Digital Preservation, Digital Archivist, ePADD, Wikidata

Personal files in transition: from the private to the public, from analog to digital, Autonomous City of Buenos Aires, April 12, 2019

Peter Chan, Digital Archivist, Stanford Libraries

Peter Chan

- Stanford Libraries, 2008 - present.
- Born-digital / Forensic Lab
  - Migrate files from storage media to managed storage
- Project member in AIMS project (2010-2011)
  - Develop methodology for stewarding born-digital archival materials.
  - National Digital Stewardship Alliance (NDSA) Innovation Award 2012 (USA)
- Project manager for ePADD project (2013-2018)
  - Software that support the appraisal, processing, discovery, and delivery of email archives.
  - National Digital Stewardship Alliance (NDSA) Innovation Award 2017 (USA)
  - Digital Preservation Coalition (DPC) Software Sustainability Institute Award for Research and Innovation 2018 (UK)

Peter Chan

- Visiting digital archivist at the Royal Library of Copenhagen, Denmark 2015
  - Archiving of emails
  - Processing and delivery of born-digital materials using forensic software
  - Preservation of video games
- Visiting digital archivist at the Computerspielemuseum, Germany 2017
  - Developing controlled vocabulary for video game
  - Publishing controlled vocabulary as Linked Open Data
- Master in Library and Information Science (2009)
  - 2 hrs lecture on digital preservation by Robin Dale (program officer at OCLC focused on trusted digital repositories in 2009; associate librarian for library service at the Library of Congress at present)

Before digital archivist

- Bachelor of Financial Administration
- Master in Business Administration
- Deloitte Haskins & Sells as management consultant
- Standard Chartered Bank as cost accountant
- Internet Systems Ltd. as banking system consultant
- Chinese University of Hong Kong as lecturer
- Bank of America Asia as VP in Operations Planning
- MyiPhoto.com - Co-founder
- AsiaPay - Operations Manager

Audience

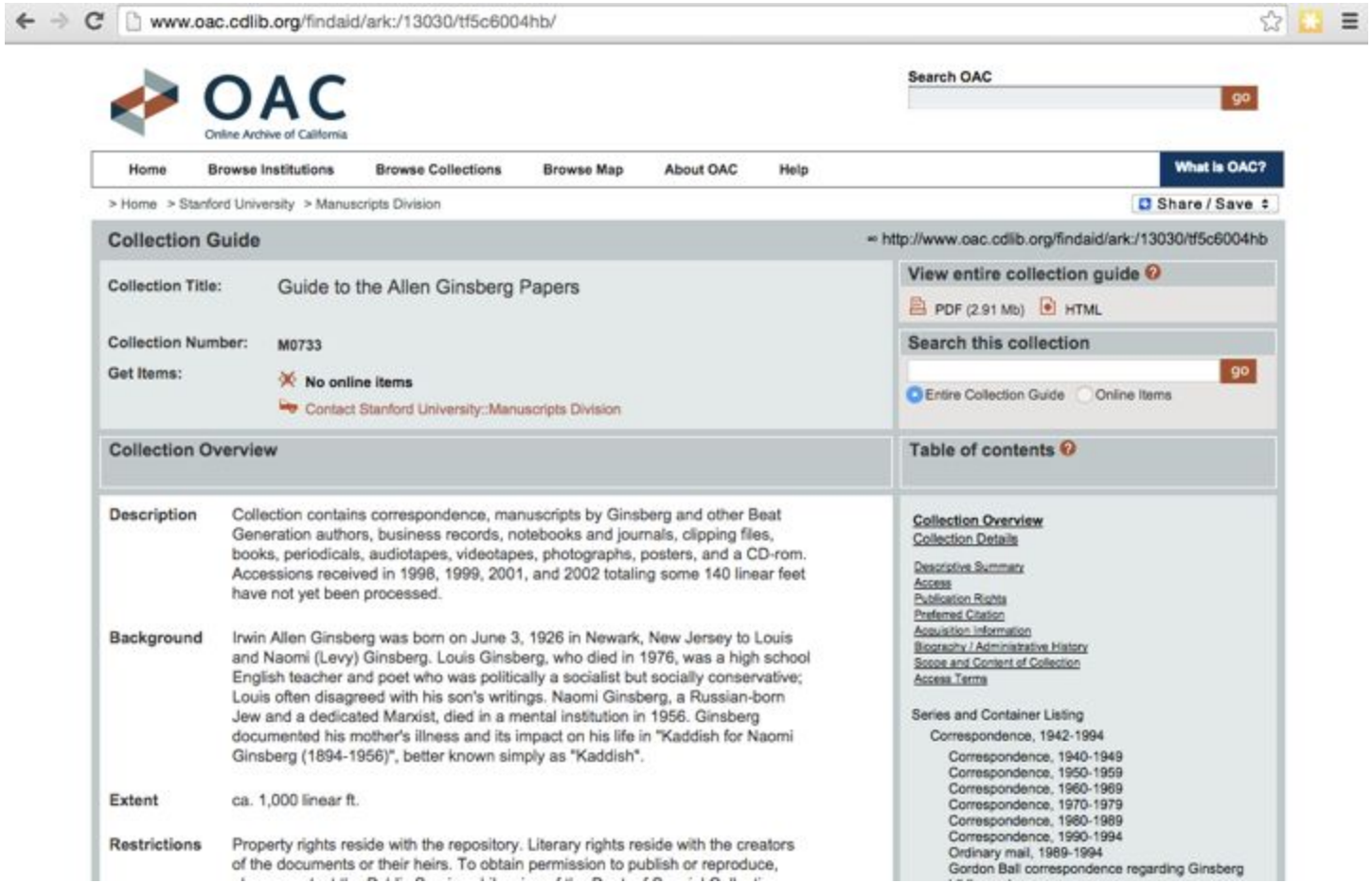
- Archivist
- Librarian
- Digital humanity scholar
- IT (Information Technology) specialist
- Administrator
- Others

Stanford Special Collections

….committed to acquiring, preserving, and providing access to primary source materials that support the research needs of the Stanford community and beyond….

Few of Our Notable Collections

- R. Buckminster Fuller Collection (an architect)
- John McCarthy papers (pioneer in AI and a faculty member)
- Allen Ginsberg papers (a poet)
- STOP AIDS Project Records (organization)
- Stephen M. Cabrinety Collection in the History of Microcomputing (computer software)


www.oac.cdlib.org/findaid/ark:/13030/tf5c6004hb/
OAC
Online Archive of California
Search OAC
go
Home
Browse Institutions
Browse Collections
Browse Map
About OAC
Help
What is OAC?
> Home > Stanford University > Manuscripts Division
Share / Save
Collection Guide
http://www.oac.cdlib.org/findaid/ark:/13030/tf5c6004hb
Collection Title: Guide to the Allen Ginsberg Papers
Collection Number: M0733
Get Items: No online items
Contact Stanford University::Manuscripts Division
View entire collection guide
PDF (2.91 Mb)
HTML
Search this collection
go
Entire Collection Guide
Online Items
Collection Overview
Table of contents
Description Collection contains correspondence, manuscripts by Ginsberg and other Beat Generation authors, business records, notebooks and journals, clipping files, books, periodicals, audiotapes, videotapes, photographs, posters, and a CD-rom. Accessions received in 1998, 1999, 2001, and 2002 totaling some 140 linear feet have not yet been processed.
Background Irwin Allen Ginsberg was born on June 3, 1926 in Newark, New Jersey to Louis and Naomi (Levy) Ginsberg. Louis Ginsberg, who died in 1976, was a high school English teacher and poet who was politically a socialist but socially conservative; Louis often disagreed with his son's writings. Naomi Ginsberg, a Russian-born Jew and a dedicated Marxist, died in a mental institution in 1956. Ginsberg documented his mother's illness and its impact on his life in "Kaddish for Naomi Ginsberg (1894-1956)", better known simply as "Kaddish".
Extent ca. 1,000 linear ft.
Restrictions Property rights reside with the repository. Literary rights reside with the creators of the documents or their heirs. To obtain permission to publish or reproduce,
Collection Overview
Collection Details
Descriptive Summary
Access
Publication Rights
Preferred Citation
Acquisition Information
Biography / Administrative History
Scope and Content of Collection
Access Terms
Series and Container Listing
Correspondence, 1942-1994
Correspondence, 1940-1949
Correspondence, 1950-1959
Correspondence, 1960-1969
Correspondence, 1970-1979
Correspondence, 1960-1989
Correspondence, 1990-1994
Ordinary mail, 1989-1994
Gordon Ball correspondence regarding Ginsberg

