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

Partners


UNIVERSITY of VIRGINIA


STANFORD
UNIVERSITY


UNIVERSITY OF Hull


Yale University

Grant Goals

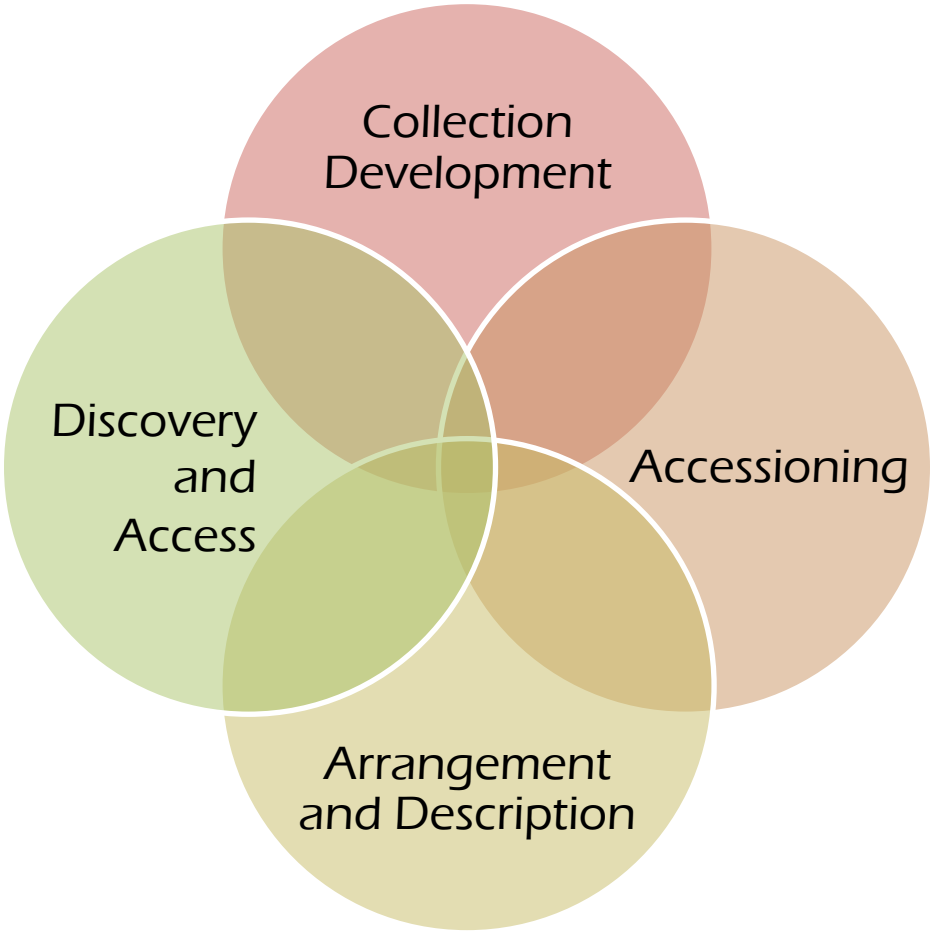
- Processing of Hybrid Collections
- Software Development
- Community Development
  - Unconference (May 2011, Charlottesville, VA)
  - UK Symposium (June 2011, London, England)
  - Workshop (August 2011, Chicago, IL)
- White Paper and Project Report

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.

University of Virginia
University of Hull
Stanford University
Yale
The Andrew W. Mellon Foundation

AIMS Framework



Collection Development

Gretchen Gueguen
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…

Neil Beagrie, "Plenty of Room at the Bottom? Personal Digital Libraries and Collections," *D-Lib Magazine* (June 2005)

Blagofaire. http://xkcd.com/239/

Do...

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 computers / 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 setting 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…

Prepare for Accessioning…

Scope and extent determined?

Method and time determined?

Pre-acquisition appraisal performed?

Test capture if needed?

Coordination with acquisition of analog material?

Enhanced curation carried out?

Development of new methodologies undertaken as needed/possible?

