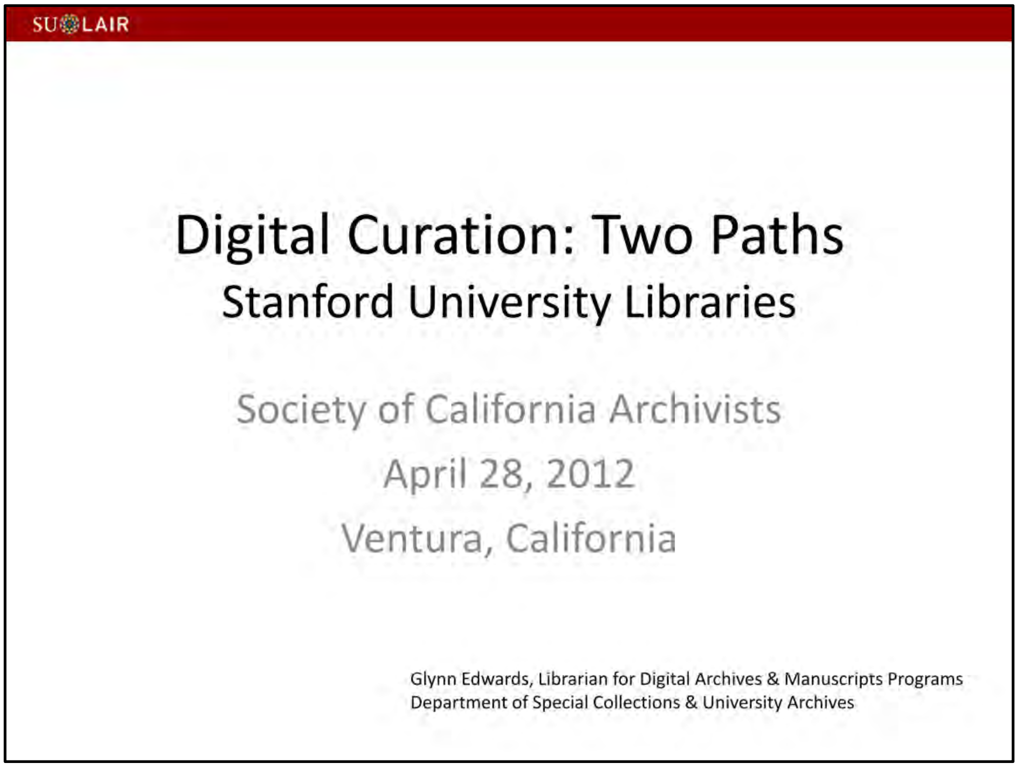


Digital Curation: Two Paths

Stanford University Libraries

Society of California Archivists

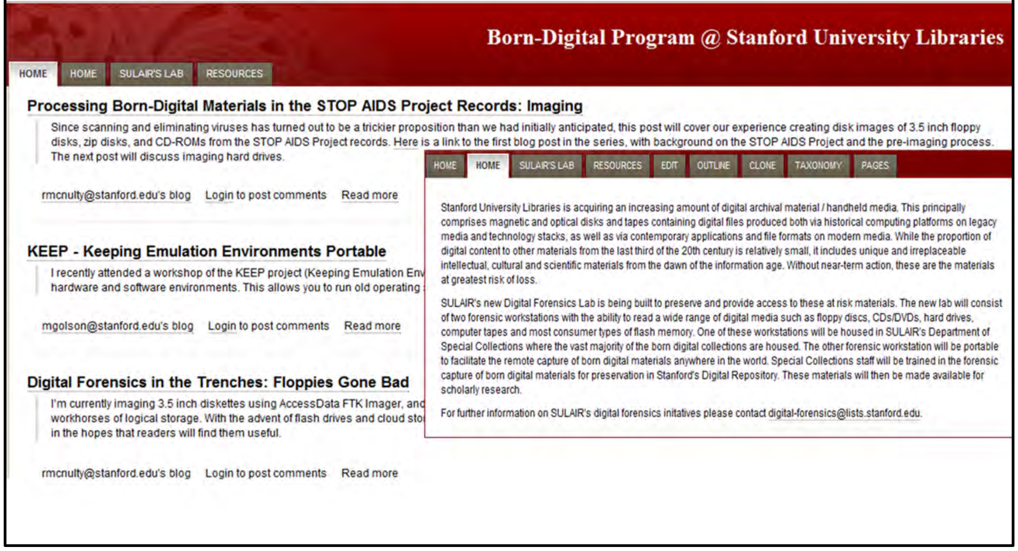
April 28, 2012

Ventura, California

Glynn Edwards, Librarian for Digital Archives & Manuscripts Programs
Department of Special Collections & University Archives

SU LAIR

Stanford University Libraries' Digital Archives Program: Overview



Processing Born-Digital Materials in the STOP AIDS Project Records: Imaging

Since scanning and eliminating viruses has turned out to be a trickier proposition than we had initially anticipated, this post will cover our experience creating disk images of 3.5 inch floppy disks, zip disks, and CD-ROMs from the STOP AIDS Project records. Here is a link to the first blog post in the series, with background on the STOP AIDS Project and the pre-imaging process. The next post will discuss imaging hard drives.

rmcnulty@stanford.edu's blog Login to post comments Read more

KEEP - Keeping Emulation Environments Portable

I recently attended a workshop of the KEEP project (Keeping Emulation Env hardware and software environments. This allows you to run old operating

mgolson@stanford.edu's blog Login to post comments Read more

Digital Forensics in the Trenches: Floppies Gone Bad

I'm currently imaging 3.5 inch diskettes using AccessData FTK Imager, and workhorses of logical storage. With the advent of flash drives and cloud sto in the hopes that readers will find them useful.

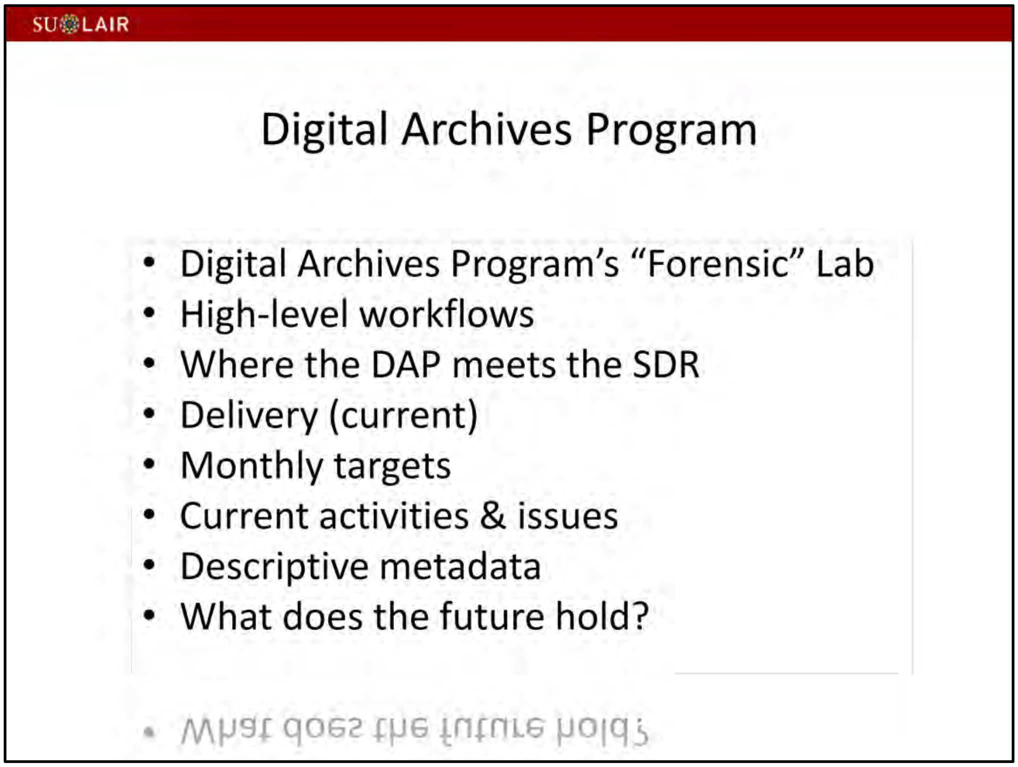
rmcnulty@stanford.edu's blog Login to post comments Read more

Stanford University Libraries is acquiring an increasing amount of digital archival material / handheld media. This principally comprises magnetic and optical disks and tapes containing digital files produced both via historical computing platforms on legacy media and technology stacks, as well as via contemporary applications and file formats on modern media. While the proportion of digital content to other materials from the last third of the 20th century is relatively small, it includes unique and irreplaceable intellectual, cultural and scientific materials from the dawn of the information age. Without near-term action, these are the materials at greatest risk of loss.

SULAIR's new Digital Forensics Lab is being built to preserve and provide access to these at risk materials. The new lab will consist of two forensic workstations with the ability to read a wide range of digital media such as floppy discs, CDs/DVDs, hard drives, computer tapes and most consumer types of flash memory. One of these workstations will be housed in SULAIR's Department of Special Collections where the vast majority of the born digital collections are housed. The other forensic workstation will be portable to facilitate the remote capture of born digital materials anywhere in the world. Special Collections staff will be trained in the forensic capture of born digital materials for preservation in Stanford's Digital Repository. These materials will then be made available for scholarly research.

For further information on SULAIR's digital forensics initatives please contact digital-forensics@lists.stanford.edu.

This morning I will present a brief overview of the (born) Digital Archives Program at the SUL.



Digital Archives Program

- Digital Archives Program’s “Forensic” Lab
- High-level workflows
- Where the DAP meets the SDR
- Delivery (current)
- Monthly targets
- Current activities & issues
- Descriptive metadata
- What does the future hold?

Mine is relatively general and will (hopefully) provide the context needed for Peter’s more detailed and technical focus. As you would imagine a lot has changed since the completion of the AIMS Project in the winter.

SULAIR

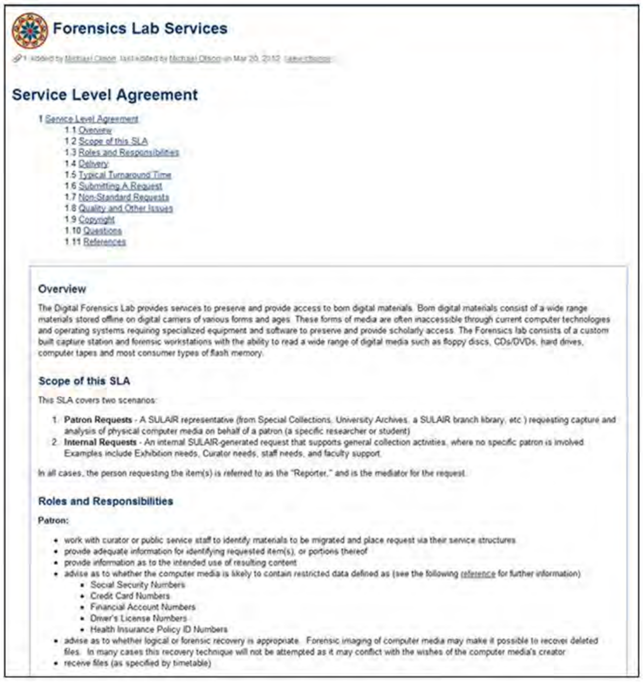
Digital Archives Program: “Forensic” Lab

One of our tasks this winter and spring has been to begin crafting our Mission Statement within SULAIR and at SU. Part of this is contingent on funding to build out the lab (equipment and staff).

