

**Key features identified:**
- need an intuitive (for archivists) graphical interface
- drag'n'drop to create the intellectual arrangement
- ability to return to original order of the material
- view some file types, add descriptive metadata etc
- high level of granularity when applying rights & permissions

Technical (acquired at accessioning) and descriptive metadata - Discovery & Access process
Discovery and Access

Peter Chan
Stanford University
What is Discovery & Access

_Discovery and Access_ refers to the systems and workflows that make _processed or unprocessed_ material and the metadata that support it available to users.
Goals of D&A

- To make material available to user communities by ensuring that they can:
  - find out about material
  - understand whether it is available for consultation and if so, how
  - access material.

- To apply appropriate access restrictions in order to protect private and sensitive information as well as intellectual property.

- To provide access to material in a format and/or environment that presents the original’s significant properties.
Case Study - Stephen Jay Gould Papers

Analog component: 550 linear feet of papers (789 boxes, 119 cartons, 30 flat boxes, and 14 map folders.

File size and number: 59.7 MB and 2,567 files.

Media formats: 98 3 ½” floppy diskettes; 61 5.25” floppy diskettes; 4 sets of punch cards*; 3 computer tapes

File Types: Computer Programs; Data sets; Documents; Spreadsheets

File Formats: ASCII Text; WordPerfect 4.2, 5.0, 5.1, 6.0, 6.1; Microsoft Word 2.0, 6.0, 97, 2000; Microsoft RTF; Microsoft Excel 4.0; Lotus 1-2-3 2.0, etc.

* During processing of the “analog” papers in 2011, another 21 sets of punch cards and more floppy diskettes were found.
D&A – EAD

Descriptive Summary

Title Series VII: Gould Born Digital Material

File Types and Formats
File Type: Computer Program; Data set; Document; Spreadsheet. File Format: ASCII, WordPerfect 4.2, 5.0, 5.1, 6.0, 6.1, Microsoft Word 2.0, 6.0, 97, 2000; Microsoft RTF; Microsoft Excel 4.0; Lotus 1-2-3 2.0

creator Gould, Stephen Jay

Additional Information

Scope and Content note This series consists primarily of the born digital material from the Stephen Jay Gould (SIG) papers. The born digital material was stored in floppy diskettes, tapes and punch cards. The labels, if any, on the computer media are in many cases too brief to tell the contents in the diskettes. The processor has to view the contents of the files to decide the proper EAD component levels which the contents belong to. Since SIG divided his works into "Articles", "Abstracts, Reviews, Letters, etc.", "Natural History Column", and "Books" in his bibliography, the processor followed this arrangement and added "Bibliography & CV", "Teaching", "Rare Books", "Punch Cards", "Misc.", and "Computer Media Photos" as other subseries. Details of the ten subseries are as follows.


Subseries II. Abstracts, Reviews, Letters, etc., is divided into 6 sub-subseries according to the publication years: 1990, 1994, 1996, 1997, 1999, "Unidentified Period". Articles which the years of publication cannot be identified were put in "Unidentified Period". There are 107 files in this subseries.


Subseries V. Bibliography & CV, contains several versions of bibliography and CV of SIG. There are 44 files in this subseries.

Subseries VI. Teaching, contains examination papers, syllabus, and students. There are 12 files in this subseries.

Subseries VII. Rare Books, contains information on the rare books owned by SIG. There are 28 files in this subseries.

Subseries VIII. Punch Cards, contains computer programs and data migrated from one box of punch cards. Data in another box of punch cards is not migrated. There are 11 files in this subseries.
D&A – Facet Browsing

Collection
Subseries
The Structure of Evolutionary Theory (227)
Subseries III: Natural History Column (169)
Subseries X: Computer Media Photos (165)
Subseries II: Abstracts, Reviews, Letters, etc (99)
Subjects
Evolution (Biology) (261)
Punctuated equilibrium (Evolution) (227)
Natural History (51)
Evolution (13)
Life-Origin (3)

Title
Filename
Content Type
Document (692)
Image (165)

Media Format
3.5 inch. floppy diskettes (472)
5.25 inch. floppy diskettes (230)
(165)

File Format:
WordPerfect 6.0 (228)
D&A – Full text search
Reconstructing (and Deconstructing) the Past

Stephen Jay Gould

Museum of Comparative Zoology
Harvard University
Cambridge, MA 02138

Framed on the Blas

The Great Exhibition of 1851 did wonders for the morale of two central figures in Victoria's England—her husband Prince Albert, who directed this magnificent show of might and industry at the Crystal Palace, and who thereby won respect and accolades from his previously suspicious subjects; and for Charles Darwin, a frequent visitor, who viewed this vast yet transparent edifice as a sign that his previously fragile notion had now become a stable and fertile ground for an intellectual revolution that he had been guarding in silence since the late 1830's.