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

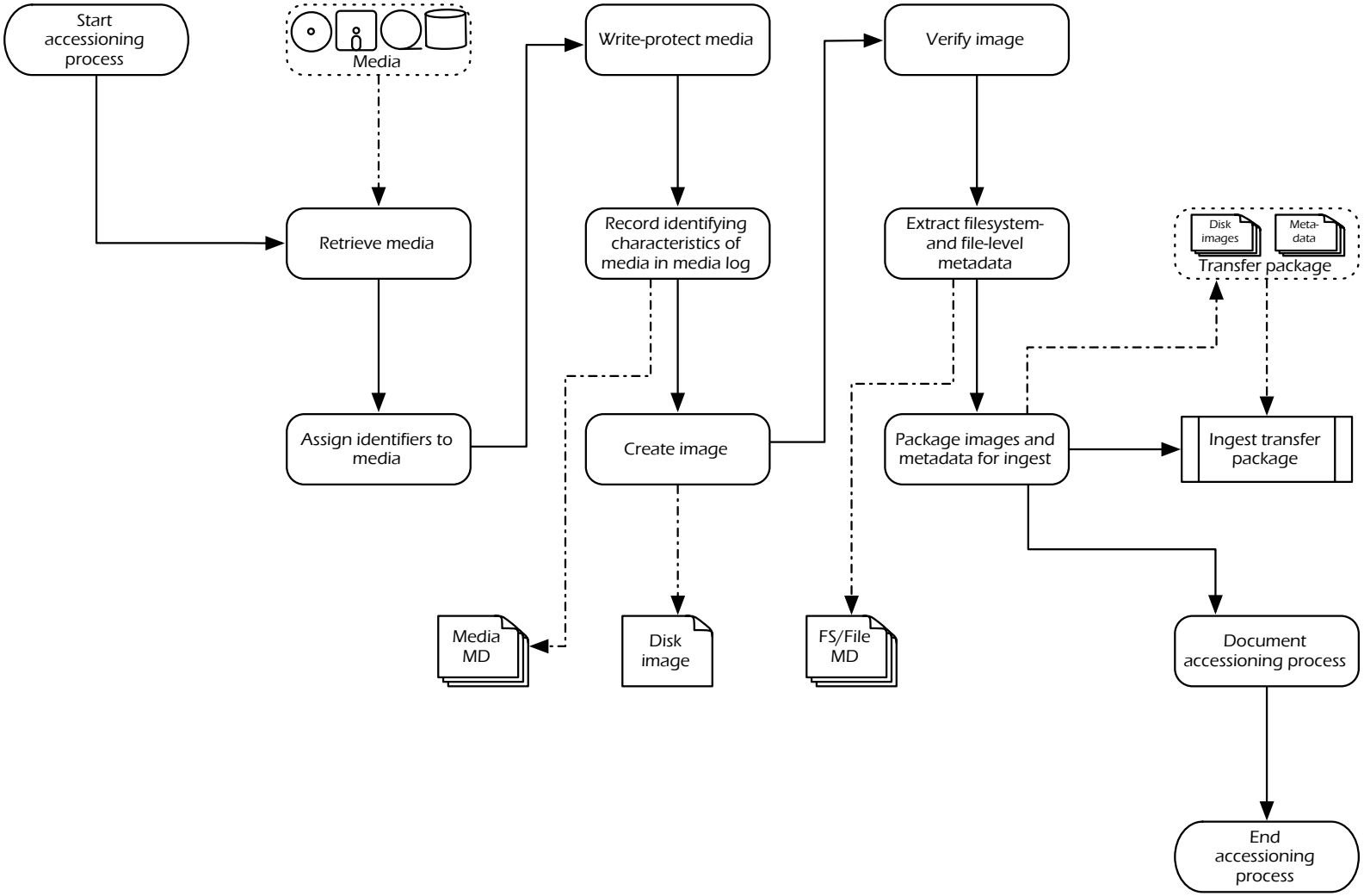
Types of Records and Media

- 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



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

AccessData FTK Imager 2.9.0.1385

File View Mode Help

Evidence Tree

- 2004-M-008.dd-0005.001
  - C910927A001 [FAT12]
    - [root]
    - [unallocated space]

File List

Name	Size	Type	Date Modified
!w0000	59 KB	Regular File	7/21/1997 7:58...
Hangman.bak	56 KB	Regular File	6/10/1997 5:23...
Hangman.bak	58 KB	Regular File	6/10/1997 5:36...
Hangman.bak	59 KB	Regular File	7/21/1997 7:57...
Hangman.tex	0 KB	Regular File	6/10/1997 5:23...
Hangman.tex	59 KB	Regular File	7/21/1997 7:58...
KENNEDY.WPD	41 KB	Regular File	10/17/1997 11:...
Q3.DIR	1 KB	Regular File	7/5/1999 11:18...
Q3.DIR.FileSlack	1 KB	File Slack	
QDATA.ABD	10 KB	Regular File	7/5/1999 11:12...

```
0000 43 39 31 30 39 32 37 41-30 30 31 28 00 00 00 00 C910927A001(····
0010 00 00 00 00 00 00 4C 77-1A 21 00 00 00 00 00 00 ·····Lw·!·····
0020 51 44 41 54 41 20 20 20-51 44 46 20 00 43 48 5A QDATA   QDF ·CHZ
0030 E5 26 31 3D 00 00 48 5A-E5 26 02 00 C0 7D 04 00 å&1=··HZå&··À}··
0040 51 44 41 54 41 20 20 20-51 53 44 20 00 B4 4C 5A QDATA   QSD ·´LZ
0050 E5 26 31 3D 00 00 48 5A-E5 26 92 02 60 49 00 00 å&1=··HZå&··`I··
0060 51 44 41 54 41 20 20 20-51 45 4C 20 00 66 4D 5A QDATA   QEL ·fMZ
0070 E5 26 31 3D 00 00 65 B7-B6 26 B7 02 00 3C 00 00 å&1=··e·¶&···<··
0080 51 44 41 54 41 20 20 20-41 42 44 20 00 16 4E 5A QDATA   ABD ··NZ
0090 E5 26 31 3D 00 00 8B 59-E5 26 D5 02 97 26 00 00 å&1=···Yå&Õ··&··
00a0 51 33 20 20 20 20 20 20-44 49 52 20 00 88 4E 5A Q3      DIR ··NZ
00b0 E5 26 31 3D 00 00 4E 5A-E5 26 E9 02 17 00 00 00 å&1=··NZå&é·····
00c0 E5 41 4E 47 4D 41 4E 20-42 4B 21 20 00 13 F9 85 åANGMAN BK! ··ù·
00d0 CA 22 CA 22 00 00 FA 85-CA 22 23 00 77 F7 00 00 Ê"Ê"··ú·Ê"#·w÷··
00e0 E5 41 4E 47 4D 41 4E 20-42 4B 21 20 00 2F 41 89 åANGMAN BK! ·/A·
00f0 CA 22 CA 22 00 00 44 89-CA 22 02 00 91 F7 00 00 Ê"Ê"··D·Ê"···÷··
0100 E5 41 4E 47 4D 41 4E 20-54 58 54 20 00 6B 65 89 åANGMAN TXT ·ke·
0110 CA 22 CA 22 00 00 66 89-CA 22 03 00 41 F9 00 00 Ê"Ê"··f·Ê"··Aù··
0120 E5 48 00 61 00 6E 00 67-00 6D 00 0F 00 B6 61 00 åH·a·n·g·m···¶a·
0130 6E 00 2E 00 74 00 65 00-78 00 00 00 00 00 FF FF n·.·t·e·x·····ÿÿ
0140 E5 41 4E 47 4D 41 4E 20-54 45 58 20 00 A0 F6 8A åANGMAN TEX · ö·
0150 CA 22 CA 22 00 00 F6 8A-CA 22 00 00 00 00 00 00 Ê"Ê"··ö·Ê"·····
0160 E5 57 30 30 30 30 20 20-20 20 20 20 08 1C F7 8A åW0000      ··÷·
0170 CA 22 CA 22 00 00 F9 8A-CA 22 04 00 8C DD 00 00 Ê"Ê"··ù·Ê"···Ý··
0180 E5 48 00 61 00 6E 00 67-00 6D 00 0F 00 22 61 00 åH·a·n·g·m···"a·
```

Cursor pos = 0; log sec = 19

Custom Content Sources

Evidence:File System|Path|File	Options

New Edit Remove Remove All Create Image

Properties | Hex Value Inter... | Custom Content...