The DAP is a hybrid program within the libraries. In some ways it is LIKE other data groups – like Branner Library which assists with GIS data or SSDS which assists with data sets. It is also like the digitization labs at SUL – the SMPL (Stanford Media Preservation Lab) and Map Lab, etc. BUT it is also an acquisition and processing unit – similar to our Manuscripts and Archival units – working in conjunction with other staff and departments within Special Collections, DLSS, etc.

SULAIR

- Communities served
  - Stanford Students
  - Stanford Staff
  - Stanford Faculty
- Types of service
  - Capture for preservation in digital repository
  - Capture for recovery (personal)
  - Training – tools, hardware, software
  - Workshops
  - Assistance with discovery of born-digital archives
- Appraisal assistance
  - Donors/sellers
  - Curators
- Processing assistance/training
- Maintaining list of hardware, software, media formats
- Troubleshooting, advising on new and old media, transfers, etc.
- Develop new tools, workflows, etc.



Forensics Lab Services

1 Added by Michael Olson, last edited by Michael Olson on Mar 20, 2012 (view change)

Service Level Agreement

Overview

The Digital Forensics Lab provides services to preserve and provide access to born digital materials. Born digital materials consist of a wide range materials stored offline on digital carriers of various forms and ages. These forms of media are often inaccessible through current computer technologies and operating systems requiring specialized equipment and software to preserve and provide scholarly access. The Forensics lab consists of a custom built capture station and forensic workstations with the ability to read a wide range of digital media such as floppy discs, CDs/DVDs, hard drives, computer tapes and most consumer types of flash memory.

Scope of this SLA

This SLA covers two scenarios:

1. **Patron Requests** - A SULAIR representative (from Special Collections, University Archives, a SULAIR branch library, etc.) requesting capture and analysis of physical computer media on behalf of a patron (a specific researcher or student)
2. **Internal Requests** - An internal SULAIR-generated request that supports general collection activities, where no specific patron is involved. Examples include Exhibition needs, Curator needs, staff needs, and faculty support.

In all cases, the person requesting the item(s) is referred to as the "Reporter," and is the mediator for the request.

Roles and Responsibilities

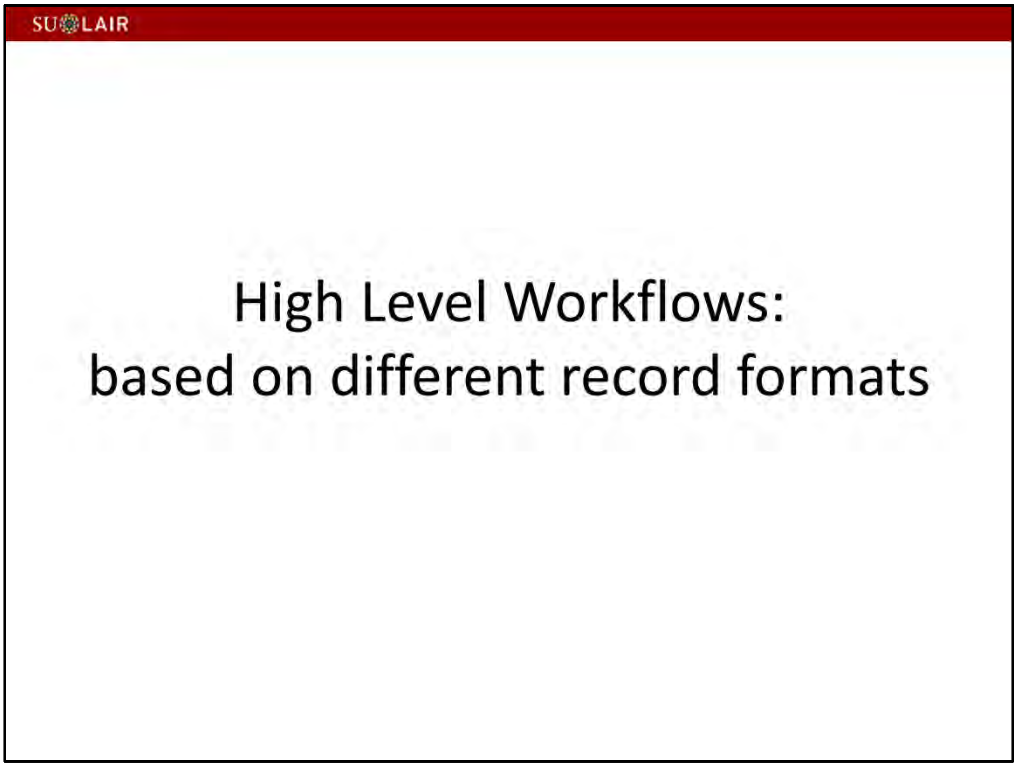
Patron:

- work with curator or public service staff to identify materials to be migrated and place request via their service structures
- provide adequate information for identifying requested item(s), or portions thereof
- provide information as to the intended use of resulting content
- advise as to whether the computer media is likely to contain restricted data defined as (see the following reference for further information)
  - Social Security Numbers
  - Credit Card Numbers
  - Financial Account Numbers
  - Driver's License Numbers
  - Health Insurance Policy ID Numbers
- advise as to whether logical or forensic recovery is appropriate. Forensic imaging of computer media may make it possible to recover deleted files. In many cases this recovery technique will not be attempted as it may conflict with the wishes of the computer media's creator
- receive files (as specified by timetable)

This slide illustrates an early DRAFT for our “service level agreement” – describing to our community what the DAP / forensic lab does.

But obviously we need to better understand and consider the following:
Understanding the community we serve
Describing the formats we capture from, the hardware, software and tools available
What type of training and/or workshops might be available
When we offer one-on-one assistance and what that encompasses

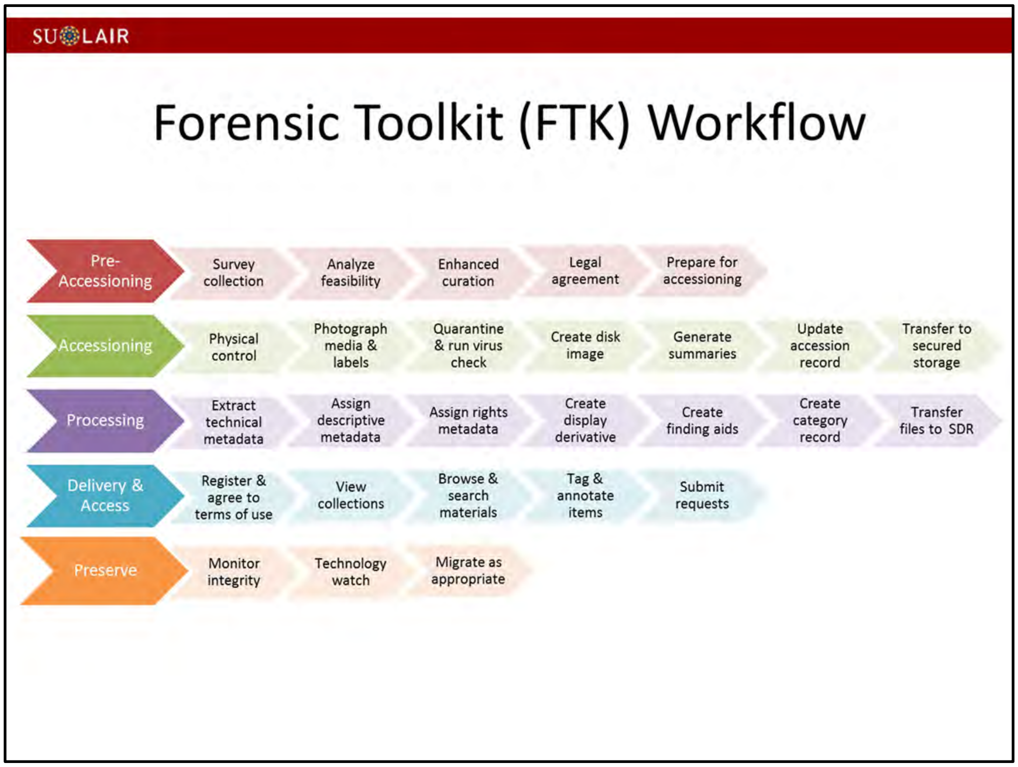
Questions: Will we attempt any cost recover? Will we capture data for personal use on campus?
What services can we offer with existing staff model?



SU LAIR

High Level Workflows: based on different record formats

Now – onto our current high-level workflows – of which we have several!



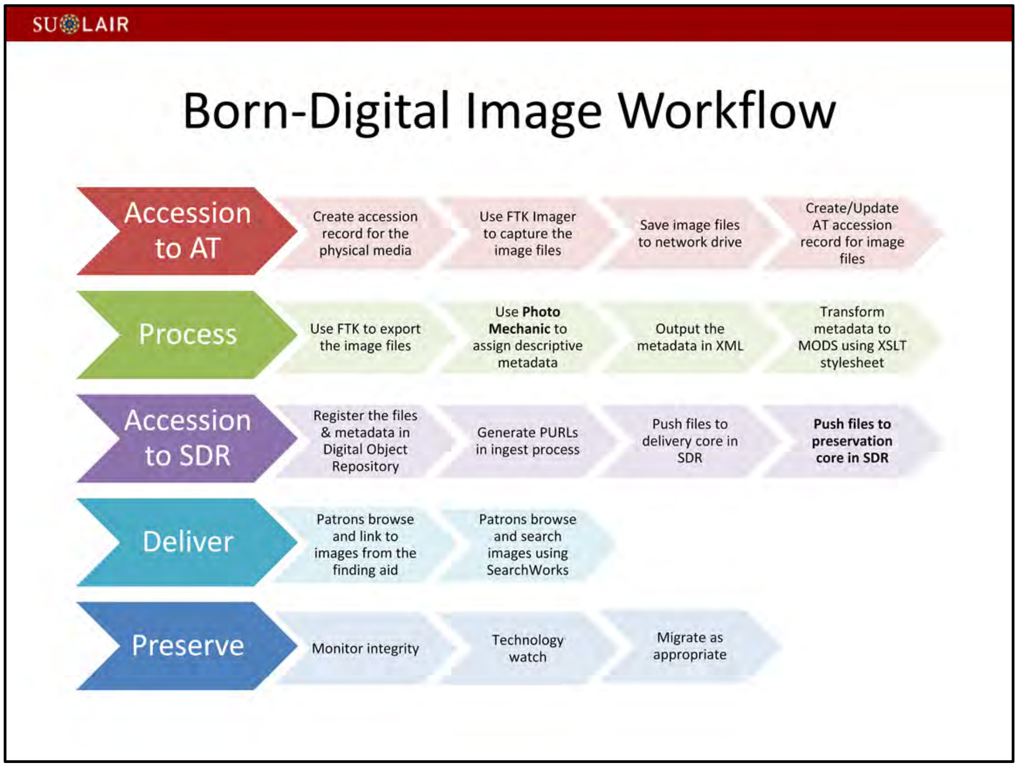
SU LAIR

Forensic Toolkit (FTK) Workflow

Pre-Accessioning	Survey collection	Analyze feasibility	Enhanced curation	Legal agreement	Prepare for accessioning		
Accessioning	Physical control	Photograph media & labels	Quarantine & run virus check	Create disk image	Generate summaries	Update accession record	Transfer to secured storage
Processing	Extract technical metadata	Assign descriptive metadata	Assign rights metadata	Create display derivative	Create finding aids	Create category record	Transfer files to SDR
Delivery & Access	Register & agree to terms of use	View collections	Browse & search materials	Tag & annotate items	Submit requests		
Preserve	Monitor integrity	Technology watch	Migrate as appropriate				

This is our basic FTK workflow for processing – primarily text- or mixed-collections. It was developed during our 2 years on the AIMS Project working with (admittedly) small born-digital materials from larger paper collections.