Born-digital Materials

- Born-digital information is distinguished from digitized, the latter describing a document created on paper that has been scanned (and possibly transformed into character data using OCR). A document created using a word processor may be described as born digital.

https://www2.archivists.org/glossary/terms/b/born-digital

Hard drives in Gitai (Amos) film archive

Punch cards in Stephen J. Gould papers

Open reel tapes in Engelbart (Douglas C.) papers

Storage media in Stephen M. Cabrinety Collection in the History of Microcomputing, ca. 1975-1995

Agrippa:
A Book of the Dead

Published in 1992

https://en.wikipedia.org/wiki/Agrippa_(A_Book_of_the_Dead)

The project manifested as a poem written by Gibson incorporated into an artist's book created by Ashbaugh; as such it was as much a work of collaborative conceptual art as poetry. Gibson stated that Ashbaugh's design "eventually included a supposedly self-devouring floppy-disk intended to display the text only once, then eat itself." Ashbaugh was gleeful at the dilemma this would pose to librarians: in order to register the copyright of the book, he had to send two copies to the United States Library of Congress, who, in order to classify it had to read it, and in the process, necessarily had to destroy it. The creators had initially intended to infect the disks with a computer virus, but declined to after considering the potential damage to the computer systems of innocents.

https://en.wikipedia.org/wiki/Agrippa_(A_Book_of_the_Dead)

Image from ZDNet

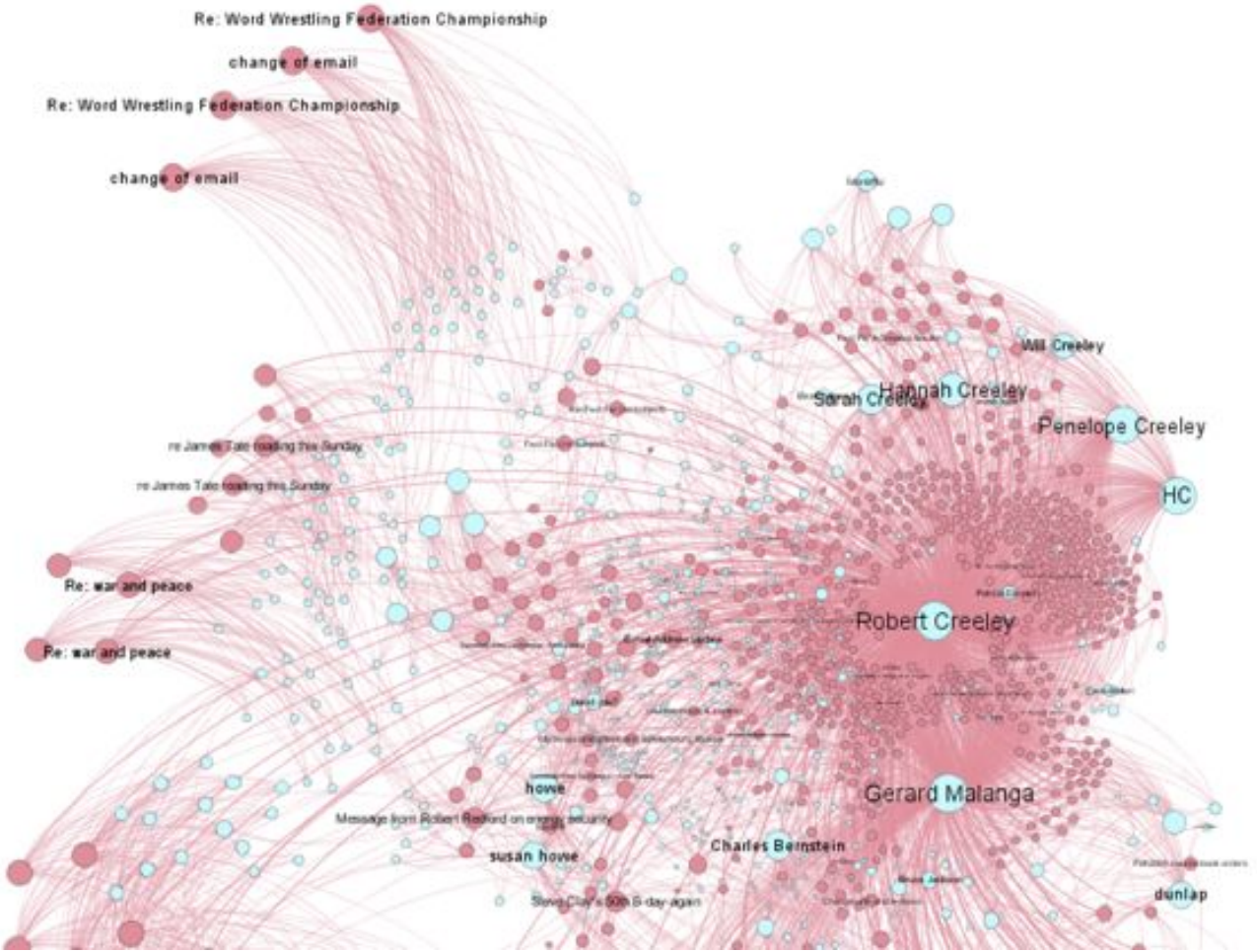
Dealing with born-digital materials provide both challenges and opportunities

Challenges

- Obsolete computer file storage media
- Processing - read each file and assign series and subseries?
- Digital preservation (print out emails)
- Staff training (both managers and archivists)
- Changes in donor agreement (exclusive or not)
- Potential destruction of collection materials and infection of computer virus

Opportunities

- Rethink arrangement and description
- Additional analysis such as social network analysis
- Delivery in the web
- Discovery not limited by author, title and subject headings
- Full text search on collection materials


Re: Word Wrestling Federation Championship
change of email
Re: Word Wrestling Federation Championship
change of email
Will Creeley
Hannah Creeley
Sarah Creeley
Penelope Creeley
HC
Re: war and peace
Re: war and peace
Robert Creeley
howe
Gerard Malanga
susan howe
Charles Bernstein
dunlap

In addition to books written by Robert Creeley, the search for “Robert Creeley” in Google show his spouse, awards he received, people related to him.

Robert Creeley

American poet

Robert White Creeley was an American poet and author of more than sixty books. He is usually associated with the Black Mountain poets, though his verse aesthetic diverged from that school's. He was close with Charles Olson, Robert Duncan, Allen Ginsberg, John Wieners and Ed Dorn. Wikipedia

Born: May 21, 1926, Arlington, MA

Died: March 30, 2005, Odessa, TX

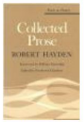
Awards: Bollingen Prize, American Book Award, MORE

Spouse: Penelope Highton (m. 1977–2005), Bobbie Louise Hawkins (m. 1960–1976)

Movies: Poetry in Motion

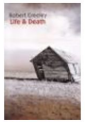
Books

View 45+ more



Collected prose
1984

Selected Poems
1976



Life & death
1993

If I were writing this
2003

For love
1962

People also search for

View 15+ more

Charles Olson

Robert Duncan

Denise Levertov

Gregory Corso

Philip Whalen

Units Dealing with Born-digital Materials

- Special Collections
- Digital Library System & Services
  - Born-digital / Forensic Lab
  - Stanford Media Preservation Lab
  - Stanford Digital Repository (SDR)