2004-M-008.dd-0005.001/C910927A001 [FAT12]/[root]

NUM

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

New | Actions | Settings | View: All Item

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

Media Log

Electronic Records on Media Accessioning Log

New | Actions | Settings

Type	Media number	Media Format	Imaging Date	Imaging Successful?	Bag Creat
	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-	CD-R		Yes	No
	2011-M-075.0008	CD-R		Yes	No
	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

Electronic Records on Media Accessioning Log: 2011-M-075.0008

Close

New Item | Edit Item | Delete Item | Manage Permissions | Alert Me

Media number	2011-M-075.0008
Media Format	CD-R
Media Density (floppies only)	N/A
Interface	N/A
Label text	Osaka Monograph Final Images Aug 29 2003 Monograph Latest Files
Manufacturer	
Serial Number (hard drives only)	
Examiner	Glick, Kevin
Imaging Successful?	Yes
Imaging Date	
Image filename	2011-M-075.0008.ISO
Source File System	ISO9660, Joliet
Image format	ISO
Imaging Software	ImgBurn
Image Fixity Function	MD5
Image Fixity Value	dbca43c94690edff07329b6687550f60
Notes	mam54 04/28/2011: Could not extract metadata using fiwalk; log file from imaging process says that the block structure is Mode 2/Form 1
Metadata Extracted?	No
Bag Created?	No
Transfer to Storage Date	
Fiscal Year	2010-11

Created at 4/27/2011 9:35 AM by Glick, Kevin
Last modified at 4/28/2011 4:26 PM by Matienzo, Mark

Close

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

Create New Bag | Open Existing Bag | Create Bag In Place | Validate Bag | Is Bag Complete | Close Bag | Save Bag | Save Bag As...

Bagger

Payload

- data
  - 2004-M-088.0001
  - 2004-M-088.0002
  - 2004-M-088.0003
  - 2004-M-088.0004
    - 2004-M-088.0004.dd
    - 2004-M-088.0004.txt
    - 2004-M-088.0004.xml
  - 2004-M-088.0005
  - 2004-M-088.0006
  - 2004-M-088.0007
  - 2004-M-088.0008
  - 2004-M-088.0009
  - 2004-M-088.0010

Tag Files

- mssa
  - bagit.txt
  - tagmanifest-md5.txt
  - bag-info.txt
  - manifest-md5.txt

Bag Info

File name: /Users/mam54/Desktop/4n6/mssa.2004-M-088
Profile: YUL_DISKIMG_ACCN_SIP_0.1 — Bag version: 0.96
Holey Bag?: false — Serialize Type?: none

Bag-Info

Standard : Add

Accession-Number	R	mssa.2004-M-088
External-Identifier	R	mssa.2004-M-088
Bag-Size		35.7 MB
Payload-Oxum		37381126.81
Source-Organization	R	Manuscripts and Archives, Yale University Library
External-Description	R	Forms part of James Tobin papers. Disk images of 3.5 inch floppy disks; originals from circa 1987-1999.
Profile Name	R	YUL_DISKIMG_ACCN_SIP_0.1
Bagging-Date		2011-04-11
Forms-Part-Of	R	mssa.ms.1746

Console

Complete: Valid: Profile Compliant:

[Wed Aug 17 21:49:58 EDT 2011]: This space will contain messages generated by the creating and updating of bags.

Mouse over the status label to see it's description

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

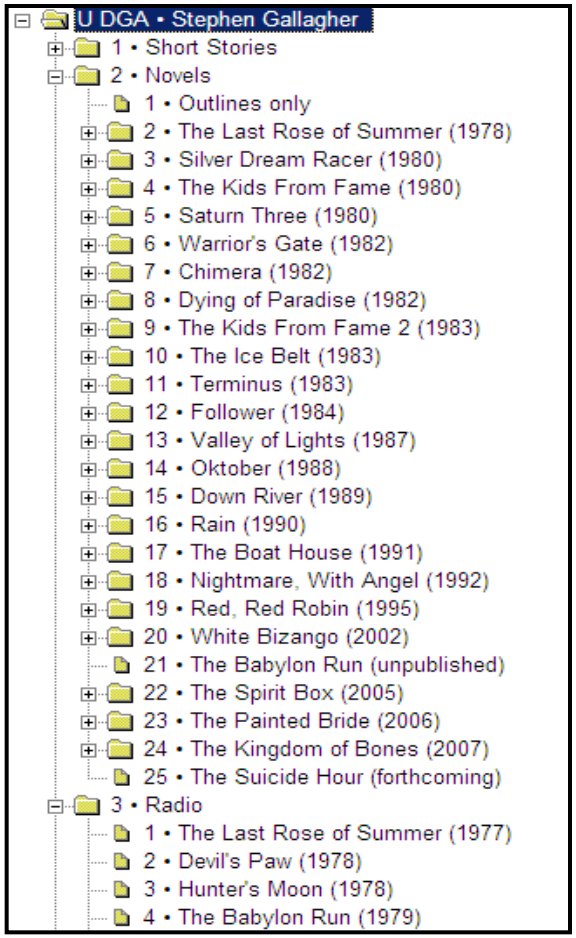
Case Study - Stephen Gallagher

Approach:

- current work higher priority in filing system
- considered each work a distinct 'project'

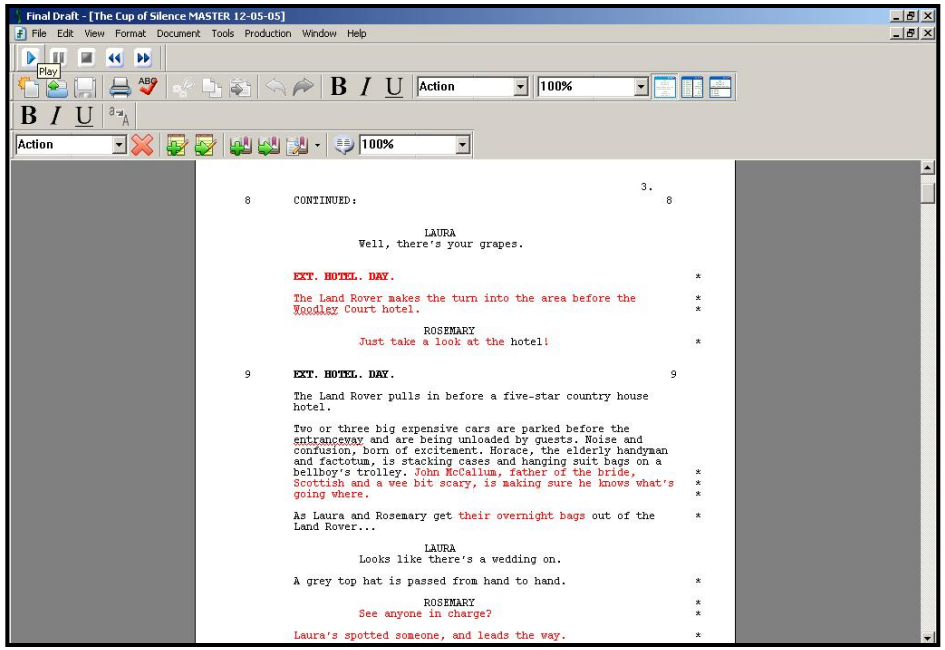
- structure reflect **his way of working** & the **archival principles of control** that creator, archivist & user can all understand