In our work since AIMS though we have found that FTK does not work best for ALL file formats – and other tools were found and workflows were developed.



Born-Digital Image Workflow

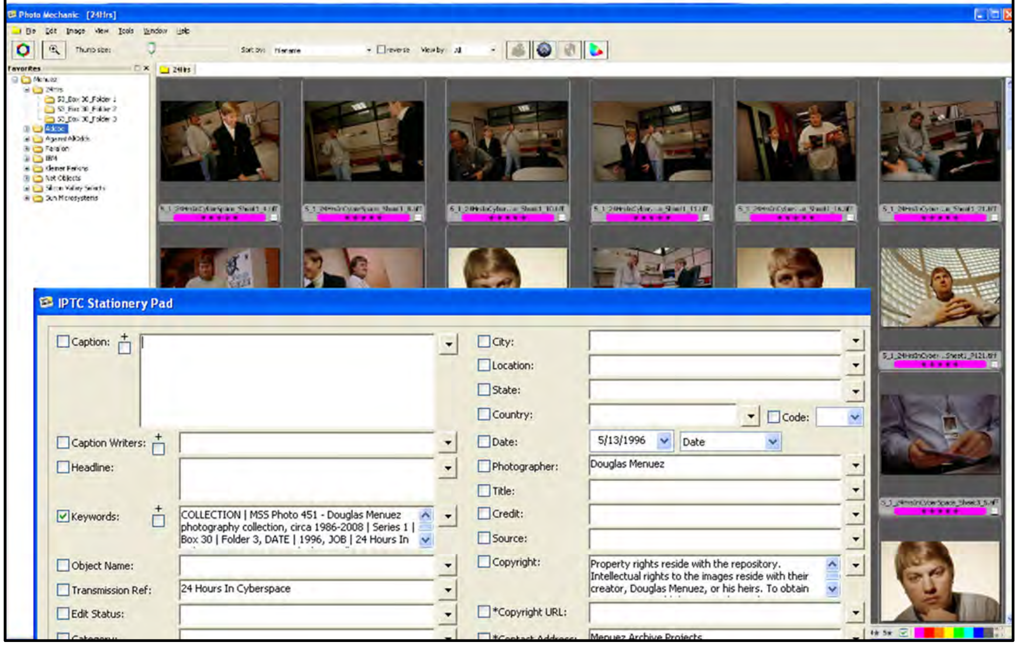
Accession to AT	Create accession record for the physical media	Use FTK Imager to capture the image files	Save image files to network drive	Create/Update AT accession record for image files
Process	Use FTK to export the image files	Use **Photo Mechanic** to assign descriptive metadata	Output the metadata in XML	Transform metadata to MODS using XSLT stylesheet
Accession to SDR	Register the files & metadata in Digital Object Repository	Generate PURLs in ingest process	Push files to delivery core in SDR	**Push files to preservation core in SDR**
Deliver	Patrons browse and link to images from the finding aid	Patrons browse and search images using SearchWorks		
Preserve	Monitor integrity	Technology watch	Migrate as appropriate	

For our digital photography collections (or large subsets) – when reformatting labs or the photographer has stored metadata in the IPTC header embedded in the digital object;

We found that that metadata was not read by FTK. So, while FTK is part of this process it is not the MAIN processing tool for this format. Instead, after testing, we selected PhotoMechanic software.

SU LAIR

PhotoMechanic



IPTC Stationery Pad

□ Caption:

□ Caption Writers:

□ Headline:

☑ Keywords: COLLECTION | MSS Photo 451 - Douglas Menuez photography collection, circa 1986-2008 | Series 1 | Box 30 | Folder 3, DATE | 1996, JOB | 24 Hours In

□ Object Name:

□ Transmission Ref: 24 Hours In Cyberspace

□ Edit Status:

□ City:

□ Location:

□ State:

□ Country: □ Code:

□ Date: 5/13/1996 Date

□ Photographer: Douglas Menuez

□ Title:

□ Credit:

□ Source:

□ Copyright: Property rights reside with the repository. Intellectual rights to the images reside with their creator, Douglas Menuez, or his heirs. To obtain

□ *Copyright URL:

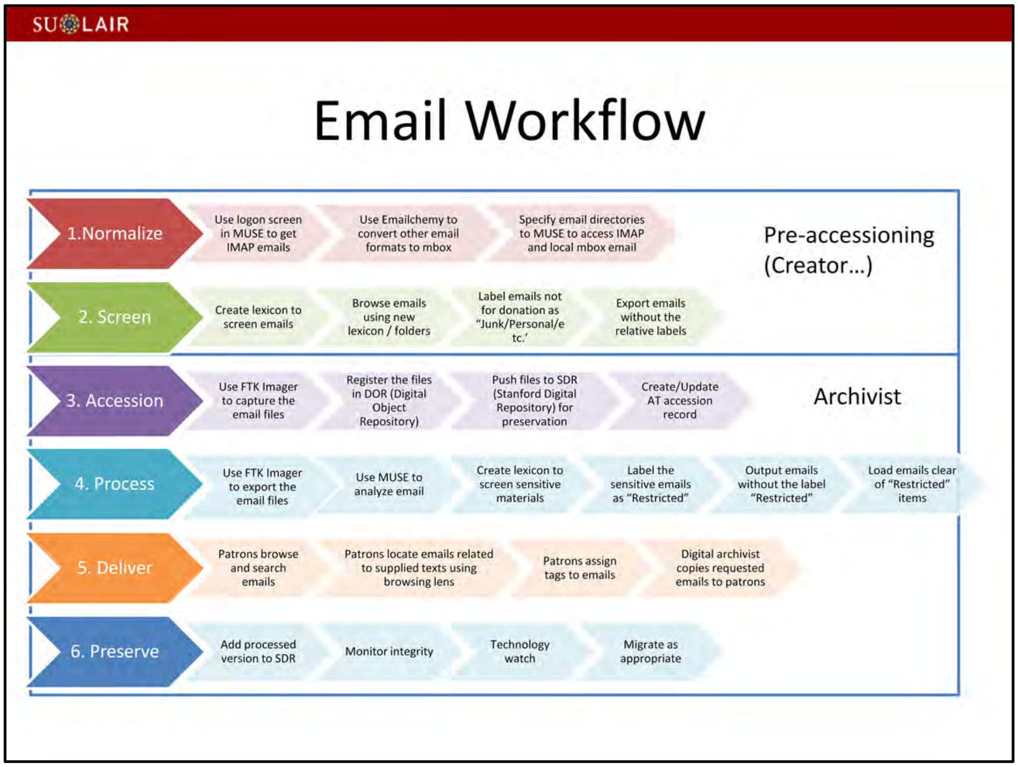
□ *Contact Address: Menuez Archive Projects

Here is an example of PhotoMechanic – these are screen shots from the Douglas Menuez collection.

The software allows us to create and use templates and to add bulk metadata across selected images.

The files and metadata are scripted out to MODS for ingestion into the SDR (Stanford Digital Repository).

This is the format that we are currently delivering online – because it is so similar to digitized photographs.



Email Workflow

Pre-accessioning (Creator…)

1.Normalize: Use logon screen in MUSE to get IMAP emails → Use Emailchemy to convert other email formats to mbox → Specify email directories to MUSE to access IMAP and local mbox email

2. Screen: Create lexicon to screen emails → Browse emails using new lexicon / folders → Label emails not for donation as "Junk/Personal/e tc.' → Export emails without the relative labels

Archivist

3. Accession: Use FTK Imager to capture the email files → Register the files in DOR (Digital Object Repository) → Push files to SDR (Stanford Digital Repository) for preservation → Create/Update AT accession record

4. Process: Use FTK Imager to export the email files → Use MUSE to analyze email → Create lexicon to screen sensitive materials → Label the sensitive emails as "Restricted" → Output emails without the label "Restricted" → Load emails clear of "Restricted" items

5. Deliver: Patrons browse and search emails → Patrons locate emails related to supplied texts using browsing lens → Patrons assign tags to emails → Digital archivist copies requested emails to patrons

6. Preserve: Add processed version to SDR → Monitor integrity → Technology watch → Migrate as appropriate

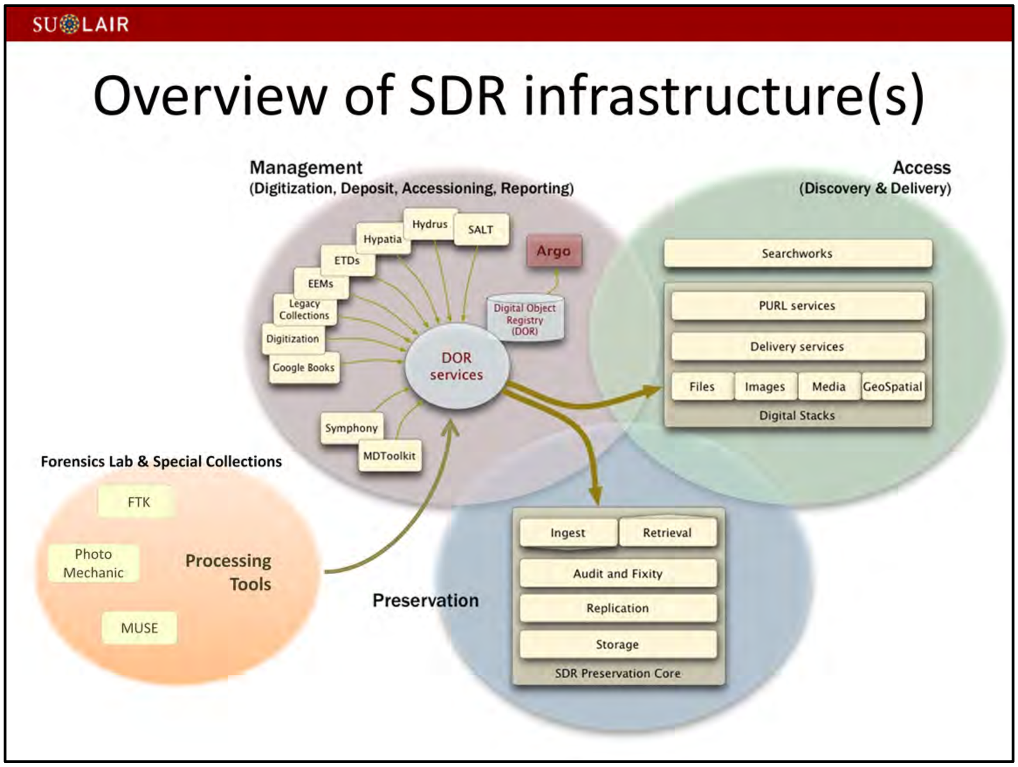
Another format that gave us difficulties in FTK were Email collections. If one email was flagged as restricted, all the emails in the "box" were flagged as restricted. There was no way to pull out one email or one thread.