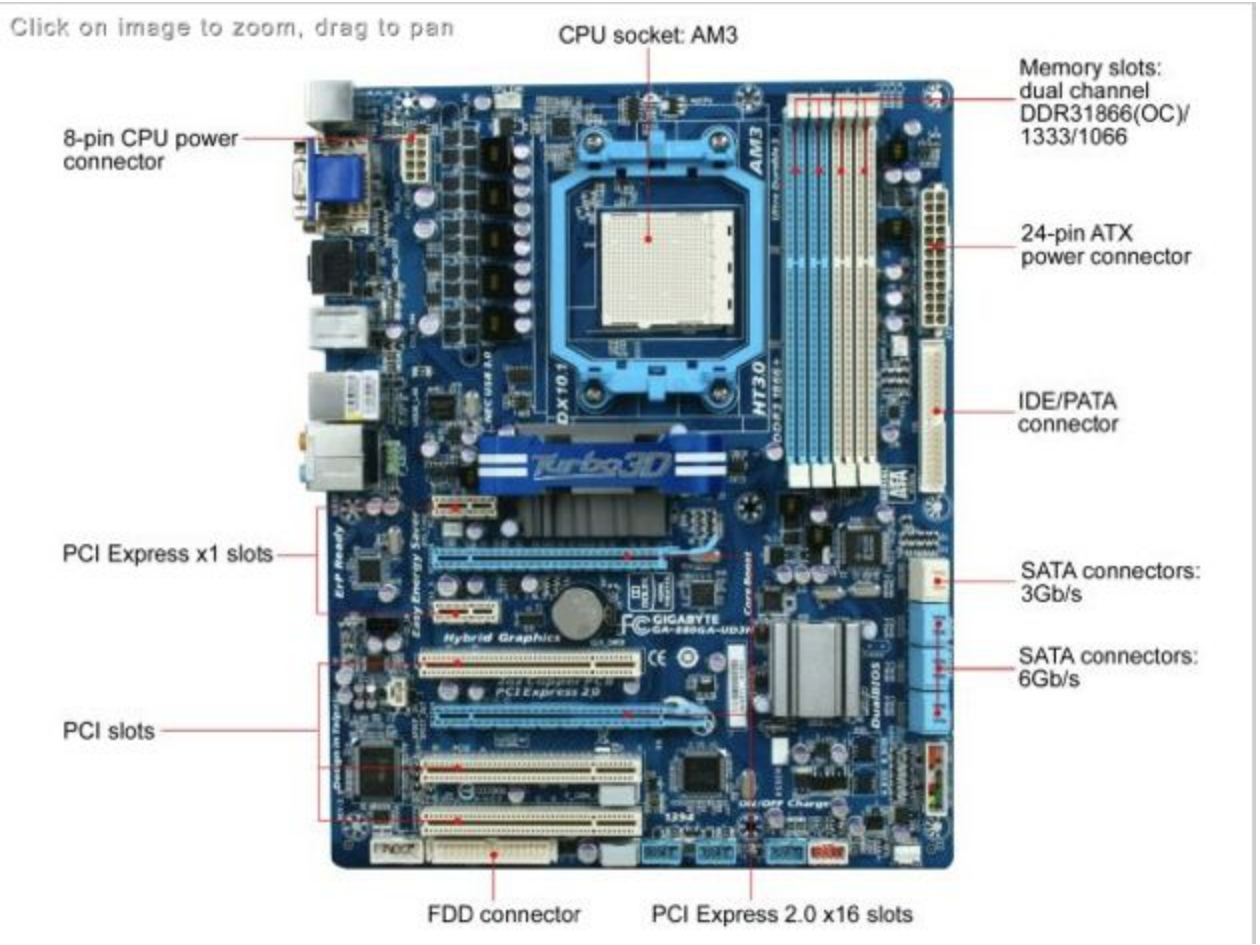
Born-digital / Forensic Lab

- Can do:
  - Iomega Zip disks (100 MB and 250 MB)
  - Floppy disc: 8-inch, 5¼-inch ("Minifloppy"), 3½-inch ("Microfloppy")
  - Multimedia Card: Compact Flash Card (CFC), Memory Stick Card (MSC), Smart Media Card (SMC), MicroDrive (MD), xD Card (xD), Memory Stick Pro (MSP), Memory Stick Pro Duo (MSPD), Secure Digital Card (SDC, SDHC, and SDXC), MicroSD - MultiMedia Card (MMC)
  - Hard Disk Drives with the following interfaces (IDE, SATA, SCSI, FireWire 800/400, USB 3.0/2.0/1.1)
  - Optical Disc: Compact Disc (CD), Digital Versatile Disc (DVD), Blu-ray Disc (BD)
- Can’t do:
  - Punch cards
  - Open reel and cartridge tapes
  - Video game cartridges
  - Damaged hard drives, floppy disc

Photo stand for photographing storage media


The Hitchhiker's Guide to the Galaxy
PLANETFALL
PLANETFALL
Verbatim
VALULIFE
CERTIFIED DISKETTE

Workstation to capture files from 5.25 inch floppy, optical, and Zip disks


Click on image to zoom, drag to pan
CPU socket: AM3
Memory slots: dual channel DDR31866(OC)/ 1333/1066
8-pin CPU power connector
24-pin ATX power connector
IDE/PATA connector
PCI Express x1 slots
SATA connectors: 3Gb/s
SATA connectors: 6Gb/s
PCI slots
FDD connector
PCI Express 2.0 x16 slots