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

Hypatia

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

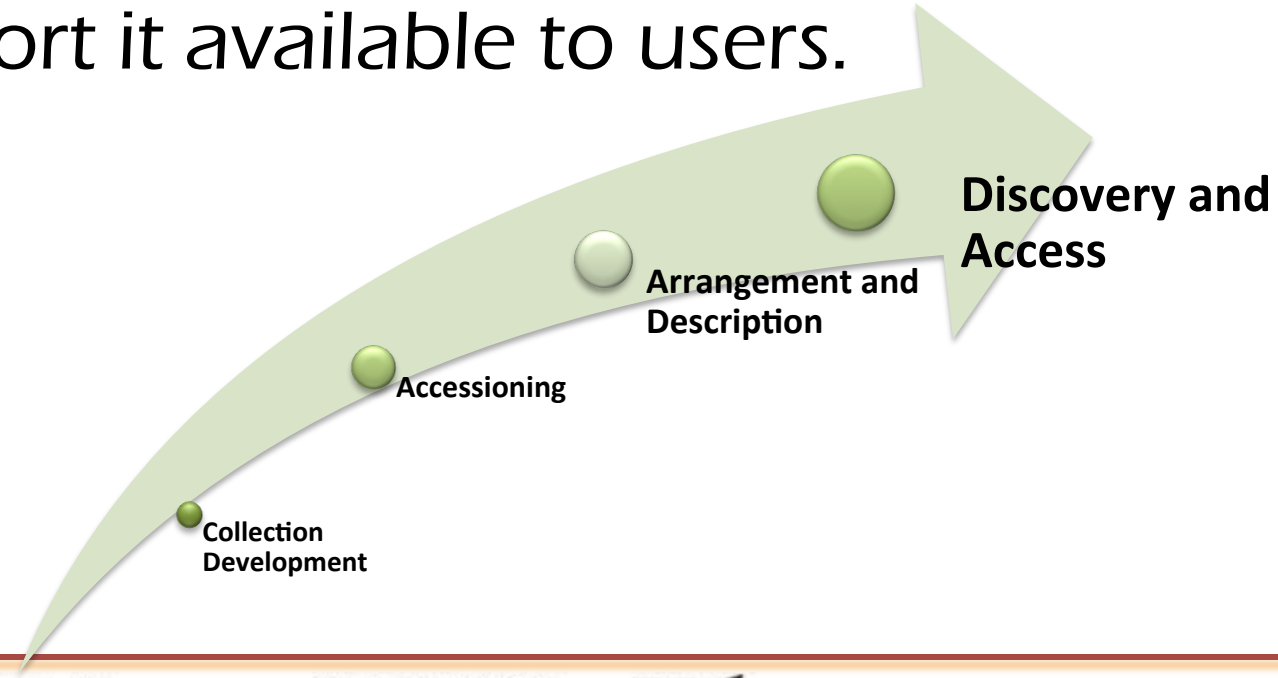
Technical (acquired at accessioning) and descriptive metadata - Discovery & Access process

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 VI: Gould Born Digital Material
File Types and Formats	File Type: Computer Program; Data set; Document; Spreadsheet. File Format: 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
creator	Gould, Stephen Jay

Additional Information

Scope and Contents note

This series consists primarily of the born digital material from the Stephen Jay Gould (SJG) 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 SJG 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 I, Articles, is divided into 15 sub-subseries according to the publication years: : 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2001, "Unidentified Period". Articles which the years of publication cannot be identified were put in "Unidentified Period". There are 99 files in this subseries.

Subseries II, Abstracts, Reviews, Letters, etc., is divided into 6 sub-subseries according to the publication years: 1992, 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 III, Natural History Column, is divided into 12 sub-subseries according to the publication years: 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, "Unidentified Period". Articles which the years of publication cannot be identified were put in "Unidentified Period". There are 171 files in this subseries.

Subseries IV, Books, contains drafts of 12 books written by SJG: The Structure of Evolutionary Theory, Full House, The Book of Life, Triumph and Tragedy in Mudville, Dinosaur in a Haystack, The Burgess Shale and the Nature of History, Time's Arrow, time's Cycle, The Lying Stones of Marrakech, Eight Little Piggies, Hidden Histories of Science, The Hedgehog, the Fox, and the Magister's Pox, The Mismeasure of Man. There are 404 files in this subseries.

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

Subseries VI, Teaching, contains examination papers, syllabus, and students' [illegible]. There are 12 [illegible] this subseries.

Subseries VII, Rare Books, contains information on the rare books owned by SJG. 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)

Suberies III: Natural History Column (169)

Subseries X Computer Media Photos (165)

Subseries II: Abstracts, Reviews, Letters, etc (99)

Subseries I: Articles (85)

more subseries

Subjects

Evolution (Biology) (281)

Punctuated equilibrium (Evolution) (227)

Natural history (51)

Evolution (13)

Life--Origin (3)

more subjects

Title

Filename

Content Type

Document (692)

Image (165)

Media Format

3.5 inch. floppy diskettes (472)

5.25 inch. floppy diskettes (20)

(163)

File Format

WordPerfect 6.0 (228)

Steven J. Gould

Browse the Collection

Biography When five-year-old Stephen Jay Gould first marveled at the towering Tyrannosaurus skeleton in the American Museum of Natural History, he decided to spend his life studying fossils. Although few children in Queens, New York, shared his early fascination for evolution, he never considered any other career but paleontology.