Conveniently, Peter discovered a software in development at SU's MobiSocial Lab in Computer Sciences – called MUSE. It was developed by Sudheendra Hangal whom we have worked closely with this past 6 months or more. I will show you a few screen shots of MUSE later.

SULAIR

Where the Digital Archives Program meets the SDR

But first, I would like to talk a little about our digital repository. Output from our various processes needs to be ingested in the SDR for long-term preservation and for delivery.



SU LAIR

Overview of SDR infrastructure(s)

Management
(Digitization, Deposit, Accessioning, Reporting)

Hydrus
SALT
Hypatia
ETDs
EEMs
Legacy Collections
Digitization
Google Books
Argo
Digital Object Registry (DOR)
DOR services
Symphony
MDToolkit

Access
(Discovery & Delivery)

Searchworks
PURL services
Delivery services
Files
Images
Media
GeoSpatial
Digital Stacks

Forensics Lab & Special Collections

FTK
Photo Mechanic
MUSE
Processing Tools

Preservation

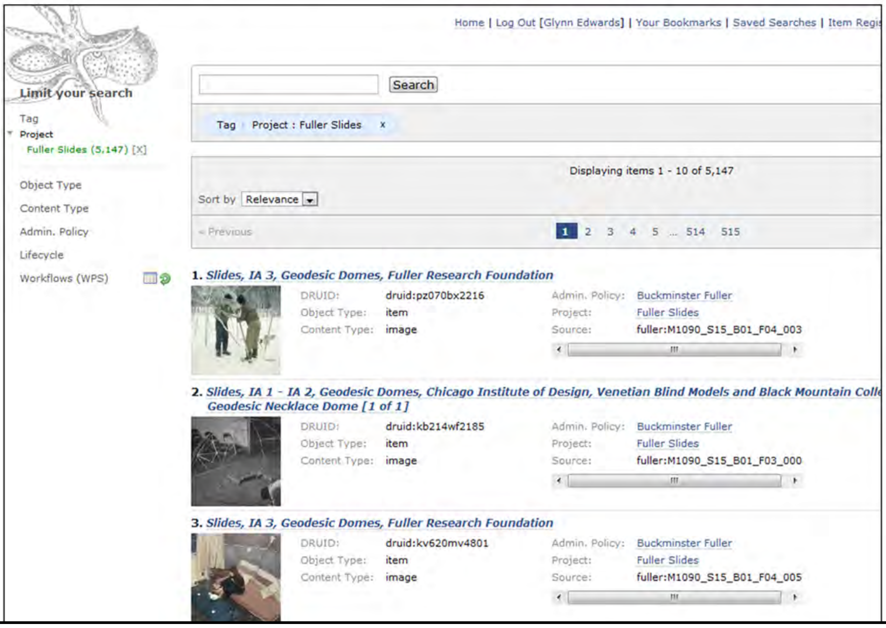
Ingest
Retrieval
Audit and Fixity
Replication
Storage
SDR Preservation Core

Here is a snapshot overview of the SDR triumvirate of Management, Access, and Preservation.
Delivery is via persistent URL's (PURLs) which resolve to "whatever" delivery environment is used for that "format"

Delivery for born-digital content in the short term is more complicated. You will note the orange oval – which sits outside of the SDR infrastructure. In it I have put the three core processing tools there as they are not yet included in the suite of tools that map data to DOR/SDR.

SULAIR

Digital Images in ARGO



Home | Log Out [Glynn Edwards] | Your Bookmarks | Saved Searches | Item Regi

Limit your search

Tag

Project

Fuller Slides (5,147) [X]

Object Type

Content Type

Admin. Policy

Lifecycle

Workflows (WPS)

Search

Tag | Project : Fuller Slides x

Displaying items 1 - 10 of 5,147

Sort by Relevance

« Previous 1 2 3 4 5 ... 514 515

1. *Slides, IA 3, Geodesic Domes, Fuller Research Foundation*

DRUID:	druid:pz070bx2216	Admin. Policy:	Buckminster Fuller
Object Type:	item	Project:	Fuller Slides
Content Type:	image	Source:	fuller:M1090_S15_B01_F04_003

2. *Slides, IA 1 - IA 2, Geodesic Domes, Chicago Institute of Design, Venetian Blind Models and Black Mountain Coll Geodesic Necklace Dome [1 of 1]*

DRUID:	druid:kb214wf2185	Admin. Policy:	Buckminster Fuller
Object Type:	item	Project:	Fuller Slides
Content Type:	image	Source:	fuller:M1090_S15_B01_F03_000

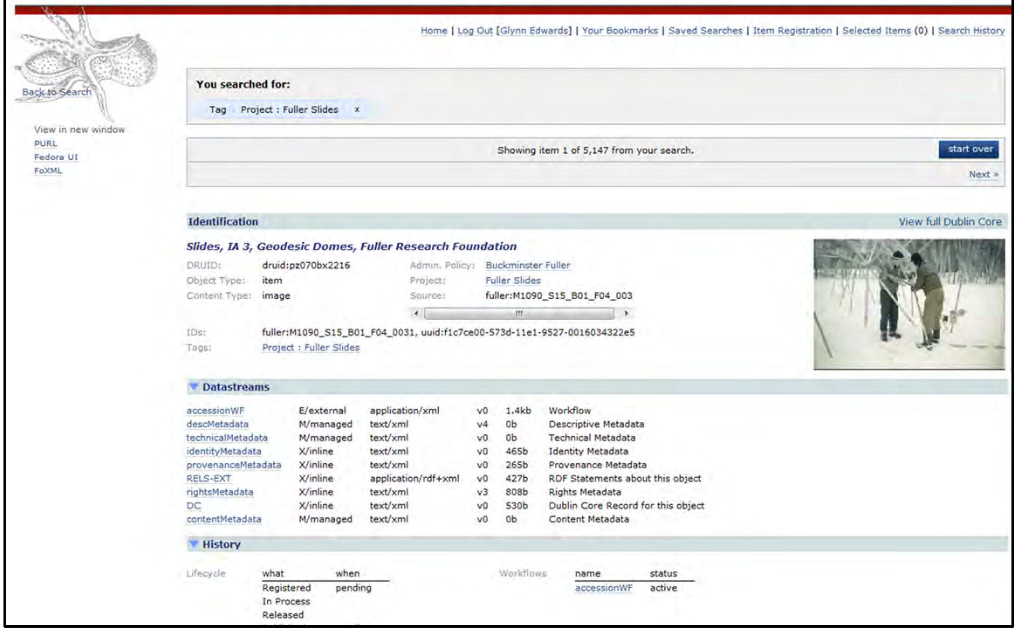
3. *Slides, IA 3, Geodesic Domes, Fuller Research Foundation*

DRUID:	druid:kv620mv4801	Admin. Policy:	Buckminster Fuller
Object Type:	item	Project:	Fuller Slides
Content Type:	image	Source:	fuller:M1090_S15_B01_F04_005

Digital photographs (PhotoMechanic data) is the only format (so far) that has been ingested into the SDR.

This is a screen shot of our Management / Administrative interface – ARGO – for the digital photos in Fuller.

Digital Images in ARGO



Home | Log Out [Glynn Edwards] | Your Bookmarks | Saved Searches | Item Registration | Selected Items (0) | Search History

Back to Search

View in new window
PURL
Fedora UI
FoXML

You searched for:

Tag Project : Fuller Slides x

Showing item 1 of 5,147 from your search.

start over

Next »

Identification View full Dublin Core

Slides, IA 3, Geodesic Domes, Fuller Research Foundation

DRUID:	druid:pz070bx2216	Admin. Policy:	Buckminster Fuller
Object Type:	item	Project:	Fuller Slides
Content Type:	image	Source:	fuller:M1090_S15_B01_F04_003

IDs: fuller:M1090_S15_B01_F04_0031, uuid:f1c7ce00-573d-11e1-9527-0016034322e5
Tags: Project : Fuller Slides

Datastreams

accessionWF	E/external	application/xml	v0	1.4kb	Workflow
descMetadata	M/managed	text/xml	v4	0b	Descriptive Metadata
technicalMetadata	M/managed	text/xml	v0	0b	Technical Metadata
identityMetadata	X/inline	text/xml	v0	465b	Identity Metadata
provenanceMetadata	X/inline	text/xml	v0	265b	Provenance Metadata
RELS-EXT	X/inline	application/rdf+xml	v0	427b	RDF Statements about this object
rightsMetadata	X/inline	text/xml	v3	808b	Rights Metadata
DC	X/inline	text/xml	v0	530b	Dublin Core Record for this object
contentMetadata	M/managed	text/xml	v0	0b	Content Metadata

History

Lifecycle

what	when
Registered	pending
In Process	
Released	

Workflows

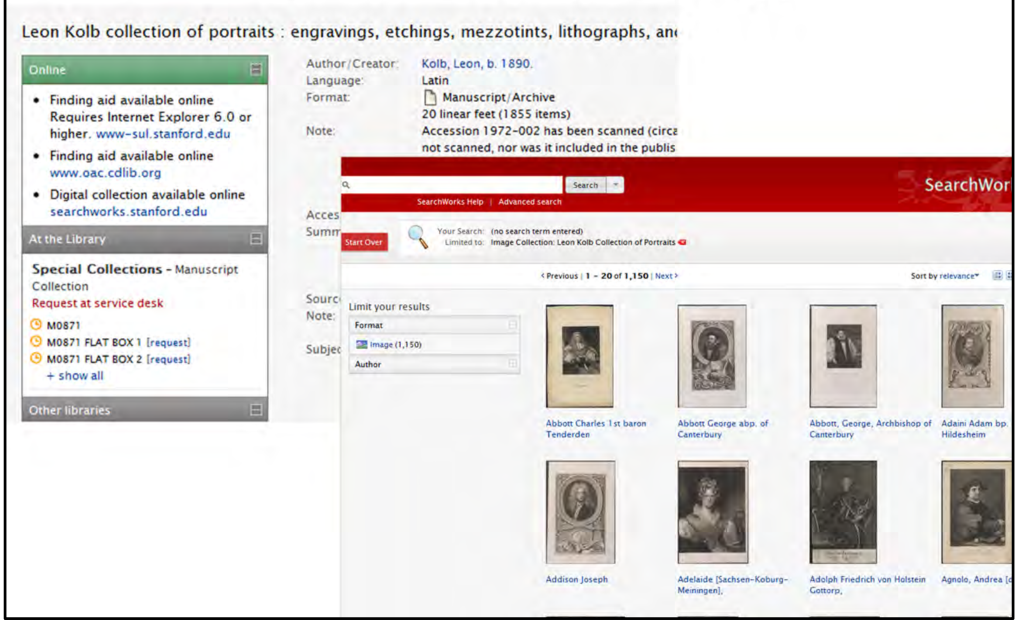
name	status
accessionWF	active

If we click on one image, you can see the Datastreams that are tied to this object; the unique IDs including a “local” one we use as a mnemonic for the physical location of the original.

ARGO object page also gives us a link to the PURL landing page for the object.