When the Exhibition closed at its original home in Hyde Park, workmen dismantled the innovative modular building of steel and glass and reerected the Crystal Palace in the suburb of Sydenham. Among the varied attractions commissioned for the grounds of the Crystal Palace's new home, none was as spurious as, innovative as, fecund as, and enduring as the set of 8ft.-tall models of prehistoric beasts built by the London sculptor Edward Hawkins (1807-1889) with the close collaboration of England's greatest anatomist Richard Owen (1804-1892), inventor of the term "dinosaur."

As Crystal Palace burned in 1856, but Hawkins's models are still in Sydenham (recently reorganized,
D&A – Tag & Annotation by Invited Persons / Public

<table>
<thead>
<tr>
<th>Title:</th>
<th>The Book of Life</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subseries:</td>
<td></td>
</tr>
<tr>
<td>Author:</td>
<td></td>
</tr>
<tr>
<td>Contributor:</td>
<td></td>
</tr>
<tr>
<td>Date:</td>
<td></td>
</tr>
<tr>
<td>Subject Values:</td>
<td>Evolution (Biology)</td>
</tr>
<tr>
<td>Archivist Tags:</td>
<td></td>
</tr>
<tr>
<td>Donor Tags:</td>
<td></td>
</tr>
<tr>
<td>Document Type:</td>
<td>Document</td>
</tr>
<tr>
<td>File Type:</td>
<td>WordPerfect 5.0</td>
</tr>
<tr>
<td>Format Type:</td>
<td>3.5 inch. flo</td>
</tr>
</tbody>
</table>
Impacts from Collection Development

• File formats: no restriction

• Computer medium: no restriction (punch card, open reel tape, 5.25 inch floppy, 3.5 inch floppy),

• File type: no restriction (computer program, data set, document, spreadsheet),

• Agreement: permission to post contents online.
Impacts from Accessioning

• Built 5.25 inch floppy capture station

• Ask Computer History Museum to read punch cards

• Open reel tapes – still outstanding
Impacts from Processing

• AccessData FTK was used to search files with restricted information, annotate files with appropriate descriptive metadata (book title, articles, etc.), and rights metadata (access restriction), generate technical metadata for the delivery platform to act upon.

• Transit Solution was used to transform files to html format for display in web.

• A XSLT program was written to transform the XSL-FO output from FTK to XML content document. A Ruby program was written to ingest the XML content document, original files, and the display derivatives to Fedora.
FTK – Bookmark and Label
FTK – Full Text, Pattern Search & Fuzzy Hash
Emulation – Design Files
Network Diagram for 50,000 Creeley Emails
MUSE: Sentiment Analysis for Emails
Hi Chris,

Thank so much for your wonderful work. We are getting nearer to what we can show people in our AIMS meeting next week. It will be great if you can get access to the whole Gould collection asap. The files on sall-powervault are generated after our Skype call and include many of the features missing in the 'Full House' files and the ETA has been updated as well.

Thanks & Regards,

Peter

----- Original Message ----- 
From: "Chris Fitzpatrick" <cfitz@stanford.edu>
To: "Peter Chan" <pchan@stanford.edu>
Cc: "Glynn Edwards" <gedwards@stanford.edu>, "sulair-aims" <sulair-aims@lists.stanford.edu>, "Lynn McRae" <lmcrate@stanford.edu>, "Tom Craemer" <tcramer@stanford.edu>
Sent: Tuesday, August 31, 2010 9:13:22 PM
Subject: hydra-aims-dev application

Hello everyone,

I've done a kinda first pass at tweaking the hydrangea application for the AIMS demo next week. I've loaded in the files that Peter sent me last week (the "Full House" files), added some minimal metadata, and processed the html output into some jp2000s and text files for the document viewer...

You can see the application here --> http://hydra-aims-dev.stanford.edu/
Want to know more?

http://born-digital-archives.blogspot.com

Gretchen Gueguen
gmg2n@virginia.edu

Mark Matienzo
mark.matienzo@yale.edu

Simon Wilson
s.wilson@hull.ac.uk

Peter Chan
pchan3@stanford.edu