Collections Inventory

Inventory of Hypatia Collections

Preparation of Collections for Hypatia

Collection Name / Institution	All Files on SUL-BRICK	Analysis	Prototype Fixture Objects (coll, set, item, file ...)	Hooks from item to file objects addressed	Ingest Processor Outputs Tested and Approved	Hypatia App Tests Fixture Objects	Collection Processed into Staging Fedora	Collection Processed into Production Fedora	Hypatia App Has Data
Xanadu / Stanford • EAD (collection and item / no FTK)	✅	✅	✅	✅	Stanford	Stanford	Stanford	Stanford	Stanford
Gould / Stanford • EAD (collection) / FTK	✅	✅	😉	✅	Stanford	Stanford	Stanford	Stanford	Stanford
Koch / Stanford • EAD (collection) / FTK	✅		Stanford	Stanford	Stanford	Stanford	Stanford	Stanford	Stanford
Creeley /Stanford • EAD (collection) / FTK	✅		Stanford	Stanford	Stanford	Stanford	Stanford	Stanford	Stanford
Gallagher / Hull • EAD (collection and item) / no FTK	✅		Uva		Uva	Stanford			
Socialist Health / Hull • EAD (collection and item) / no FTK	✅		Uva		Uva	Stanford			
Tobin / Yale • EAD (collection and item) / no FTK	✅	✅	😉		Uva	Stanford			
Turner / Yale • EAD (collection and item) / no FTK	✅		Uva		Uva	Stanford			
Cheuse / UVa • EAD (collection and item), FTK	✅	✅	Uva		Uva	Stanford			

General conversion and data mapping

- Hypatia EAD conversion analysis

Stanford

Collection Name	Estimated Size of Collection in Hypatia
M1437 Gould	2.5 GB
M1292 Xanadu	5.0 GB
M0662 Creeley	3.0 GB
M1584 Koch	35 GB

Stephen Jay Gould

The collection was re-processed due to a change in storage location and new ideas on relationships between files and EAD.

Stanford FTK to Hypatia object mapping

Processed files are currently stored in

\\sul-wallaby\ForensicsLab\01-OBJECT_POOL\M1437 Stephen Jay Gould\M1437 Gould

and in Sul-Brick/sulguest/Stanford/M1437 Gould

Directory Structure is as follows:

- Computer Media Photo
- EAD
- FTK html
- FTL xml
- Disk Image
- Transit Solution

"FTK html" folder is used to store report from AccessDataFTK in html.

"FTK xml" folder is used to store report from AccessDataFTK in xml.

"Logical Image" folder is used to store the logical images and the audit logs of disk imaging.

"Transit Solution" folder is used to store the html version of the original files created by Transit Solution.

Xanadu

A Collection consists of 6 hard drives. A Marc record for the collection is available in SearchWorks; a very basic finding aid describes the contents of the collection.

Contents of the collection are currently stored on \\sul-wallaby\ForensicsLab\01-OBJECT_POOL\M1292 Xanadu

Xanadu EAD and Hypatia fixture objects

Directory Structure is as follows:

- Disk Images
- Computer Media Photo
- EAD

The Disk Images folder contains 3 forensic disk images from 3 physical hard drives. The forensic disk images are named CMxx.dd with the "CM" standing for computer media. This folder also contains two additional metadata files for each forensic disk image. The first is a .txt file that contains technical metadata about the forensic imaging process (example CM01.001\). The second is a .csv file that lists the partitions and files contained on the hard drive (example CM01.001\). This file also contains the root path, creation dates, and whether the file was deleted on the media and subsequentially recovered.

The Photo Images of Drives folder contains digital photographs of the source media (JPEG), in this case images of the front and back of the harddrives.

The EAD folder contains the Encoded Archival Description file for the Xanadu collection (example EAD\). This file currently does not contain any pointers to where the hard drives are physically located in the collection. We are also currently missing reference identifiers to the computer media in the finding aid. I believe this is just an oversight but I'm following up with Special Collections to determine why they are missing.