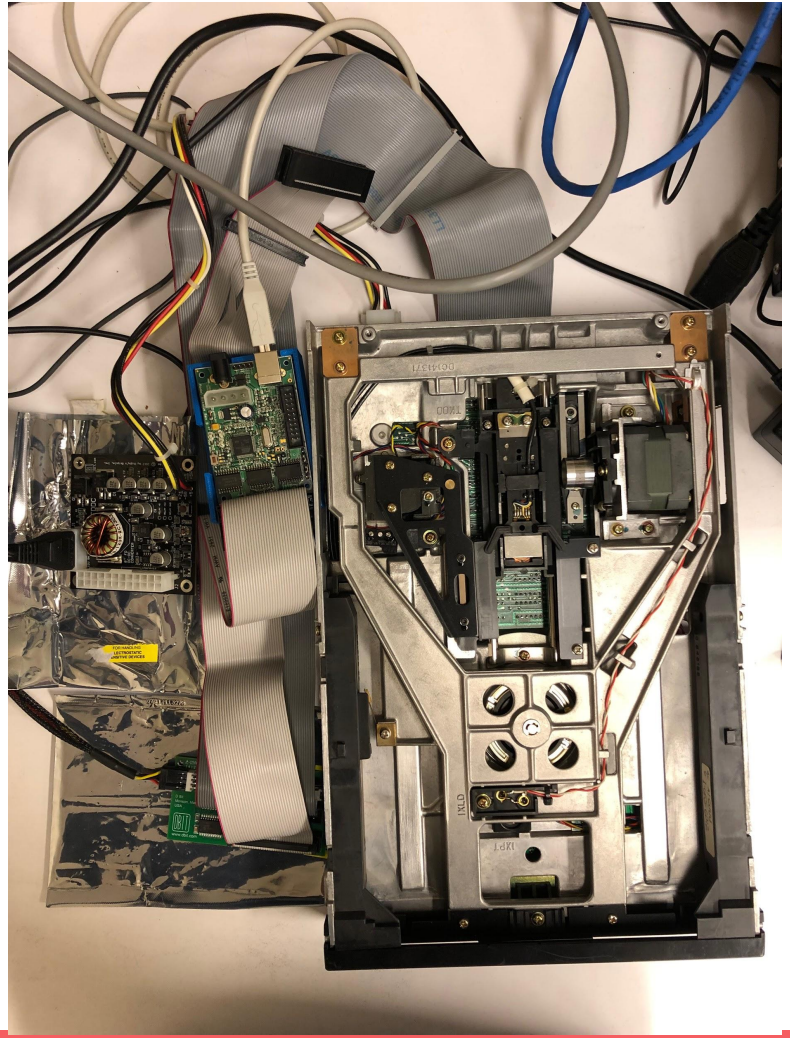
Dear Peter,

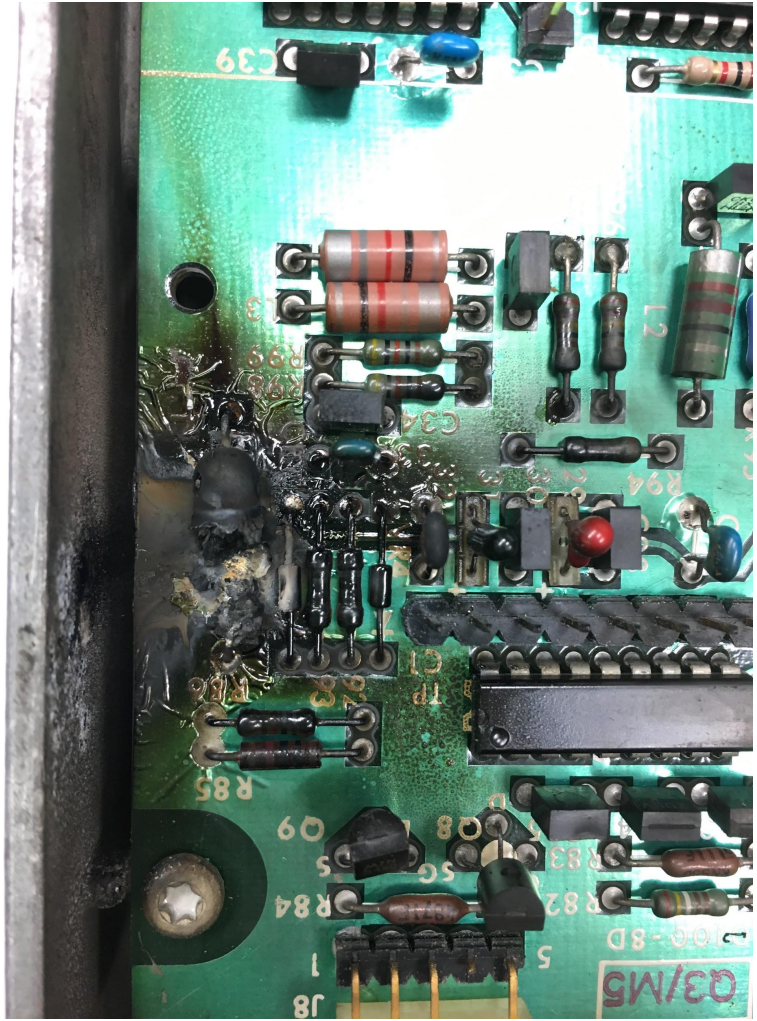
Unfortunately we do not manufacture any motherboards now a days which can support the 5.25 floppy. The interface are different than 3.5 and they are becoming obsolete and are no longer available on the newer motherboards.

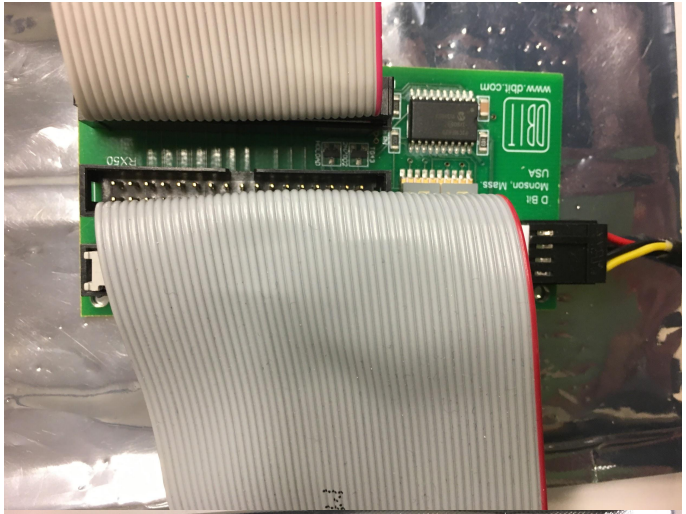
2010

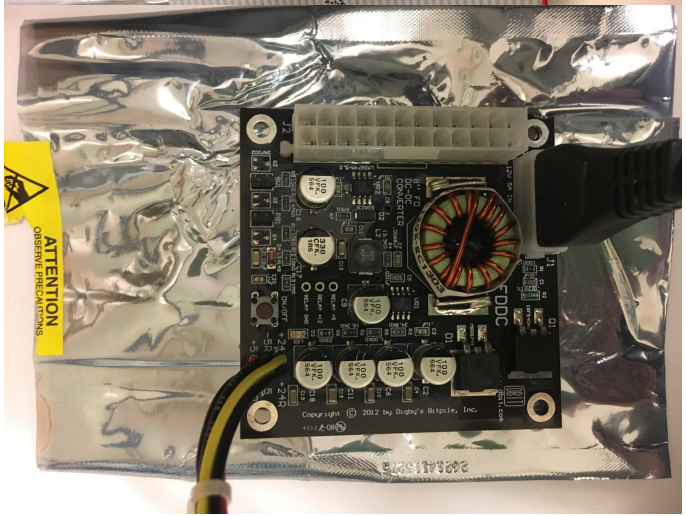
8-inch, 5.25-inch, and 3.5-inch floppy

8 inch floppy disk drive






www.dbit.com
DBIT
RX50


ATTENTION
OBSERVE PRECAUTIONS
FDDC
Copyright © 2012 by Digby's Bitpile, Inc.


200 PSI WORKING PRESSURE - TESTED TO 400 LB. 1987


Disc No:
SYNCRON.
SF2
SC0222_CM001
Corporate Express

KryoFlux

FRED (Forensic Recovery Evidence Device) to capture hard drives

Forensic Toolkit Suite

$4,114.85

SMS *

- 1 Year (+$1,222.76)
- 2 Years (+$2,445.52)
- 3 Years (+$3,668.28)