SU LAIR

Digital Image (delivery)



Leon Kolb collection of portraits : engravings, etchings, mezzotints, lithographs, an

Online

- Finding aid available online Requires Internet Explorer 6.0 or higher. www-sul.stanford.edu
- Finding aid available online www.oac.cdlib.org
- Digital collection available online searchworks.stanford.edu

At the Library

Special Collections - Manuscript Collection
Request at service desk
M0871
M0871 FLAT BOX 1 [request]
M0871 FLAT BOX 2 [request]
+ show all

Other libraries

Author/Creator: Kolb, Leon, b. 1890.
Language: Latin
Format: Manuscript/Archive
20 linear feet (1855 items)
Note: Accession 1972-002 has been scanned (circa
not scanned, nor was it included in the publis

Acces
Summ

Sourc
Note:

Subjec

Your Search: (no search term entered)
Limited to: Image Collection: Leon Kolb Collection of Portraits

‹ Previous | 1 - 20 of 1,150 | Next ›

Sort by relevance

Limit your results

Format

Image (1,150)

Author

Abbott Charles 1st baron Tenderden

Abbott George abp. of Canterbury

Abbott, George, Archbishop of Canterbury

Adaini Adam bp. Hildesheim

Addison Joseph

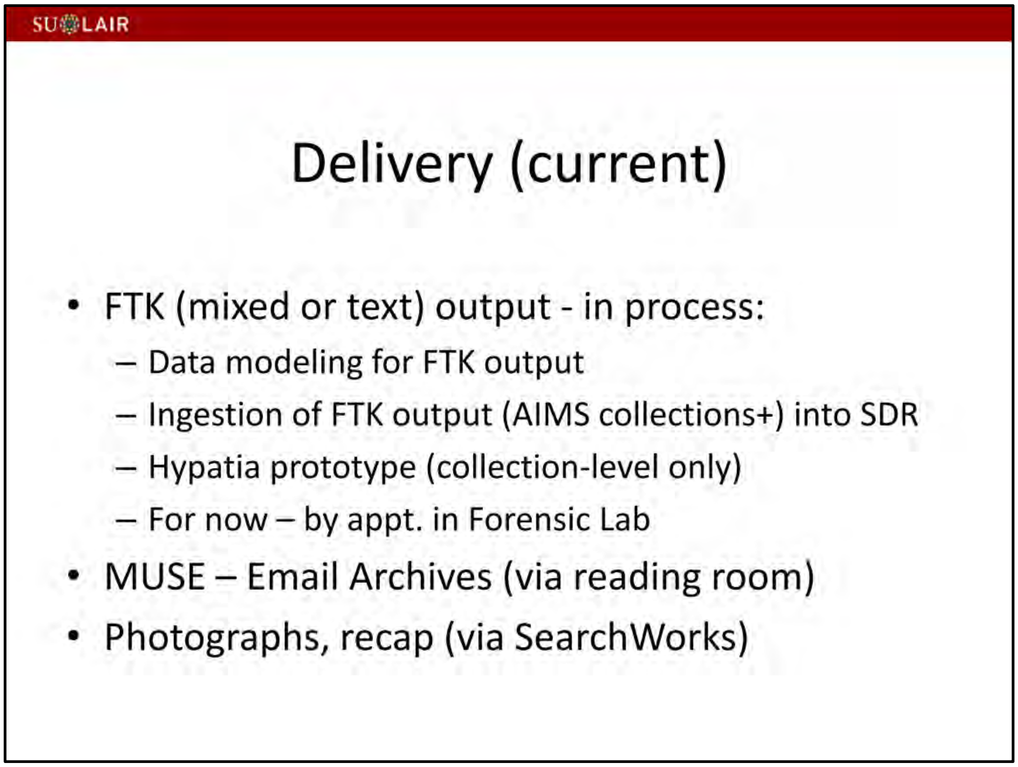
Adelaide [Sachsen-Koburg-Meiningen],

Adolph Friedrich von Holstein Gottorp,

Agnolo, Andrea [c

Discovery and access to the digital photographs is via our online catalog (SearchWorks).

The collection-level record includes a link to the set of digital objects; but the objects themselves come up as well if you search for that collection.



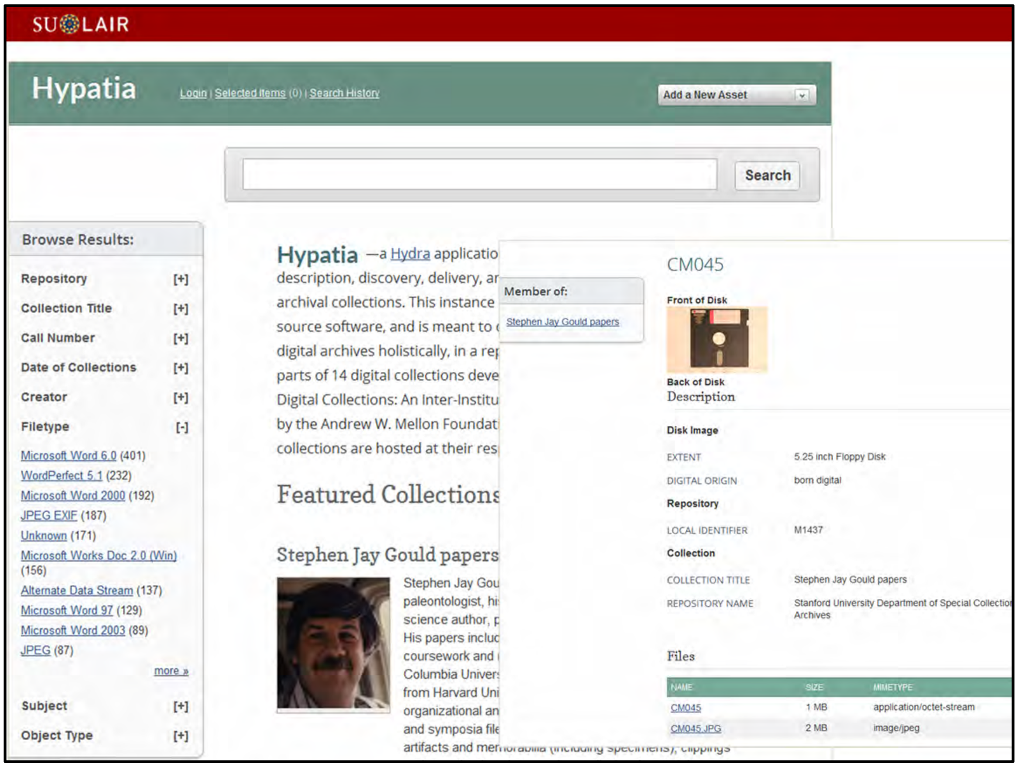
SU LAIR

Delivery (current)

- FTK (mixed or text) output - in process:
  - Data modeling for FTK output
  - Ingestion of FTK output (AIMS collections+) into SDR
  - Hypatia prototype (collection-level only)
  - For now – by appt. in Forensic Lab
- MUSE – Email Archives (via reading room)
- Photographs, recap (via SearchWorks)

While DLSS (Digital Libraries Systems & Services Dept. in SUL) is working on creating a data model for ingesting FTK output – slated to be completed this summer …

What this means currently is that access to files processed on the AIMS Project can only be delivered via FTK by appointment with our resident Digital Archivist – Peter Chan.



Hypatia —a Hydra applicatio
description, discovery, delivery, ar
archival collections. This instance
source software, and is meant to
digital archives holistically, in a rep
parts of 14 digital collections deve
Digital Collections: An Inter-Institu
by the Andrew W. Mellon Foundat
collections are hosted at their res

Featured Collections

Stephen Jay Gould papers

Stephen Jay Gou
paleontologist, hi
science author, p
His papers incluc
coursework and
Columbia Univer
from Harvard Uni
organizational an
and symposia file
artifacts and memorabilia (including specimens), clippings

CM045

Description

Disk Image

EXTENT	5.25 inch Floppy Disk
DIGITAL ORIGIN	born digital

Repository

LOCAL IDENTIFIER	M1437

Collection

COLLECTION TITLE	Stephen Jay Gould papers
REPOSITORY NAME	Stanford University Department of Special Collectio Archives

Files

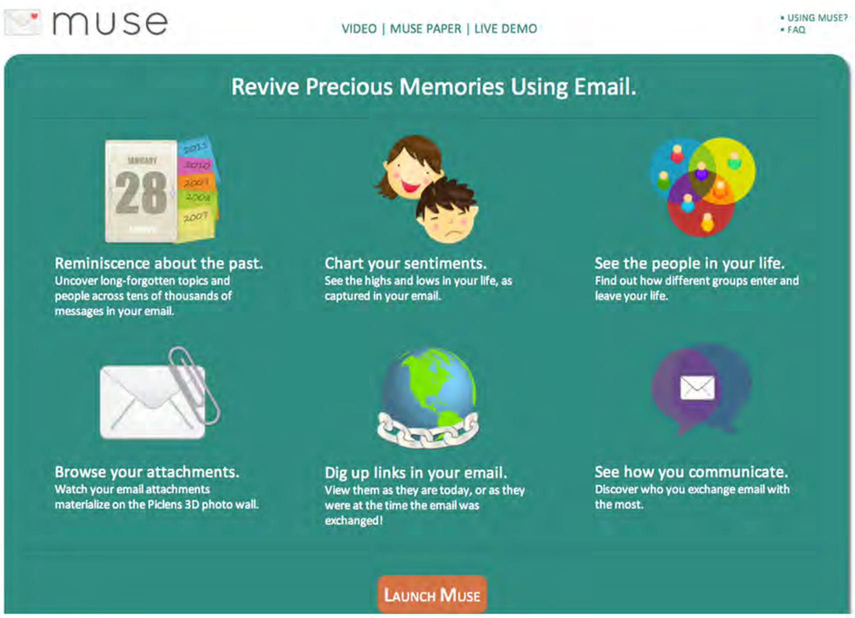
NAME	SIZE	MIMETYPE
CM045	1 MB	application/octet-stream
CM045.JPG	2 MB	image/jpeg

This screenshot is for a prototype – called Hypatia – developed at the tail end of the AIMS Project last fall.

Collection-level records were ingested along with a set of photographs of the physical media – but not the files themselves.

SU LAIR

Email: appraisal, processing, discovery & delivery



muse

VIDEO | MUSE PAPER | LIVE DEMO

- USING MUSE?
- FAQ

Revive Precious Memories Using Email.

Reminiscence about the past.
Uncover long-forgotten topics and people across tens of thousands of messages in your email.

Chart your sentiments.
See the highs and lows in your life, as captured in your email.

See the people in your life.
Find out how different groups enter and leave your life.