Now professor at Harvard University and curator of its Museum of Comparative Zoology, Gould attended Antioch College, then returned to Manhattan, for graduate work in paleontology at Columbia University. For his doctoral thesis he investigated variation and evolution in an obscure Burmudian land snail, anchoring his later theorizing in intense scrutiny of a single group of organisms, as Darwin had done with Barnacles

D&A – Full text search

hydrangea

Log Out [archivist1]

Add an AIMS Document

Search

Limit your search

Collection
Steven J. Gould (857) [remove]

Subseries
Subjects
Filename
Content Type
Media Format
File Format

You searched for: No Keywords Collection Steven J. Gould

Start over

« Previous Next »

1 2 3 4 5 6 7 8 9 ... 21 22

Sort by relevance

Show 10 per page

857 documents

Title	Creator	Deposited	Type	files	Status
druid:940		Sun Sep 05 02:36:19 UTC 2010	Document	1	Created
druid:944		Sun Sep 05 02:41:25 UTC 2010	Document	1	Created
druid:948		Sun Sep 05 02:42:04 UTC 2010	Document	1	Created
druid:952		Sun Sep 05 02:42:44 UTC 2010	Document	1	Created
druid:956		Sun Sep 05 02:43:24 UTC 2010	Document	1	Created
druid:960		Sun Sep 05 02:48:35 UTC 2010	Document	1	Created
druid:964		Sun Sep 05 02:49:10 UTC 2010	Document	1	Created
druid:968		Sun Sep 05 02:49:49 UTC 2010	Document	1	Created
druid:974		Sun Sep 05 16:16:25 UTC 2010	Document	0	Created
druid:976		Sun Sep 05 16:17:31 UTC 2010	Document	0	Created
druid:978		Sun Sep 05 16:18:08 UTC 2010	Document	0	Created
druid:980		Sun Sep 05 16:18:46 UTC 2010	Document	0	Created
druid:983		Sun Sep 05 16:19:22 UTC 2010	Document	0	Created
druid:986		Sun Sep 05 16:20:00 UTC 2010	Document	0	Created
druid:989		Sun Sep 05 16:20:31 UTC 2010	Document	0	Created
druid:992		Sun Sep 05 16:21:02 UTC 2010	Document	0	Created

D&A – See Contents on Web

« Back to Results

This document refers to:

Collection

Steven J. Gould (857)

aims_document (1)

Subseries

The Structure of Evolutionary Theory (227)

Suberies III: Natural History Column (169)

Subseries X Computer Media Photos (165)

Subseries II: Abstracts, Reviews, Letters, etc (99)

Subseries I: Articles (85)

Bibliography (29)

Dinosaur in a Haystack (27)

Subseries VII SJG Rare Books (21)

Full House (13)

Subseries VI Teaching (12)

CV (5)

The Book of Life (3)

Triumph and Tragedy in Mudville (2)

more subseries

Subjects

Evolution (Biology) (281)

Punctuated equilibrium (Evolution) (227)

< Previous #2 of 3 Next >

Document Viewer

full screen

DOCUMENT ANNOTATIONS TEXT Zoom Level Search

p. 1

Reconstructing (and Deconstructing) the Past

Stephen Jay Gould

Museum of Comparative Zoology

Harvard University

Cambridge, MA 02138

I Framed on the Bias

The Great Exhibition of 1851 did wonders for the morale of two central figures in Victoria's England -- for 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 nation 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 spectacular, as innovative, as fecund, and as enduring as the set of life-sized models of prehistoric beasts built by the London sculptor Waterhouse Hawkins (1807-1889) with the close collaboration of England's greatest anatomist Richard Owen (1804-1892), inventor of the term "dinosaur."

The Crystal Palace burned in 1936, but Hawkins's models are still in Sydenham (recently lightened,

D&A – Tag & Annotation by Invited Persons / Public

19NATHID.BK!.wpd (1)
19NATHID.wpd (1)
21NATHI6.wpd (1)
22NATHI6.wpd (1)
295.doc (1)
296.doc (1)
297.doc (1)
299.doc (1)
2SJGBIBL.BK!.wpd (1)
2SJGBIBL.WPD (1)
300.doc (1)
301.doc (1)
302.doc (1)
9BASEBAB.wp5 (1)
ACKCER3 (1)
AFRISILE.wp (1)
ALLOMIND.BK!.wpd (1)
ALLOMIND.WPD (1)
ANDERSON.W50.wp5 (1)
BABERUTH.wp5 (1)
BACON.wp5 (1)
BASBA'93.BAK.wp5 (1)
BASBA'93.wp5 (1)
BASE'94.wp5 (1)
BELLCURV.wp5 (1)
BIBA..doc (1)
BIBA.BK!.wp5 (1)
BIBA.doc (1)
BIBA.wp5 (1)
BIBA.wpd (1)
BIBB..doc (1)
BIBB.BK!.wp5 (1)
BIBB.wp5 (1)
BIBB.wpd (1)
BIBC..doc (1)
BIBC.wp5 (1)
BIBD..doc (1)
BIBD.wp5 (1)

Original Document (PDF)

Description

Title:

Subseries: The Book of Life

Annotation:

Author: +

Contributor: +

Date:

Subject Values: Evolution (Biology)
Life--Origin x
Scientific Illustration x

Archivist Tags: +

Donor Tags: +

Document Type: Document

File Type: WordPerfect 5.0

Format Type: 3.5 inch. flo

Download Source Files

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

AccessData Forensic Toolkit Version: 3.2.0.32216 Database: localhost Case: M1437 Stephen Jay Gould

File Edit View Evidence Filter Tools Manage Help

Filter: - unfiltered - Filter Manager...

Explore Overview Email Graphics Bookmarks Live Search Index Search Volatile

Evidence Items

Evidence
- CM004.ad1
  - A::A:\
- CM005.ad1
- CM006.ad1
- CM007.ad1
- CM008.ad1
- CM009.ad1
  - A::A:\
- CM010.ad1
  - A::A:\
- CM011.ad1
- CM012.ad1
- CM013.ad1
- CM014.ad1
- CM015.ad1
- CM016.ad1
- CM017.ad1
- CM018.ad1
- CM019.ad1
- CM020.ad1
- CM021.ad1
- CM022.ad1
- CM023.ad1
- CM024.ad1
- CM025.ad1
- CM026.ad1
- CM027.ad1
- CM028.ad1
- CM029.ad1

File Content

Hex Text Filter

Fig. I-1 ... l940. This original illustration in
the Museum ... econstruction in National Geography
for February ... groups. Above Sidneyia, the
largest anima ... belong to the unique creature
Anomalocaris ... the rear end of a bivalved arthropod
(the large cre ...
Fig. I-2 ... a l985 article by Briggs and
Whittington ... right, and the figure is dominated
by two specin ... dd organisms. Three Aysheaia feed
on sponges a ... tom just left of Aysheaia. Two
Wiwaxia graz ...