1x Forensic Toolkit Suite	**$4,114.85**

Subtotal $4,114.85

1 Add To Cart

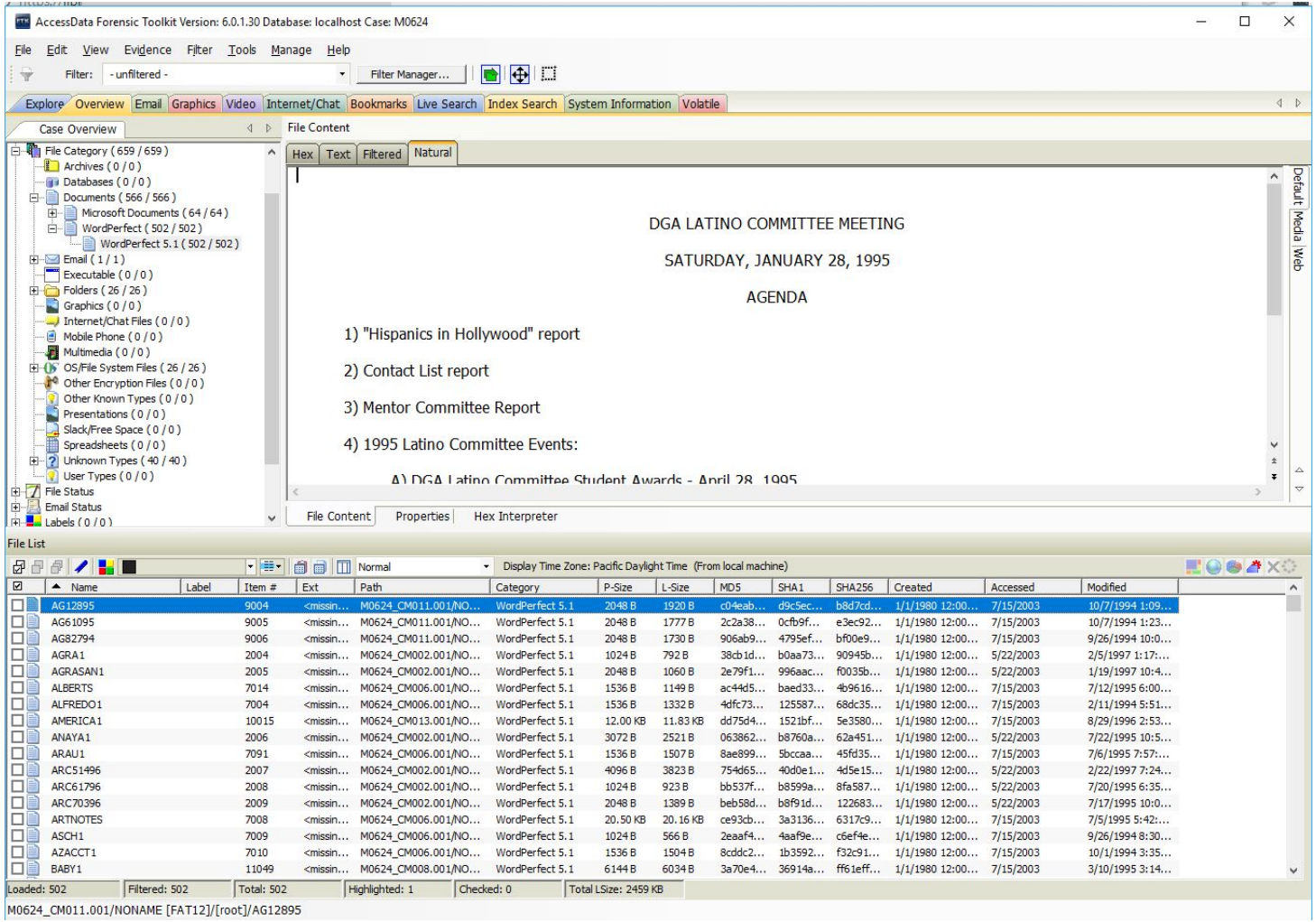
Add to Wishlist

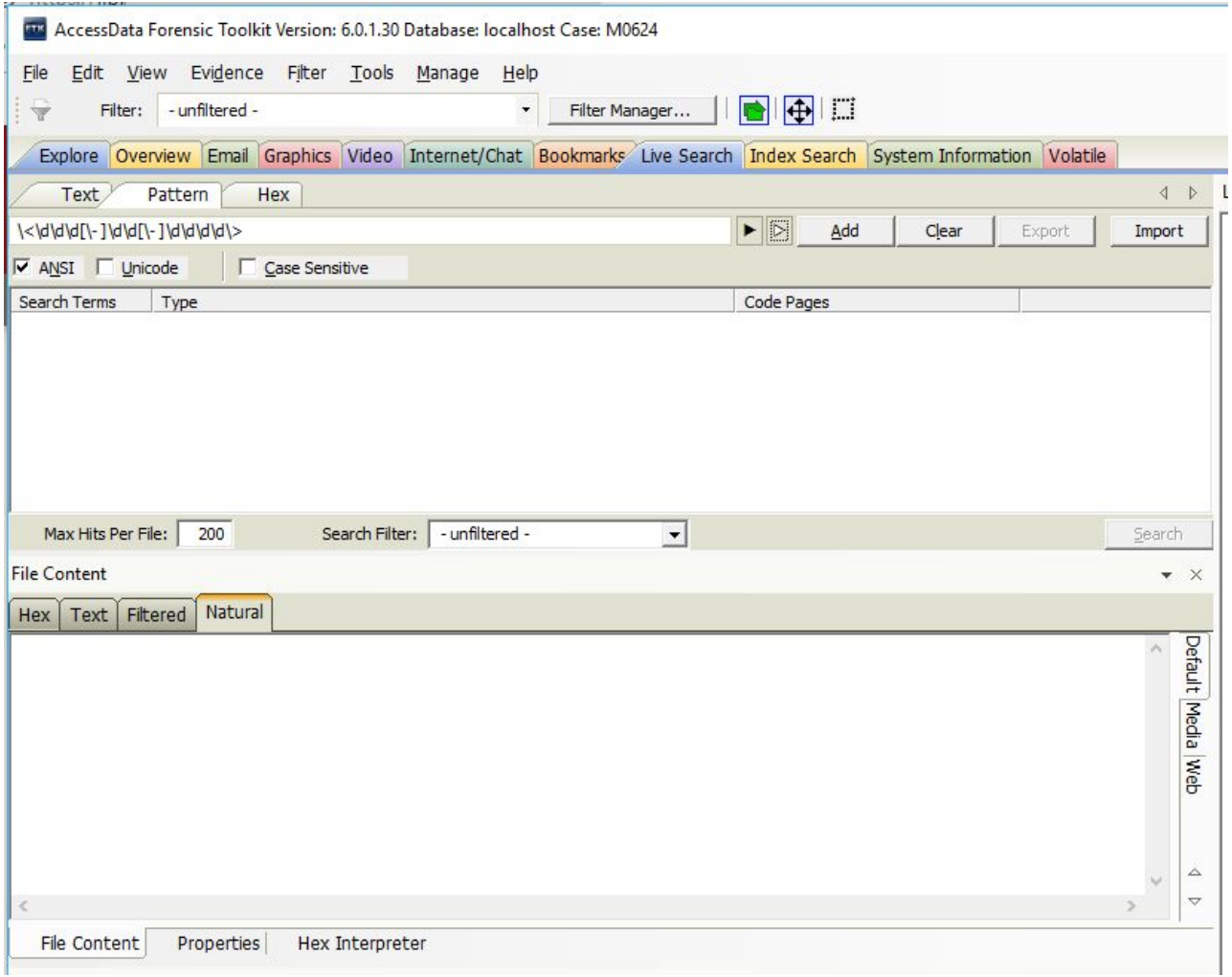
File Items

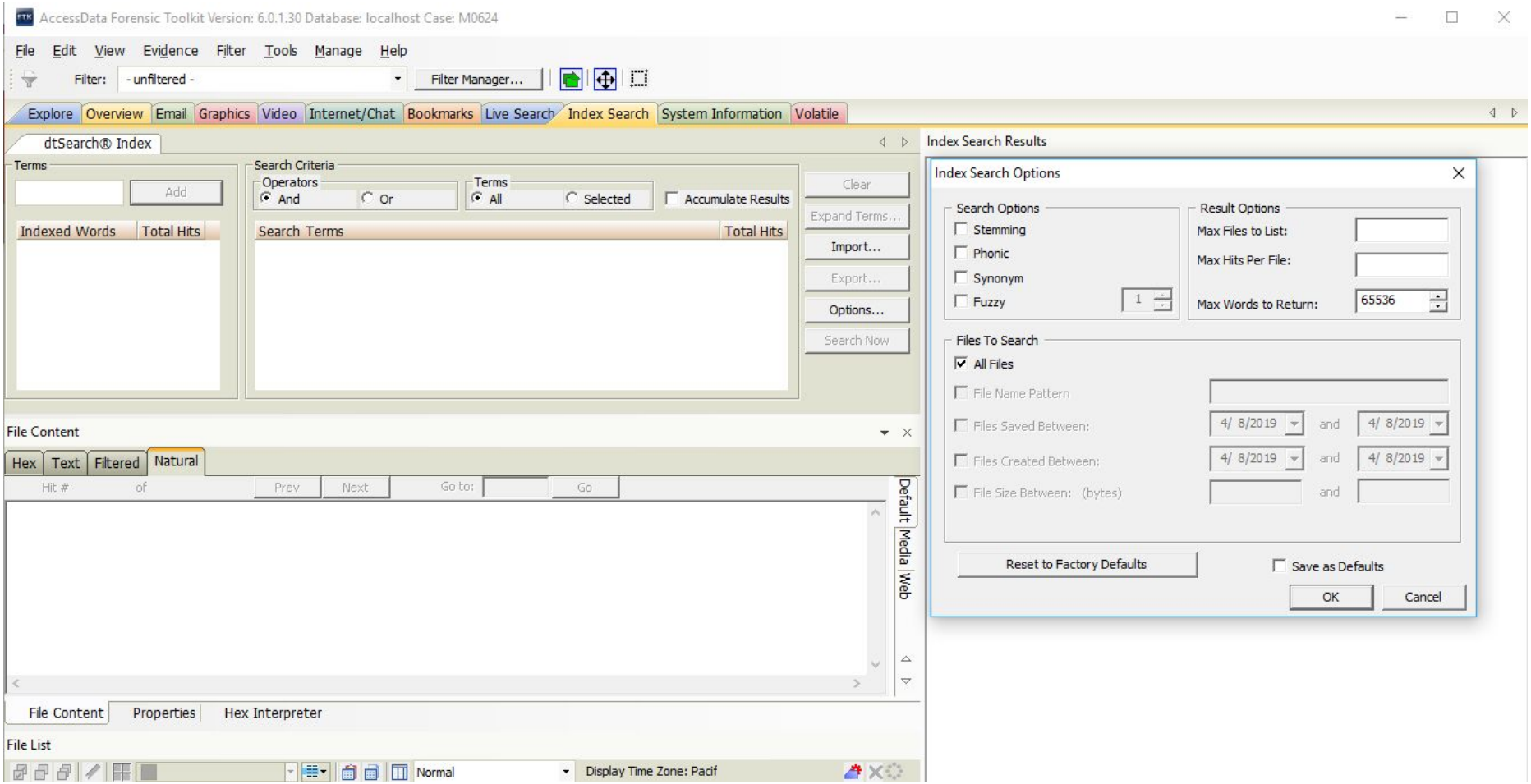
Evidence Items: 576
Checked Items: 0
Unchecked Items: 313911