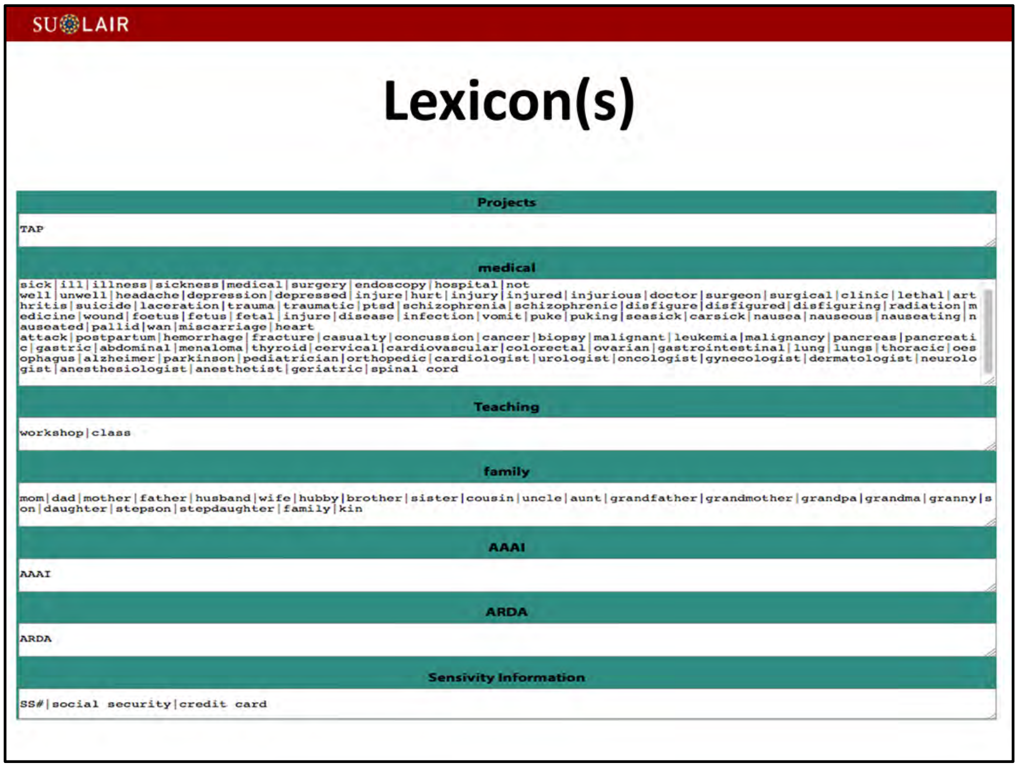
Browse your attachments.
Watch your email attachments materialize on the Piclens 3D photo wall.

Dig up links in your email.
View them as they are today, or as they were at the time the email was exchanged!

See how you communicate.
Discover who you exchange email with the most.

I'd like to switch now to MUSE again – and show you a few quick screen shots.

Besides being used and tested for Processing Email Archives, MUSE will be tested as an appraisal tool, a pre-accessioning tool, and for delivery via the reading room in Special Collections.



Lexicon(s)

Projects

TAP

medical

```
sick|ill|illness|sickness|medical|surgery|endoscopy|hospital|not
well|unwell|headache|depression|depressed|injure|hurt|injury|injured|injurious|doctor|surgeon|surgical|clinic|lethal|art
hritis|suicide|laceration|trauma|traumatic|ptsd|schizophrenia|schizophrenic|disfigure|disfigured|disfiguring|radiation|m
edicine|wound|foetus|fetus|fetal|injure|disease|infection|vomit|puke|puking|seasick|carsick|nausea|nauseous|nauseating|n
auseated|pallid|wan|miscarriage|heart
attack|postpartum|hemorrhage|fracture|casualty|concussion|cancer|biopsy|malignant|leukemia|malignancy|pancreas|pancreati
c|gastric|abdominal|menaloma|thyroid|cervical|cardiovascular|colorectal|ovarian|gastrointestinal|lung|lungs|thoracic|oes
ophagus|alzheimer|parkinson|pediatrician|orthopedic|cardiologist|urologist|oncologist|gynecologist|dermatologist|neurolo
gist|anesthesiologist|anesthetist|geriatric|spinal cord
```

Teaching

workshop|class

family

```
mom|dad|mother|father|husband|wife|hubby|brother|sister|cousin|uncle|aunt|grandfather|grandmother|grandpa|grandma|granny|s
on|daughter|stepson|stepdaughter|family|kin
```

AAAI

AAAI

ARDA

ARDA

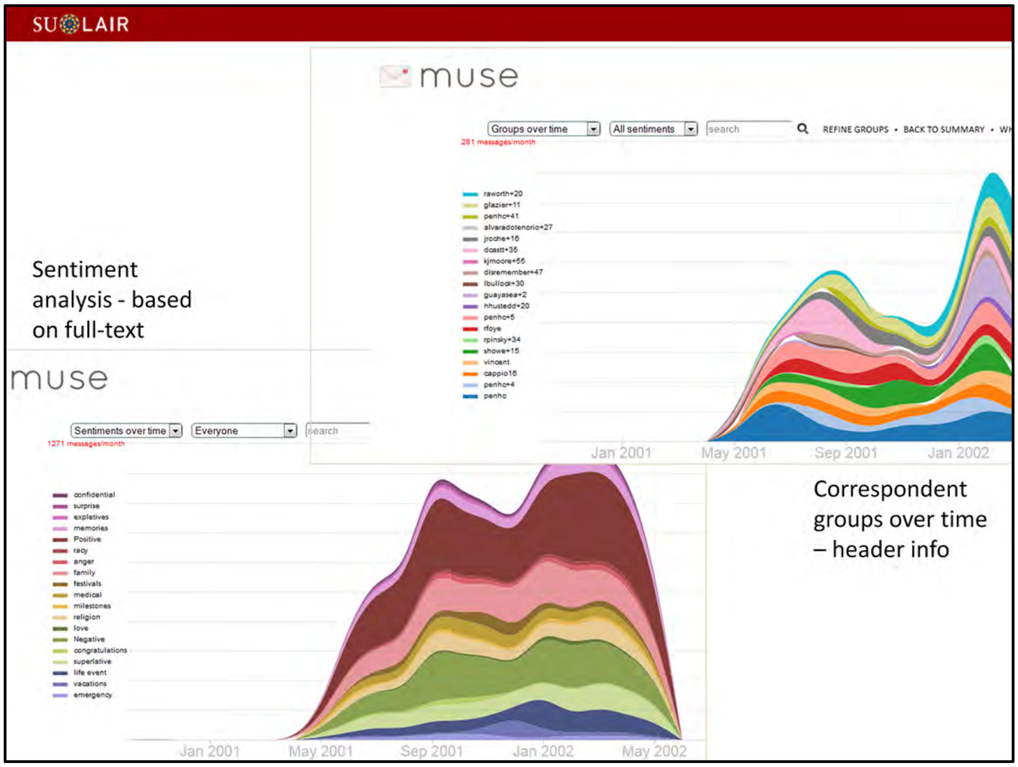
Sensivity Information

SS#|social security|credit card

One of the underlying structures for MUSE is the built-in Lexicon – which is using natural language processing to group “sentiments”

In working with Sudheendra, we requested the ability to build our own Lexicons or to edit the existing one. That capability was added and MUSE will now accept multiple lexicons.

This is one Peter created for one of our collections



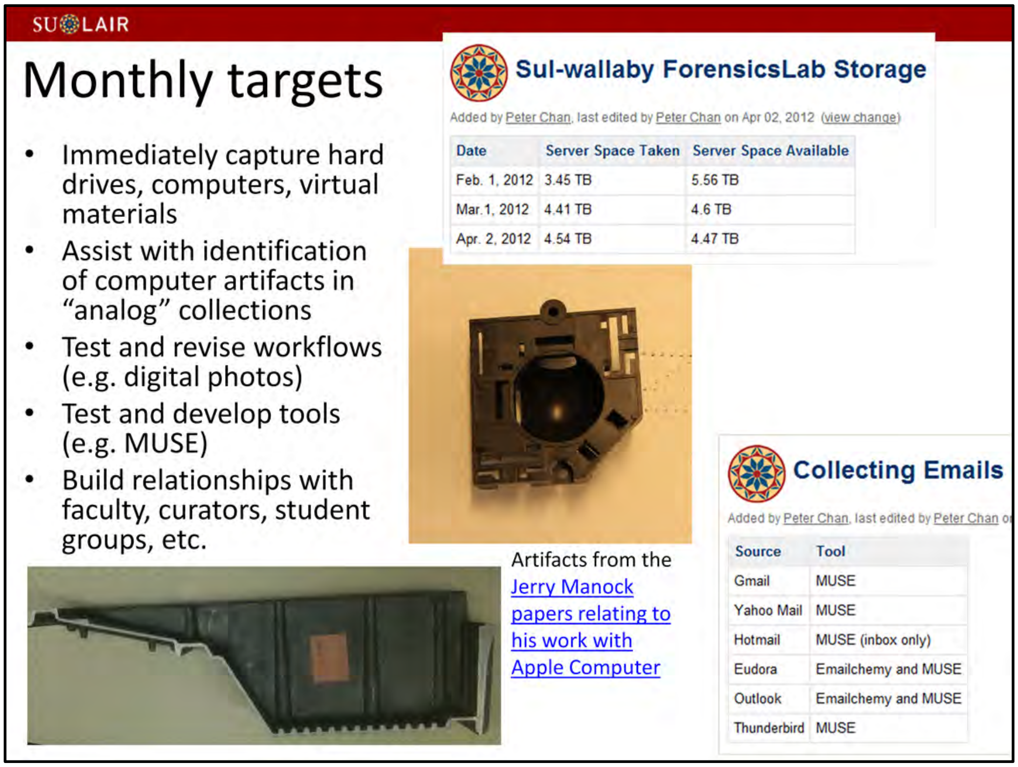
Sentiment analysis - based on full-text

Correspondent groups over time – header info

This final view of MUSE – shows two visualizations based on the contents/data.

The first is the Sentiment analysis based on full-text analysis of similarly grouped terms – for family, emotions, etc.
If you click on any of the colors (here presented over time) you will bring up the actual emails associated with that “emotion” and can scroll through them.

The second is Correpondents over time – which pulls data from the header information.



SU LAIR

Monthly targets

- Immediately capture hard drives, computers, virtual materials
- Assist with identification of computer artifacts in "analog" collections
- Test and revise workflows (e.g. digital photos)
- Test and develop tools (e.g. MUSE)
- Build relationships with faculty, curators, student groups, etc.