File Content

Add Files to Bookmark

Files to Add

All Highlighted All Checked All Listed

1 item selected.

Name	Path
FIGCAP	CM004.ad1\A::A:\\FIGCAP

File Comment

Also include

Email Attachments

Parent Email

Bookmark Selection in File

Select Existing Bookmark

- 0 Series VI: Stephan Jay Gould Born-Digital Material
  - 01 Subseries I: Articles
  - 02 Subseries II: Abstracts, Reviews, Letters, etc.
  - 03 Subseries III: Natural History Column
  - 04 Subseries IV Books
  - 05 Subseries V Bibliography & CV
  - 06 Subseries VI Teaching
  - 07 Subseries VII SJG Rare Books

OK Cancel

Default Media Web

File List

Normal (2) Display Time

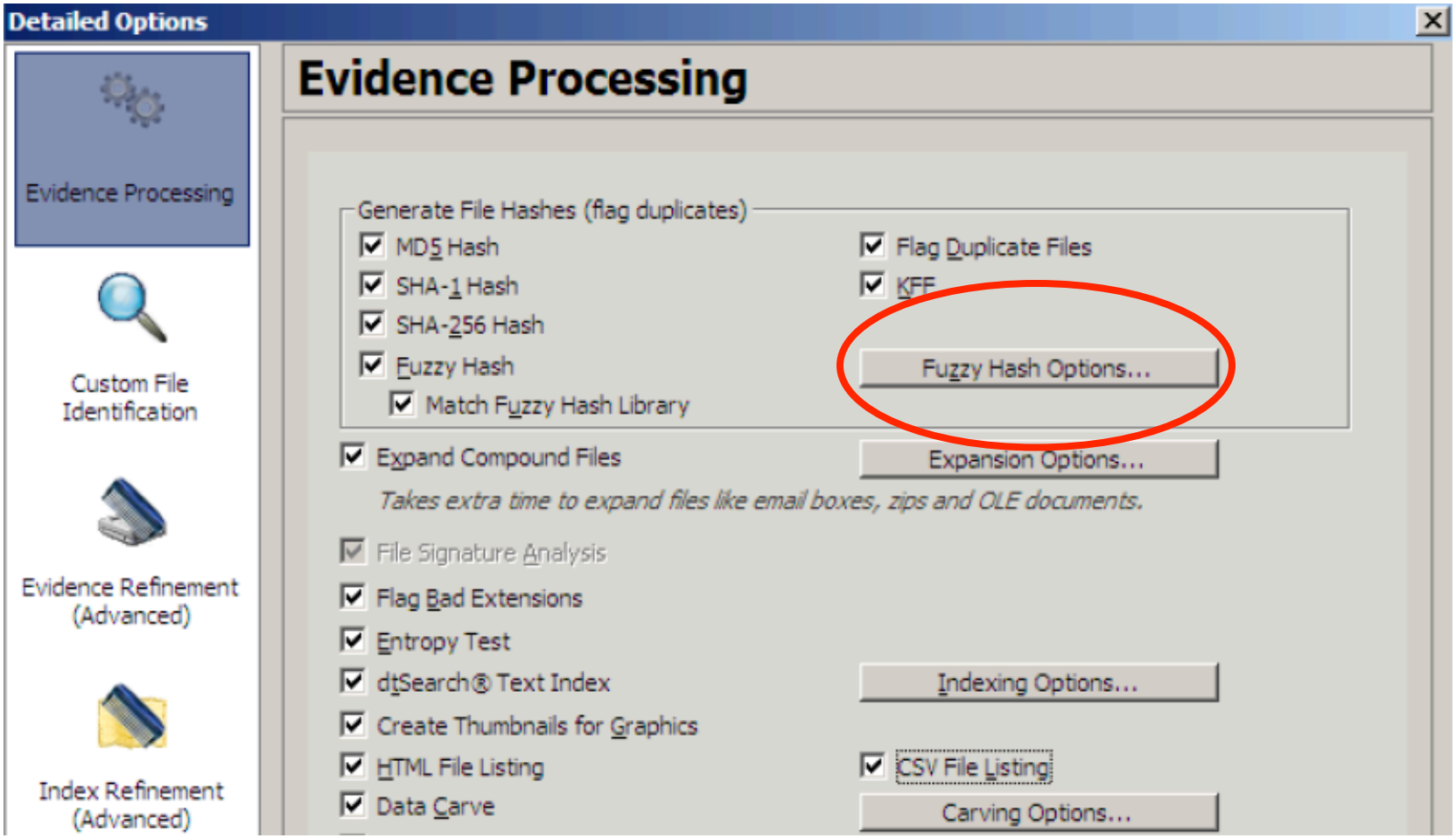
Name	Path	Item #	Ext	Category	L Size	MD5	Created	Accessed	Modified	Duplicate File	Flagge...
BOX3-L	/CM004.ad1/A::A:\/BOX3-L	101007	<missing?>	Unknown	5823 B	DDA8980654401E493A7D92503B6C1437	12/31/1979 ...	12/31/1979 ...	11/29/198...		False
BURBIB	/CM004.ad1/A::A:\/BURBIB	101009	<missing?>	Unknown	19.38 KB	D3608AEE5A0BEAD98551A9EE70301C38	12/31/1979 ...	12/31/1979 ...	11/29/198...		False
FIGCAP	/CM004.ad1/A::A:\/FIGCAP	101003	<missing?>	WordPerfect 4.2	29.07 KB	D791992949FC7F38DB5230E319A525CD	12/31/1979 ...	12/31/1979 ...	12/20/198...		False
PREF	/CM004.ad1/A::A:\/PREF	101008	<missing?>	Unknown	17.71 KB	560B645115FB76777E999701B8D83172	12/31/1979 ...	12/31/1979 ...	4/28/1988 ...		False
TABCON	/CM004.ad1/A::A:\/TABCON	101005	<missing?>	Unknown	4072 B	E57DEE5C5B3CF4B87ADDC6F6887D218C	12/31/1979 ...	12/31/1979 ...	4/28/1988 ...		False
TABLE	/CM004.ad1/A::A:\/TABLE	101004	<missing?>	WordPerfect 4.2	1284 B	5E1CE315E7DB05609C5C31733FFF5982	12/31/1979	12/31/1979	10/20/198		False

Loaded: 8 | Filtered: 8 | Total: 8 | Highlighted: 1 | Checked: 0 | Total LSize: 90.97 KB