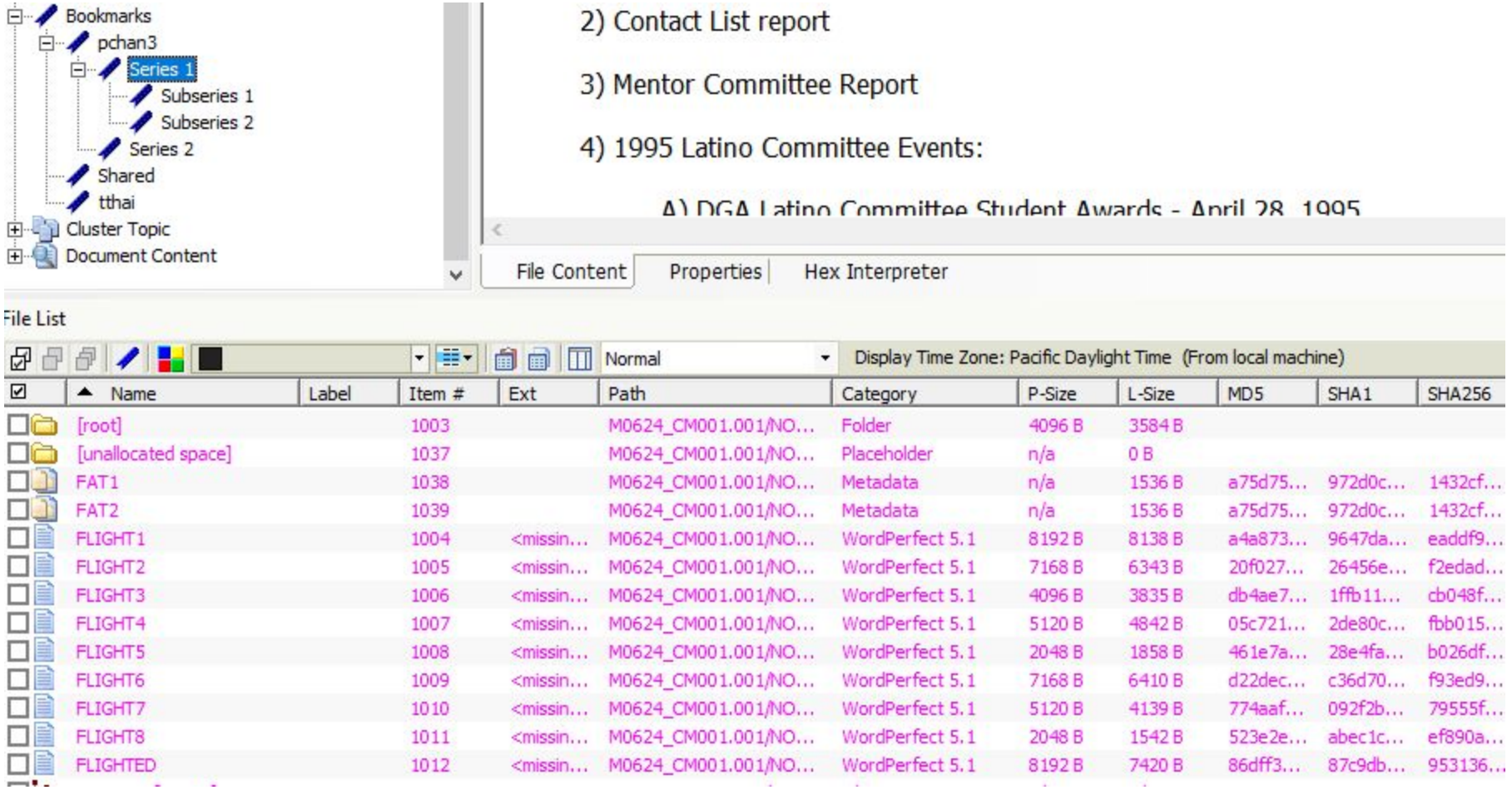
File Category

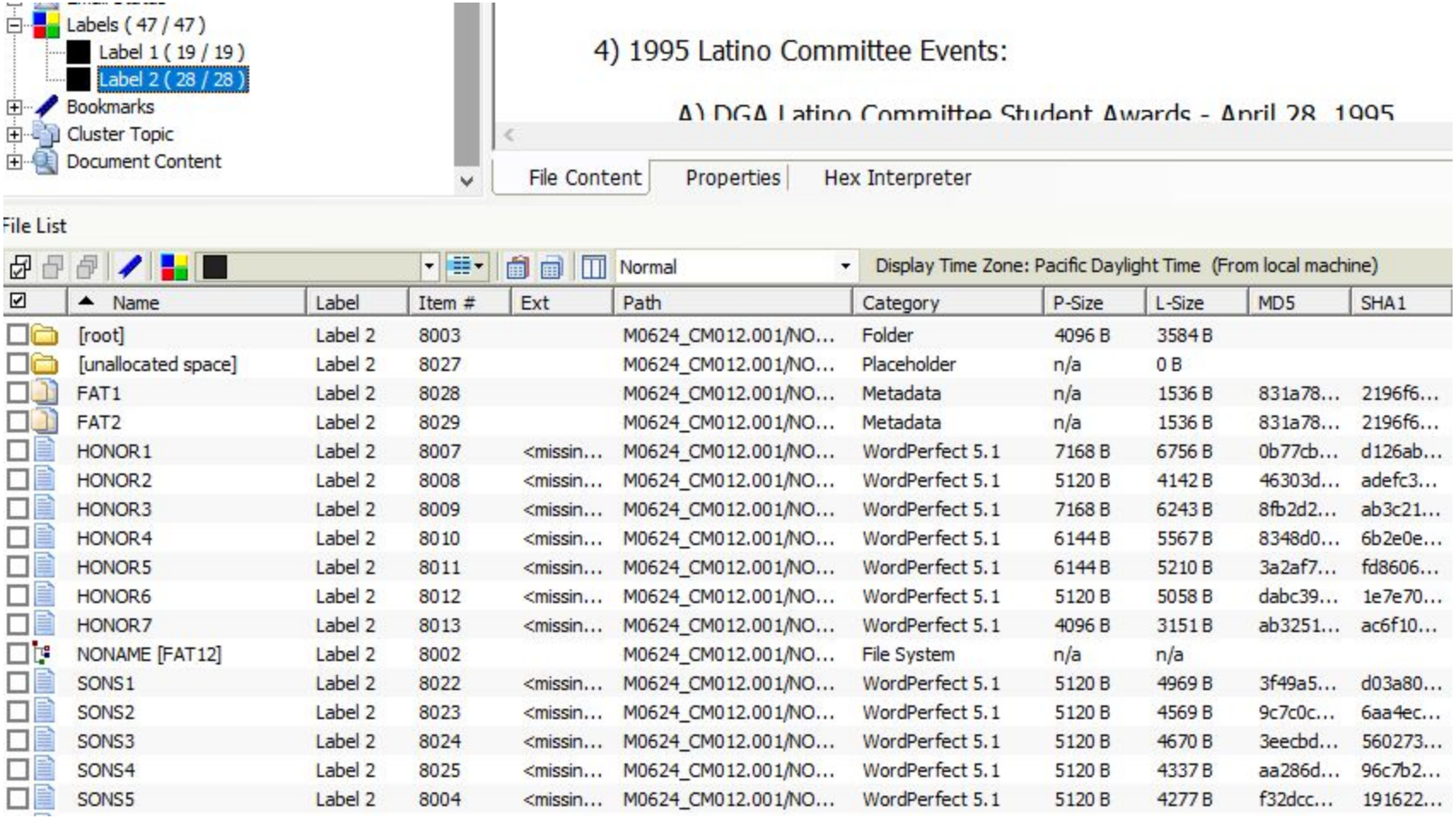
Archives: 1004
Databases: 5
Documents: 31580
Email: 214
Executable: 6034
Folders: 14293
Graphics: 128485
Internet/Chat Files: 0
Mobile Phone: 0
Multimedia: 62820
OS/File System Files: 1298
Other Encryption Files: 68
Other Known Types: 20
Presentations: 0
Slack/Free Space: 44
Spreadsheets: 52
Unknown Types: 67994
User Types: 0