Sul-wallaby ForensicsLab Storage

Added by Peter Chan, last edited by Peter Chan on Apr 02, 2012 (view change)

Date	Server Space Taken	Server Space Available
Feb. 1, 2012	3.45 TB	5.56 TB
Mar.1, 2012	4.41 TB	4.6 TB
Apr. 2, 2012	4.54 TB	4.47 TB

Artifacts from the Jerry Manock papers relating to his work with Apple Computer

Collecting Emails

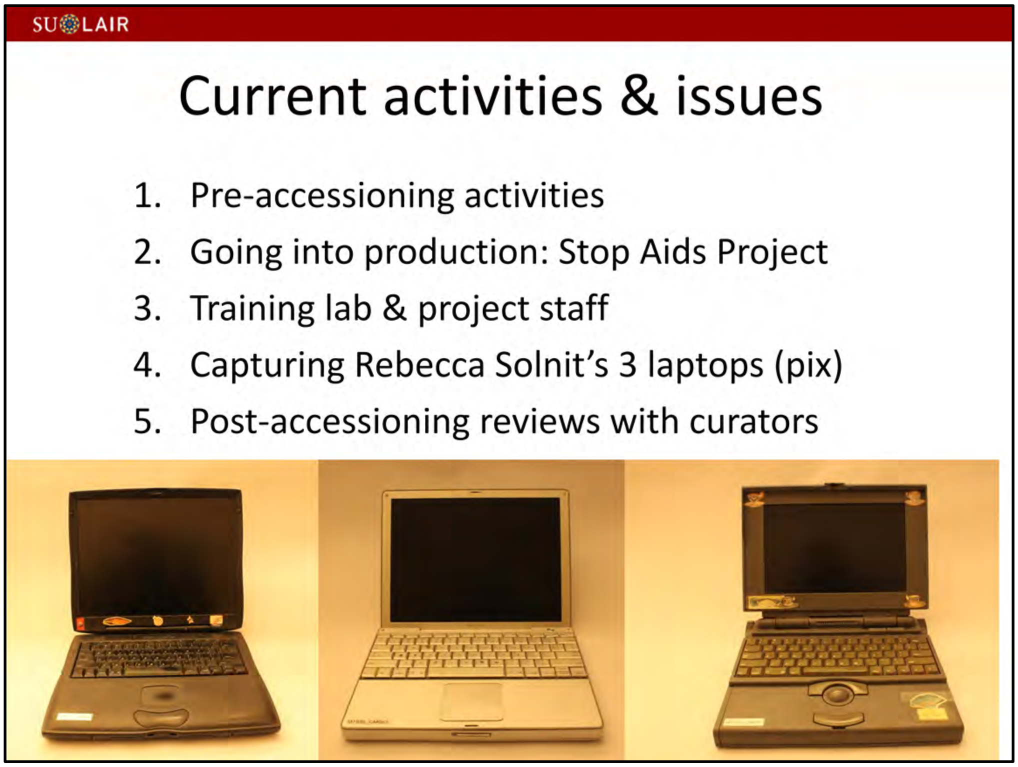
Added by Peter Chan, last edited by Peter Chan o

Source	Tool
Gmail	MUSE
Yahoo Mail	MUSE
Hotmail	MUSE (inbox only)
Eudora	Emailchemy and MUSE
Outlook	Emailchemy and MUSE
Thunderbird	MUSE

After AIMS was over, and with Peter now base-funded, we began setting ATTAINABLE monthly targets – or goals as well as longer term projects and priorities.

Some examples of monthly targets are:
Capturing data from hard drives, computers and virtual transfers
Assist with identification of computer artifacts – as in the Jerry Manock papers

Longer term goals are:
Test and revise workflows
Continue to seek out and test new tools, etc.
Build relationships with faculty, curators, student groups, etc.

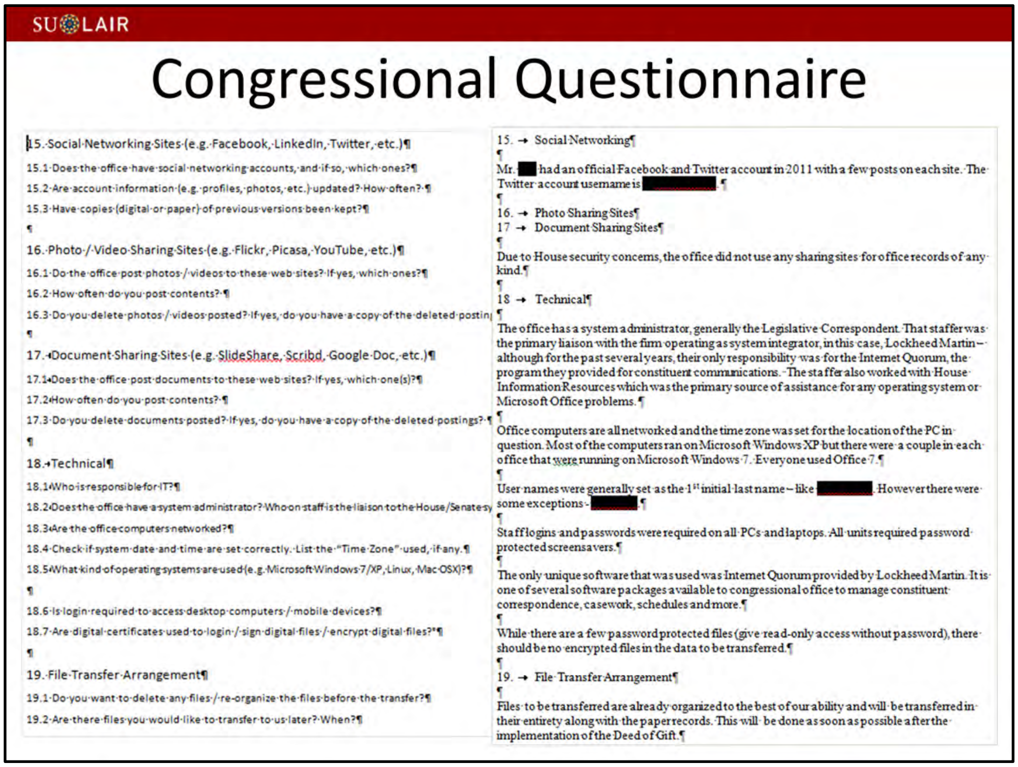


Current activities & issues

1. Pre-accessioning activities
2. Going into production: Stop Aids Project
3. Training lab & project staff
4. Capturing Rebecca Solnit's 3 laptops (pix)
5. Post-accessioning reviews with curators

Our work often gets sidetracked by issues or problems encountered while trying to capture hard drives. Here are three laptops from the Rebecca Solnit collection – one, an old MAC laptop, requires a cable/connector we did not receive nor have.

Peter has to be very creative in trying to resolve these issues – often going far from Stanford for help.



Congressional Questionnaire

15. Social Networking Sites (e.g. Facebook, LinkedIn, Twitter, etc.)

15.1 Does the office have social networking accounts, and if so, which ones?

15.2 Are account information (e.g. profiles, photos, etc.) updated? How often?

15.3 Have copies (digital or paper) of previous versions been kept?

16. Photo / Video Sharing Sites (e.g. Flickr, Picasa, YouTube, etc.)

16.1 Do the office post photos / videos to these web sites? If yes, which ones?

16.2 How often do you post contents?

16.3 Do you delete photos / videos posted? If yes, do you have a copy of the deleted postin

17. Document Sharing Sites (e.g. SlideShare, Scribd, Google Doc, etc.)

17.1 Does the office post documents to these web sites? If yes, which one(s)?

17.2 How often do you post contents?

17.3 Do you delete documents posted? If yes, do you have a copy of the deleted postings?

18. Technical

18.1 Who is responsible for IT?

18.2 Does the office have a system administrator? Who on staff is the liaison to the House/Senate sy

18.3 Are the office computers networked?

18.4 Check if system date and time are set correctly. List the "Time Zone" used, if any.

18.5 What kind of operating systems are used (e.g. Microsoft Windows 7/XP, Linux, Mac OSX)?

18.6 Is login required to access desktop computers / mobile devices?

18.7 Are digital certificates used to login / sign digital files / encrypt digital files?"

19. File Transfer Arrangement

19.1 Do you want to delete any files / re-organize the files before the transfer?

19.2 Are there files you would like to transfer to us later? When?

15. → Social Networking

Mr. [redacted] had an official Facebook and Twitter account in 2011 with a few posts on each site. The Twitter account username is [redacted].

16. → Photo Sharing Sites
17 → Document Sharing Sites

Due to House security concerns, the office did not use any sharing sites for office records of any kind.

18 → Technical

The office has a system administrator, generally the Legislative Correspondent. That staffer was the primary liaison with the firm operating as system integrator, in this case, Lockheed Martin – although for the past several years, their only responsibility was for the Internet Quorum, the program they provided for constituent communications. The staffer also worked with House Information Resources which was the primary source of assistance for any operating system or Microsoft Office problems.

Office computers are all networked and the time zone was set for the location of the PC in question. Most of the computers ran on Microsoft Windows XP but there were a couple in each office that were running on Microsoft Windows 7. Everyone used Office 7.

User names were generally set as the 1st initial last name – like [redacted]. However there were some exceptions - [redacted].

Staff logins and passwords were required on all PCs and laptops. All units required password protected screensavers.

The only unique software that was used was Internet Quorum provided by Lockheed Martin. It is one of several software packages available to congressional office to manage constituent correspondence, casework, schedules and more.

While there are a few password protected files (give read-only access without password), there should be no encrypted files in the data to be transferred.

19. → File Transfer Arrangement

Files to be transferred are already organized to the best of our ability and will be transferred in their entirety along with the paper records. This will be done as soon as possible after the implementation of the Deed of Gift.

Another project came about when one of our curators was taking in a congressional collection.

We created a survey / questionnaire for staffers to fill in to assist with understanding the scope, files, extent, etc. of the digital material. I posted it on the congressional papers roundtable listserv and rec'd interest but most did not hold out much hope of getting responses.

In this example – we rec'd a very detailed 5 page response – which has been a wonderful resource for us as we proceed with his digital component.

Going into production: Stop Aids Project

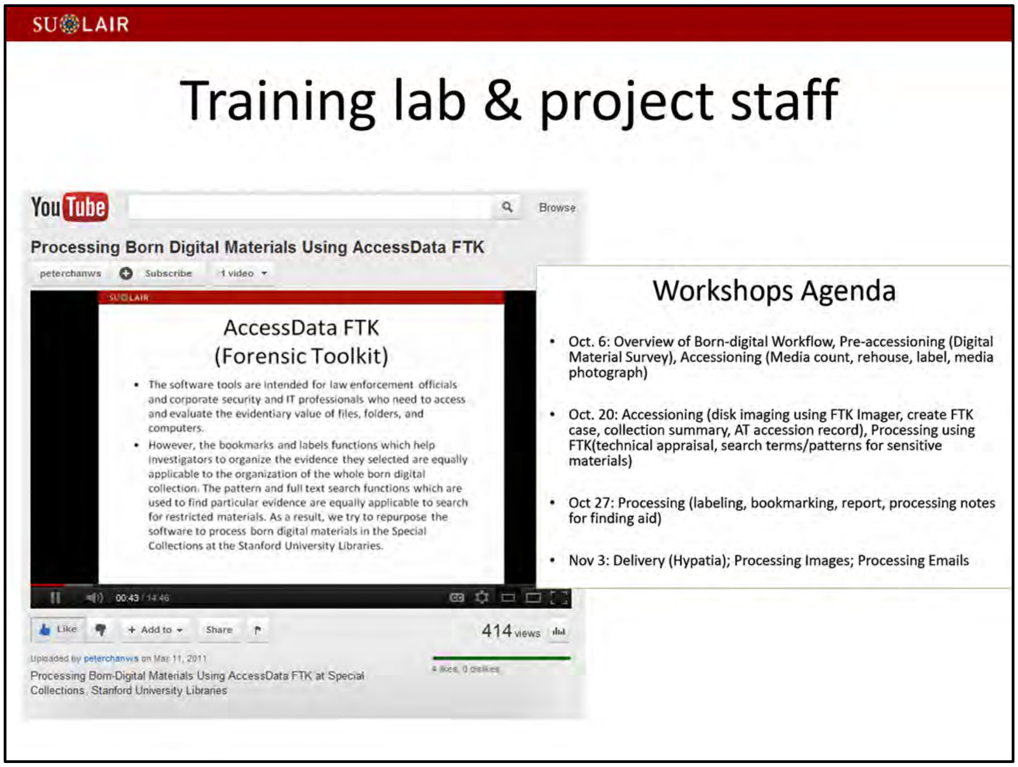
After AIMS we rec'd a grant to process the Stop AIDS Project records and digital content.