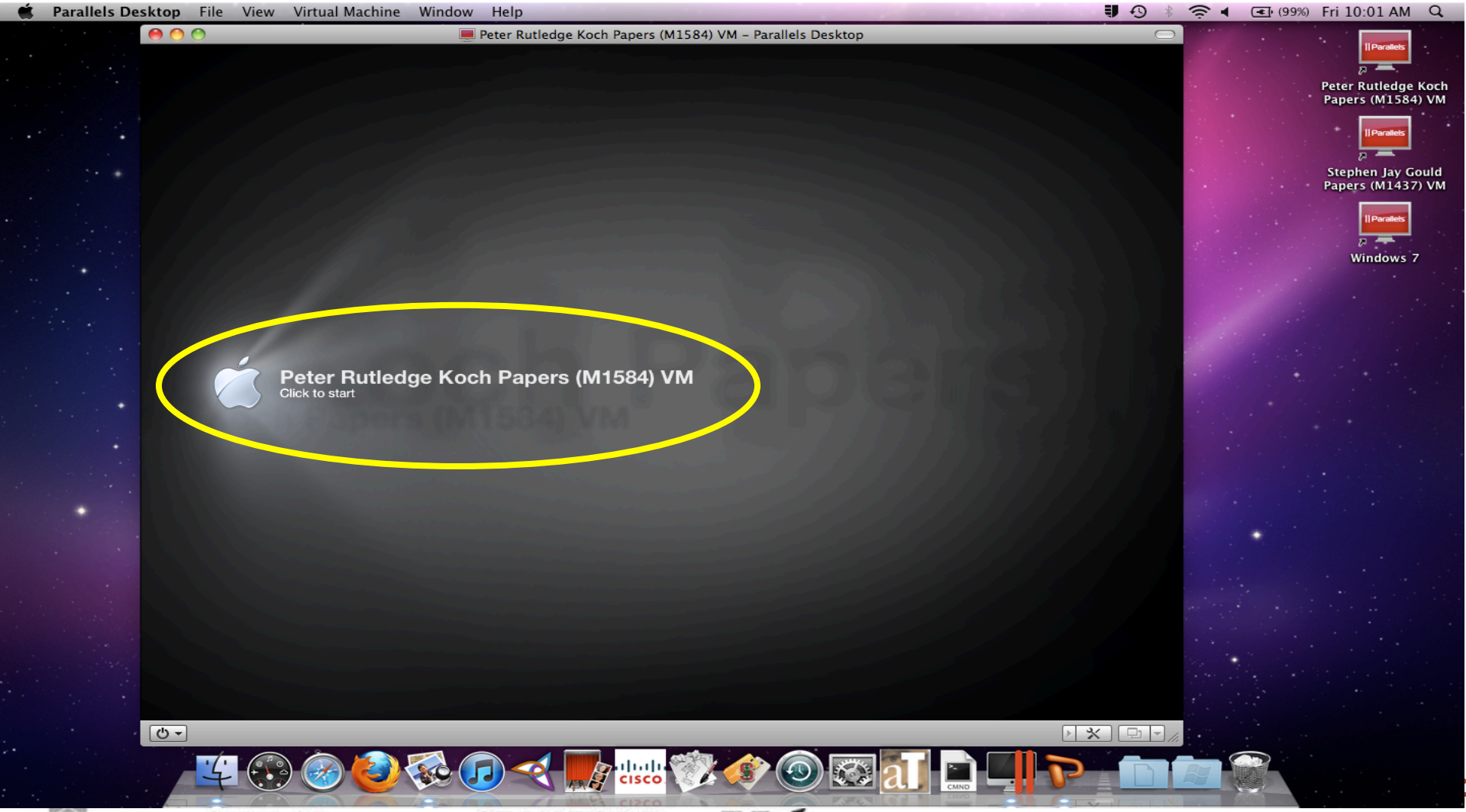
/CM004.ad1/A::A:\/FIGCAP

Ready | Explore Tab Filter: [None]

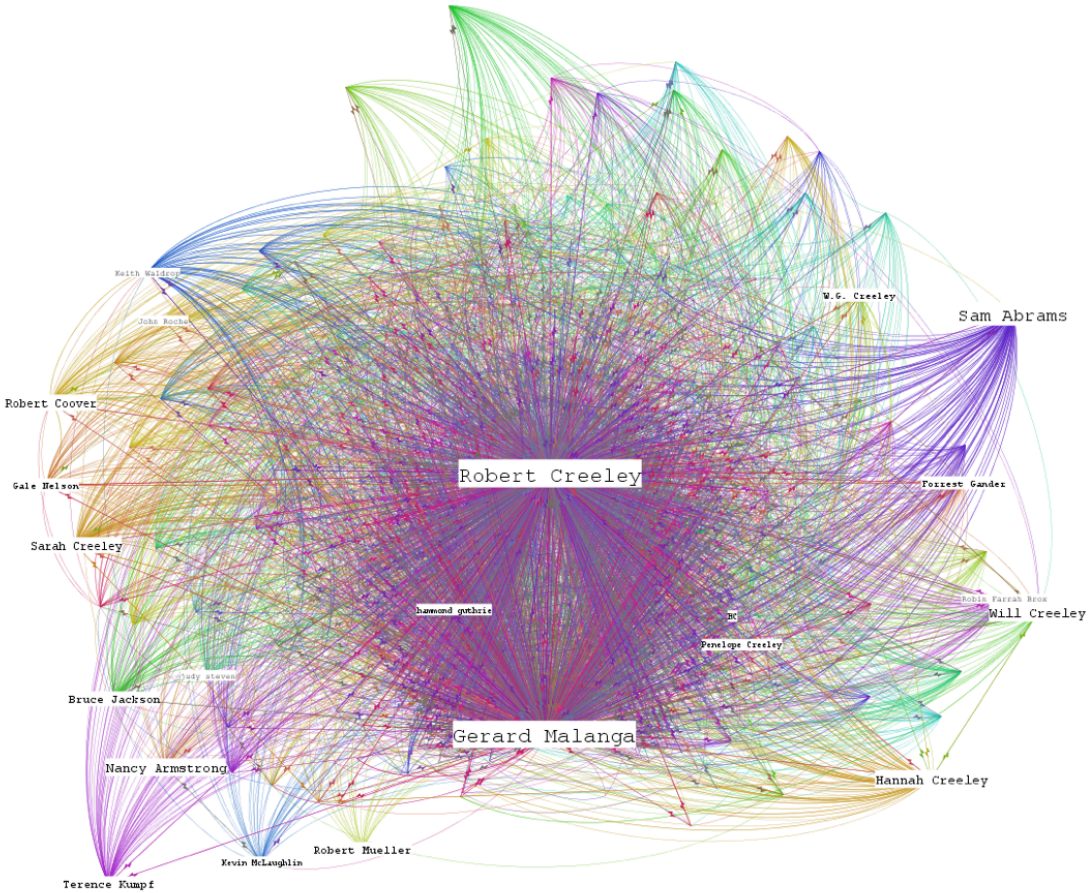
FTK – Full Text, Pattern Search & Fuzzy Hash