AccessData Forensic Toolkit Version: 6.0.1.30 Database: localhost Case: M0624
File Edit View Evidence Filter Tools Manage Help
Filter: - unfiltered - Filter Manager...
Explore Overview Email Graphics Video Internet/Chat Bookmarks Live Search Index Search System Information Volatile
Case Overview
File Category (659 / 659)
Archives (0 / 0)
Databases (0 / 0)
Documents (566 / 566)
Microsoft Documents (64 / 64)
WordPerfect (502 / 502)
WordPerfect 5.1 (502 / 502)
Email (1 / 1)
Executable (0 / 0)
Folders (26 / 26)
Graphics (0 / 0)
Internet/Chat Files (0 / 0)
Mobile Phone (0 / 0)
Multimedia (0 / 0)
OS/File System Files (26 / 26)
Other Encryption Files (0 / 0)
Other Known Types (0 / 0)
Presentations (0 / 0)
Slack/Free Space (0 / 0)
Spreadsheets (0 / 0)
Unknown Types (40 / 40)
User Types (0 / 0)
File Status
Email Status
Labels (0 / 0)
File Content
Hex Text Filtered Natural
DGA LATINO COMMITTEE MEETING
SATURDAY, JANUARY 28, 1995
AGENDA
1) "Hispanics in Hollywood" report
2) Contact List report
3) Mentor Committee Report
4) 1995 Latino Committee Events:
A) DGA Latino Committee Student Awards - April 28, 1995
Default Media Web
File Content Properties Hex Interpreter
File List
Normal
Display Time Zone: Pacific Daylight Time (From local machine)
Name Label Item # Ext Path Category P-Size L-Size MD5 SHA1 SHA256 Created Accessed Modified
AG12895 9004 <missin... M0624_CM011.001/NO... WordPerfect 5.1 2048 B 1920 B c04eab... d9c5ec... b8d7cd... 1/1/1980 12:00... 7/15/2003 10/7/1994 1:09...
AG61095 9005 <missin... M0624_CM011.001/NO... WordPerfect 5.1 2048 B 1777 B 2c2a38... 0cfb9f... e3ec92... 1/1/1980 12:00... 7/15/2003 10/7/1994 1:23...
AG82794 9006 <missin... M0624_CM011.001/NO... WordPerfect 5.1 2048 B 1730 B 906ab9... 4795ef... bf00e9... 1/1/1980 12:00... 7/15/2003 9/26/1994 10:0...
AGRA1 2004 <missin... M0624_CM002.001/NO... WordPerfect 5.1 1024 B 792 B 38cb1d... b0aa73... 90945b... 1/1/1980 12:00... 5/22/2003 2/5/1997 1:17:...
AGRASAN1 2005 <missin... M0624_CM002.001/NO... WordPerfect 5.1 2048 B 1060 B 2e79f1... 996aac... f0035b... 1/1/1980 12:00... 5/22/2003 1/19/1997 10:4...
ALBERTS 7014 <missin... M0624_CM006.001/NO... WordPerfect 5.1 1536 B 1149 B ac44d5... baed33... 4b9616... 1/1/1980 12:00... 7/15/2003 7/12/1995 6:00...
ALFREDO1 7004 <missin... M0624_CM006.001/NO... WordPerfect 5.1 1536 B 1332 B 4dfc73... 125587... 68dc35... 1/1/1980 12:00... 7/15/2003 2/11/1994 5:51...
AMERICA1 10015 <missin... M0624_CM013.001/NO... WordPerfect 5.1 12.00 KB 11.83 KB dd75d4... 1521bf... 5e3580... 1/1/1980 12:00... 7/15/2003 8/29/1996 2:53...
ANAYA1 2006 <missin... M0624_CM002.001/NO... WordPerfect 5.1 3072 B 2521 B 063862... b8760a... 62a451... 1/1/1980 12:00... 5/22/2003 7/22/1995 10:5...
ARAU1 7091 <missin... M0624_CM006.001/NO... WordPerfect 5.1 1536 B 1507 B 8ae899... 5bccaa... 45fd35... 1/1/1980 12:00... 7/15/2003 7/6/1995 7:57:...
ARC51496 2007 <missin... M0624_CM002.001/NO... WordPerfect 5.1 4096 B 3823 B 754d65... 40d0e1... 4d5e15... 1/1/1980 12:00... 5/22/2003 2/22/1997 7:24...
ARC61796 2008 <missin... M0624_CM002.001/NO... WordPerfect 5.1 1024 B 923 B bb537f... b8599a... 8fa587... 1/1/1980 12:00... 5/22/2003 7/20/1995 6:35...
ARC70396 2009 <missin... M0624_CM002.001/NO... WordPerfect 5.1 2048 B 1389 B beb58d... b8f91d... 122683... 1/1/1980 12:00... 5/22/2003 7/17/1995 10:0...
ARTNOTES 7008 <missin... M0624_CM006.001/NO... WordPerfect 5.1 20.50 KB 20.16 KB ce93cb... 3a3136... 6317c9... 1/1/1980 12:00... 7/15/2003 7/5/1995 5:42:...
ASCH1 7009 <missin... M0624_CM006.001/NO... WordPerfect 5.1 1024 B 566 B 2eaaf4... 4aaf9e... c6ef4e... 1/1/1980 12:00... 7/15/2003 9/26/1994 8:30...
AZACCT1 7010 <missin... M0624_CM006.001/NO... WordPerfect 5.1 1536 B 1504 B 8cddc2... 1b3592... f32c91... 1/1/1980 12:00... 7/15/2003 10/1/1994 3:35...
BABY1 11049 <missin... M0624_CM008.001/NO... WordPerfect 5.1 6144 B 6034 B 3a70e4... 36914a... ff61eff... 1/1/1980 12:00... 7/15/2003 3/10/1995 3:14...
Loaded: 502 Filtered: 502 Total: 502 Highlighted: 1 Checked: 0 Total LSize: 2459 KB
M0624_CM011.001/NONAME [FAT12]/[root]/AG12895


AccessData Forensic Toolkit Version: 6.0.1.30 Database: localhost Case: M0624
File Edit View Evidence Filter Tools Manage Help
Filter: - unfiltered - Filter Manager...
Explore Overview Email Graphics Video Internet/Chat Bookmarks Live Search Index Search System Information Volatile
Text Pattern Hex
\<\d\d\d[\-]\d\d[\-]\d\d\d\d\>
Add Clear Export Import
ANSI Unicode Case Sensitive
Search Terms Type Code Pages
Max Hits Per File: 200 Search Filter: - unfiltered - Search
File Content
Hex Text Filtered Natural
Default Media Web
File Content Properties Hex Interpreter