SAP was the production "test" of the workflows we'd developed earlier.

At this point the data has been captured and awaits processing.

Some interesting stats though:
Data loss from CDs was very high! These were back-ups primarily but we were only able to capture about 4% of hundreds. 3." floppies had a success rate of 60% and zip disks 96% capture.

Dataset is nearly 500K files and over 850 GB – much larger than the "AIMS" collections.



Training lab & project staff

Processing Born Digital Materials Using AccessData FTK

AccessData FTK (Forensic Toolkit)

- The software tools are intended for law enforcement officials and corporate security and IT professionals who need to access and evaluate the evidentiary value of files, folders, and computers.
- However, the bookmarks and labels functions which help investigators to organize the evidence they selected are equally applicable to the organization of the whole born digital collection. The pattern and full text search functions which are used to find particular evidence are equally applicable to search for restricted materials. As a result, we try to repurpose the software to process born digital materials in the Special Collections at the Stanford University Libraries.

Uploaded by peterchanws on Mar 11, 2011

Processing Born-Digital Materials Using AccessData FTK at Special Collections, Stanford University Libraries

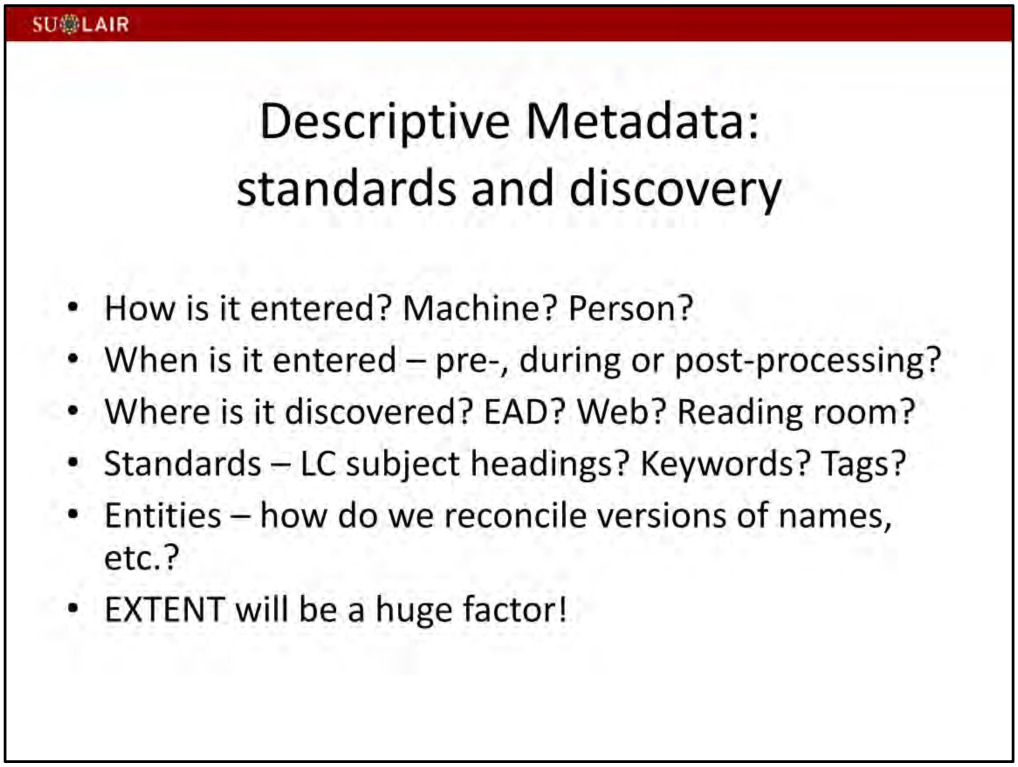
Workshops Agenda

- Oct. 6: Overview of Born-digital Workflow, Pre-accessioning (Digital Material Survey), Accessioning (Media count, rehouse, label, media photograph)
- Oct. 20: Accessioning (disk imaging using FTK Imager, create FTK case, collection summary, AT accession record), Processing using FTK(technical appraisal, search terms/patterns for sensitive materials)
- Oct 27: Processing (labeling, bookmarking, report, processing notes for finding aid)
- Nov 3: Delivery (Hypatia); Processing Images; Processing Emails

As we needed to process the digital content as part of the grant, Peter had to quickly pull together training sessions for staff.

He developed FOUR workshops – attended by SAP staff and other Special Collections and Preserving Virtual Worlds staff

Our goal is to consolidate a lot of this material on our website over the summer.



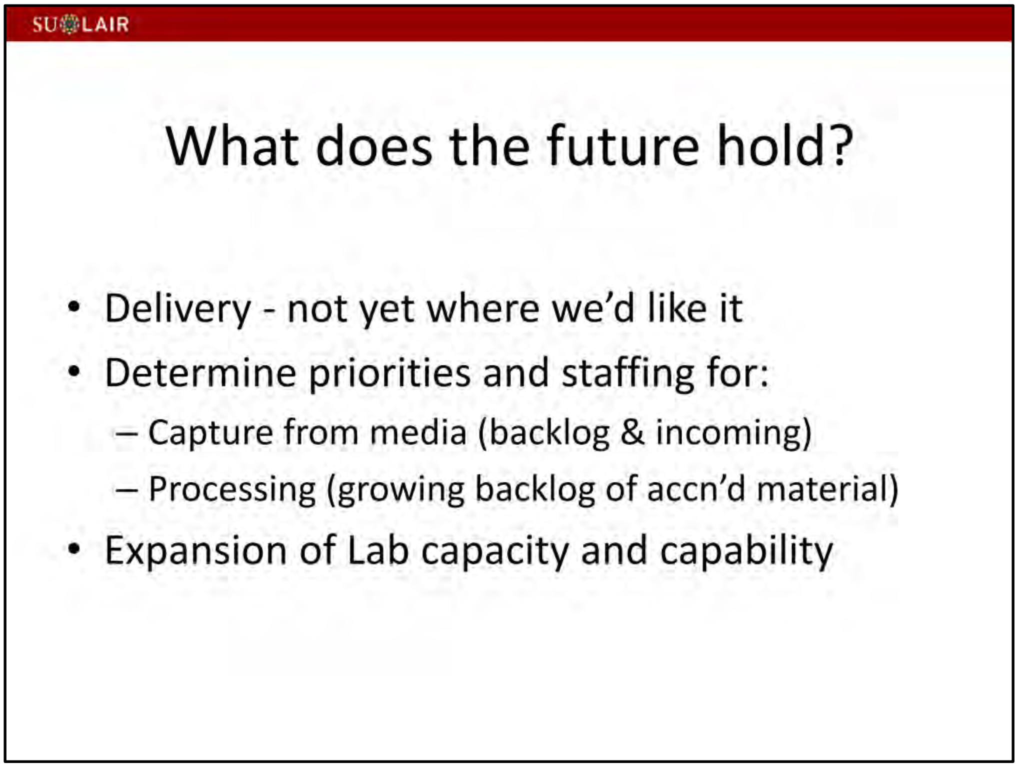
SULAIR

Descriptive Metadata: standards and discovery

- How is it entered? Machine? Person?
- When is it entered – pre-, during or post-processing?
- Where is it discovered? EAD? Web? Reading room?
- Standards – LC subject headings? Keywords? Tags?
- Entities – how do we reconcile versions of names, etc.?
- EXTENT will be a huge factor!

The processing of the SAP records this summer will test our initial workflows greatly.

The size of the digital collection far surpases the AIMS collections in complexity and number of files.



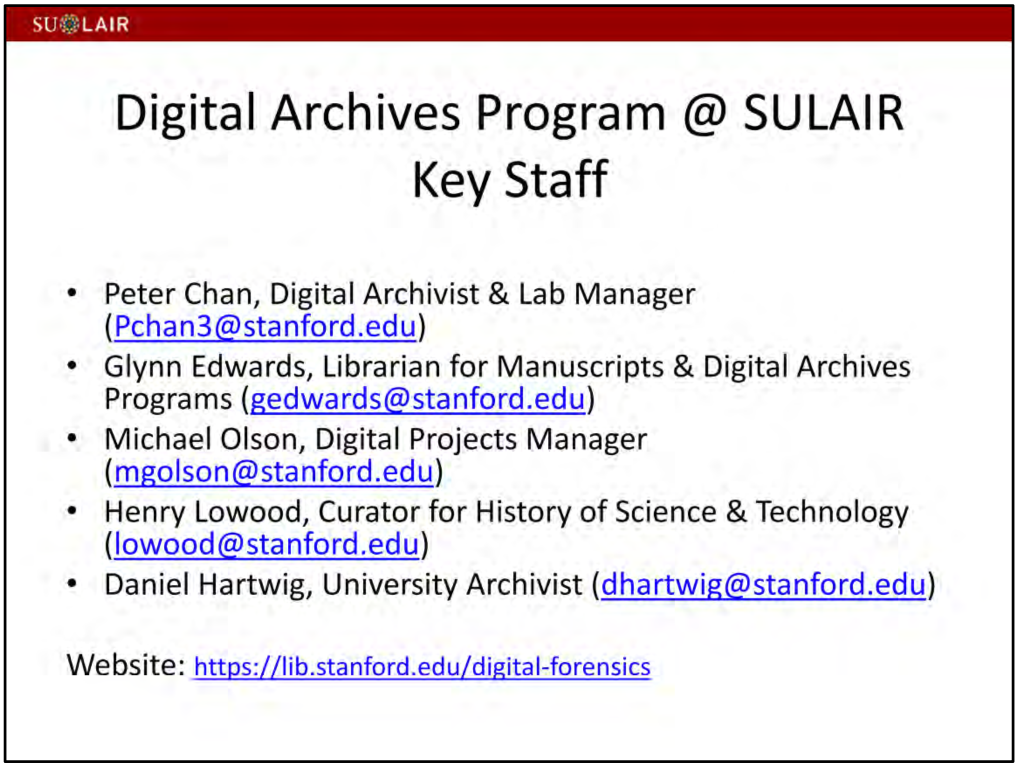
What does the future hold?

- Delivery - not yet where we'd like it
- Determine priorities and staffing for:
  - Capture from media (backlog & incoming)
  - Processing (growing backlog of accn'd material)
- Expansion of Lab capacity and capability

The ultimate goal of delivery would be to discover across collections and formats. That said we are far from searching across digitized and born-digital metadata and full-text at this point.

We also need to determine how we will: capture data from our growing backlog of physical media; process the data (staff and workflows)

It all hinges on how our lab is staffed and our capability to expand capacity and tools as needed.



Digital Archives Program @ SULAIR Key Staff

- Peter Chan, Digital Archivist & Lab Manager (Pchan3@stanford.edu)
- Glynn Edwards, Librarian for Manuscripts & Digital Archives Programs (gedwards@stanford.edu)
- Michael Olson, Digital Projects Manager (mgolson@stanford.edu)
- Henry Lowood, Curator for History of Science & Technology (lowood@stanford.edu)
- Daniel Hartwig, University Archivist (dhartwig@stanford.edu)

Website: https://lib.stanford.edu/digital-forensics

The team involved on the AIMS Project form the core of the DAP – Peter, Michael, me
To that we have added the expertise of Henry Lowood and Daniel Hartwig.