Emulation – Design Files



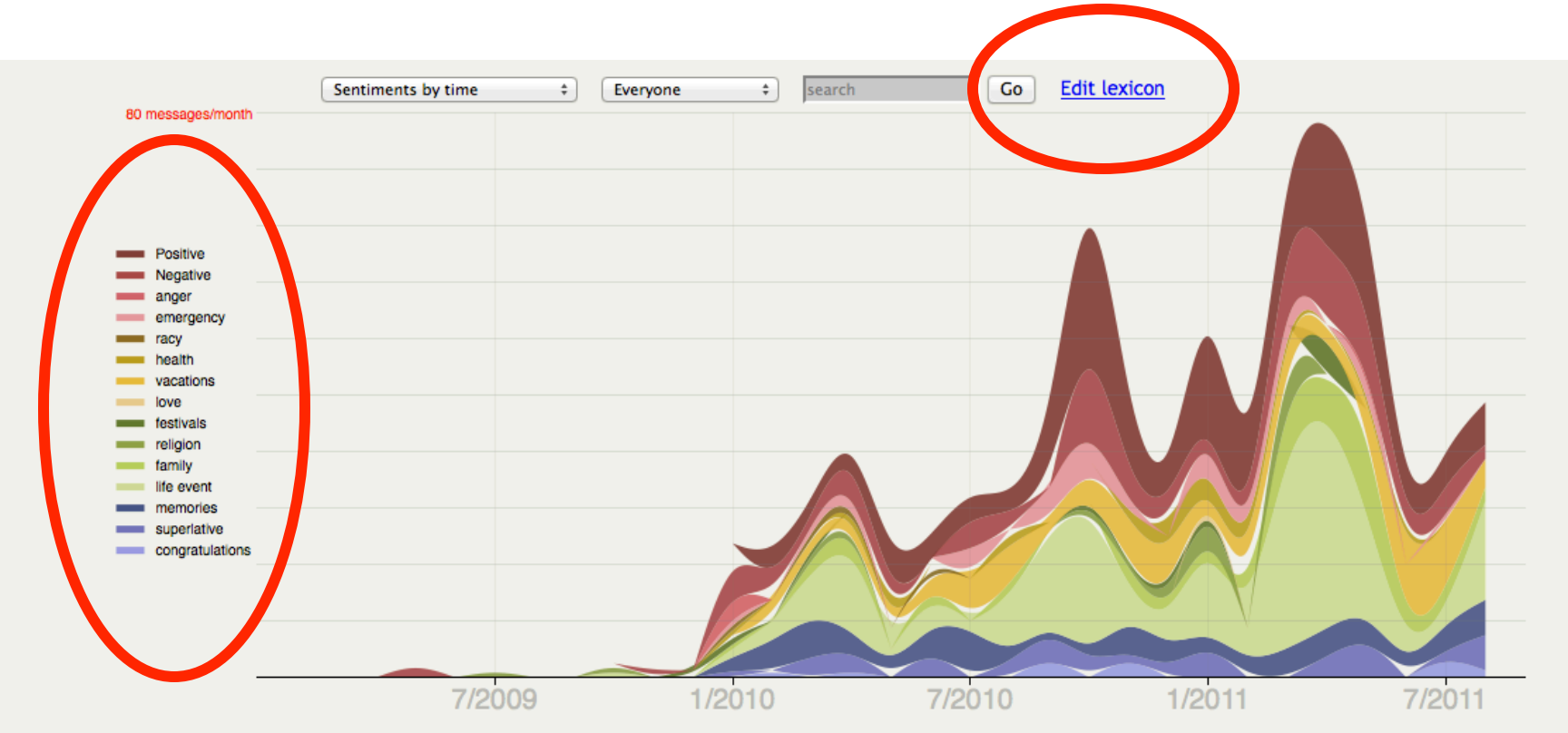
Network Diagram for 50,000 Creeley Emails



AIMS

University of Virginia | University of Hull | Stanford University | Yale | The Andrew W. Mellon Foundation

MUSE: Sentiment Analysis for Emails



MUSE: See Individual Email

Sentiments

- Positive (150)
- life event (33)
- Negative (18)
- family (15)
- vacations (15)

More

Groups

- gedwards (37)
- gedwards+1 (24)
- emeeks (12)
- twl8n (9)
- bjd2b+14 (8)

More

People

- Glynn Edwards (72)
- Michael G Olson (39)
- Elijah Meeks (13)
- Tom Laudeman (12)
- Tom Cramer (10)

More

Click to toggle filter

23/150

Date: August 31, 2010 11:59pm
From: Peter Chan <pchan3@stanford.edu>
To: Chris Fitzpatrick <cfitz@stanford.edu>
Glynn Edwards <gedwards@stanford.edu>,
Cc: sulair-aims <sulair-aims@lists.stanford.edu>, Lynn McRae <lmcrae@stanford.edu>, Tom Cramer <tcramer@stanford.edu>
Group: cfitz+4
Subject: **Re: hydra-aims-dev application**

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 sul-powervault are generated after our Skype call and include many of the features missing in the "Full House" files and the EAD has been updated as well.
Thanks & Regards,
Peter
----- Original Message -----
From: "Chris Fitzpatrick" <cfitz@stanford.edu>
To: "Peter Chan" <pchan3@stanford.edu>
Cc: "Glynn Edwards" <gedwards@stanford.edu>, "sulair-aims" <sulair-aims@lists.stanford.edu>, "Lynn McRae" <lmcrae@stanford.edu>, "Tom Cramer" <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/

University of Virginia · University of Hull · Stanford University · Yale · The Andrew W. Mellon Foundation

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

University of Virginia · University of Hull · Stanford University · Yale · The Andrew W. Mellon Foundation