Yale

Summary

Collection title	Number of files/objects	Total Extent in (mega /giga)bytes	Extent to be transferred for development	EAD filename	Level of description of born-digital material
James Tobin papers	27 disk images + metadata (approx 80 files total)	36 MB	36 MB	mssa.ms.1746.bpg.xml	Disks are described individually within EAD as separate components
Henry Ashby Turner papers	~5-10	~200 MB	~80 MB	mssa.ms.1691.bpg.xml	Components represent individual digital objects within a specific subseries
Love Makes a Family records	TBC	~36 GB	TBC	mssa.ms.1962.bpg.xml	Only described at high-level aggregations
Pelli Clarke Pelli records	TBC	~6 GB	TBC	mssa.ms.1939.bpg.xml	Currently completely undescribed
New Haven Oral Histories	TBC	~101 GB	TBC	mssa.ru.1055.bpg.xml	Described as individual "interviews" - audio file + MS Word document
James Welch papers (Beinecke)	TBC	TBC	TBC	beinecke.welch.bpg.xml	TBC

James Tobin papers

- Assets loaded on sul-brick; in directory /home/sulguest3/Yale/mssa.ms.1746. This directory is a BagIt bag.
- All of the assets are related to sub-components within the Computer diskettes (3.5 inch) subcomponent of Accession 2004-M-088.
- Within this directory, each directory has the format 2004-M-088.nnnn (e.g. 2004-M-088.0001)
- Directory names correlate with unitids in the EAD for components that represent individual disk.
- Each directory has three files: a disk image (.dd extension); an imaging log file (.txt); and filesystem level metadata extracted from the disk image (.xml; comparable to the CSV files created by FTK Imager)

Henry Ashby Turner papers

- Assets loaded on sul-brick; in directory /home/sulguest3/Yale/mssa.ms.1691 - there are only 2 files.

- Each file asset is associated with a specific component; in other words, only two components have assets associated with them. The assets are a Microsoft Access database and a FileMaker Pro database.
- The components that have an asset associated with them contain a dao element. This element's xlink:href attribute is a file URI that points to the location on sul-brick (this is a hack, but it should be sufficient)

Virginia

Summary

Collection title	Number of files/objects	Total Extent in (mega/giga)bytes	Extent to be transferred for development	EAD filename	Level of description of born-digital material
Alan Cheuse papers	EAD + FTK output (metadata, plus approx 1,400 files)	approx 55 MB	approx 55 MB	uva10726.xml	disk images were processed using FTK. Labels assigned to FTK objects correspond with values in <unitid> tags. those <unitid>s are listed below.

unitids:

- e002001
- e002002
- e002003
- e002004
- e002005
- e002006
- e002007
- e002007b
- e007
- e0100 – e0144
    - EXCEPT e0136…this disk is unreadable, no FTK content
- e0557-- e0557t
    - EXCEPT e0557r…the disk is unreadable
- e0422 – e0429
    - EXCEPT e0421, e0421a and e0423…unreadable disks

Hull

Files transferred via external hard drive/USB pen drive so no physical media to photograph

Collection title	Number of files /objects	Total Extent (mega /giga) bytes	Extent to be transferred for development	EAD filename	Level of description of born-digital material
Stephen Gallagher	paper records (7.5m) 14,320 digital files (excluding 39 Amstrad disks still to be read)	n/a 13.6 GB tbc	~200 MB files have been rearranged into intellectual order for the demo	U DGA.xml - current (beta) structure of the collection ONLY	Currently working through the material, with detailed series descriptions - novel/screenplay etc being created in CALM
Socialist Health Association	paper records (6.5m) 2558 digital files	n/a 670MB	TBC	U DSM.xml - paper based material ONLY	Preliminary cursory look only - scheduled to start this shortly - focus has been Stephen Gallagher due to the larger volume & complexity