AccessData Forensic Toolkit Version: 6.0.1.30 Database: localhost Case: M0624
File Edit View Evidence Filter Tools Manage Help
Filter: - unfiltered - Filter Manager...
Explore Overview Email Graphics Video Internet/Chat Bookmarks Live Search Index Search System Information Volatile
dtSearch® Index
Terms
Add
Indexed Words Total Hits
Search Criteria
Operators And Or
Terms All Selected
Accumulate Results
Search Terms Total Hits
Clear
Expand Terms...
Import...
Export...
Options...
Search Now
File Content
Hex Text Filtered Natural
Hit # of Prev Next Go to: Go
Default Media Web
File Content Properties Hex Interpreter
File List
Normal
Display Time Zone: Pacif
Index Search Results
Index Search Options
Search Options
Stemming
Phonic
Synonym
Fuzzy 1
Result Options
Max Files to List:
Max Hits Per File:
Max Words to Return: 65536
Files To Search
All Files
File Name Pattern
Files Saved Between: 4/ 8/2019 and 4/ 8/2019
Files Created Between: 4/ 8/2019 and 4/ 8/2019
File Size Between: (bytes) and
Reset to Factory Defaults
Save as Defaults
OK
Cancel


Bookmarks
pchan3
Series 1
Subseries 1
Subseries 2
Series 2
Shared
tthai
Cluster Topic
Document Content
2) Contact List report
3) Mentor Committee Report
4) 1995 Latino Committee Events:
File Content
Properties
Hex Interpreter
File List
Normal
Display Time Zone: Pacific Daylight Time (From local machine)
Name
Label
Item #
Ext
Path
Category
P-Size
L-Size
MD5
SHA1
SHA256
[root] 1003 M0624_CM001.001/NO... Folder 4096 B 3584 B
[unallocated space] 1037 M0624_CM001.001/NO... Placeholder n/a 0 B
FAT1 1038 M0624_CM001.001/NO... Metadata n/a 1536 B a75d75... 972d0c... 1432cf...
FAT2 1039 M0624_CM001.001/NO... Metadata n/a 1536 B a75d75... 972d0c... 1432cf...
FLIGHT1 1004 <missin... M0624_CM001.001/NO... WordPerfect 5.1 8192 B 8138 B a4a873... 9647da... eaddf9...
FLIGHT2 1005 <missin... M0624_CM001.001/NO... WordPerfect 5.1 7168 B 6343 B 20f027... 26456e... f2edad...
FLIGHT3 1006 <missin... M0624_CM001.001/NO... WordPerfect 5.1 4096 B 3835 B db4ae7... 1ffb11... cb048f...
FLIGHT4 1007 <missin... M0624_CM001.001/NO... WordPerfect 5.1 5120 B 4842 B 05c721... 2de80c... fbb015...
FLIGHT5 1008 <missin... M0624_CM001.001/NO... WordPerfect 5.1 2048 B 1858 B 461e7a... 28e4fa... b026df...
FLIGHT6 1009 <missin... M0624_CM001.001/NO... WordPerfect 5.1 7168 B 6410 B d22dec... c36d70... f93ed9...
FLIGHT7 1010 <missin... M0624_CM001.001/NO... WordPerfect 5.1 5120 B 4139 B 774aaf... 092f2b... 79555f...
FLIGHT8 1011 <missin... M0624_CM001.001/NO... WordPerfect 5.1 2048 B 1542 B 523e2e... abec1c... ef890a...
FLIGHTED 1012 <missin... M0624_CM001.001/NO... WordPerfect 5.1 8192 B 7420 B 86dff3... 87c9db... 953136...



Name	Label	Item #	Ext	Path	Category	P-Size	L-Size	MD5	SHA1
[root]	Label 2	8003		M0624_CM012.001/NO...	Folder	4096 B	3584 B		
[unallocated space]	Label 2	8027		M0624_CM012.001/NO...	Placeholder	n/a	0 B		
FAT1	Label 2	8028		M0624_CM012.001/NO...	Metadata	n/a	1536 B	831a78...	2196f6...
FAT2	Label 2	8029		M0624_CM012.001/NO...	Metadata	n/a	1536 B	831a78...	2196f6...
HONOR1	Label 2	8007	<missin...	M0624_CM012.001/NO...	WordPerfect 5.1	7168 B	6756 B	0b77cb...	d126ab...
HONOR2	Label 2	8008	<missin...	M0624_CM012.001/NO...	WordPerfect 5.1	5120 B	4142 B	46303d...	adefc3...
HONOR3	Label 2	8009	<missin...	M0624_CM012.001/NO...	WordPerfect 5.1	7168 B	6243 B	8fb2d2...	ab3c21...
HONOR4	Label 2	8010	<missin...	M0624_CM012.001/NO...	WordPerfect 5.1	6144 B	5567 B	8348d0...	6b2e0e...
HONOR5	Label 2	8011	<missin...	M0624_CM012.001/NO...	WordPerfect 5.1	6144 B	5210 B	3a2af7...	fd8606...
HONOR6	Label 2	8012	<missin...	M0624_CM012.001/NO...	WordPerfect 5.1	5120 B	5058 B	dabc39...	1e7e70...
HONOR7	Label 2	8013	<missin...	M0624_CM012.001/NO...	WordPerfect 5.1	4096 B	3151 B	ab3251...	ac6f10...
NONAME [FAT12]	Label 2	8002		M0624_CM012.001/NO...	File System	n/a	n/a		
SONS1	Label 2	8022	<missin...	M0624_CM012.001/NO...	WordPerfect 5.1	5120 B	4969 B	3f49a5...	d03a80...
SONS2	Label 2	8023	<missin...	M0624_CM012.001/NO...	WordPerfect 5.1	5120 B	4569 B	9c7c0c...	6aa4ec...
SONS3	Label 2	8024	<missin...	M0624_CM012.001/NO...	WordPerfect 5.1	5120 B	4670 B	3eecbd...	560273...
SONS4	Label 2	8025	<missin...	M0624_CM012.001/NO...	WordPerfect 5.1	5120 B	4337 B	aa286d...	96c7b2...
SONS5	Label 2	8004	<missin...	M0624_CM012.001/NO...	WordPerfect 5.1	5120 B	4277 B	f32dcc...	191622...

AccessData FTK - Pros

- Generate technical metadata (file format, checksum, etc.)
- Ability to assign files to bookmarks and series
- Build-in file viewer for few hundred file formats
- Regular expression search (e.g. social security no.)
- Full text search
- Maintain file integrity

AccessData FTK - Cons

- Cost
- Complexibility
- Cannot remove files from the case
- Proprietary software
- Designed for forensic use cases for law enforcement agencies
- Not designed for archival process

https://www.youtube.com/watch?v=hDAhbR8dyp8

Stanford Media Preservation Lab

Digitise audio and video materials

Reformat file formats for preservation and delivery

Details at

https://library.stanford.edu/research/digitization-services/labs/stanford-media-preservation-lab/capture-specs

Stanford Digital Repository (SDR)

- The SDR is a service supporting long-term management of scholarly information resources at Stanford.
- Administrative controls enable depositors to specify content licenses, control content release through embargo, and manage public access levels for finding, viewing and downloading content.
- Metadata describing the content is indexed for search and discovery in SearchWorks, and copies of ingested content are provided via persistent URLs (PURLs) to authorized users.
- Each digital object in SDR is stored redundantly in geo-diverse locations and audited systematically to ensure bit preservation.

Reading Room Computer

Digital Preservation

“The year is 2045, and my grandchildren (as yet unborn) are exploring the attic of my house (as yet unbought). They find a letter dated 1995 and a CD-ROM (compact disk). The letter claims that the disk contains a document that provides the key to obtaining my fortune (as yet unearned). My grandchildren are understandably excited, but they have never seen a CD before—except in old movies—and even if they can somehow find a suitable disk drive, how will they run the software necessary to interpret the information on the disk? How can they read my obsolete digital document?”

”Ensuring the Longevity of Digital Information" Jeff Rothenberg, January 1995 edition of Scientific American (Vol. 272, Number 1, pp. 42-7)

Backup is NOT preservation!

It is important to understand that backup alone is not digital preservation. It is only a part of it.

More importantly, how good are your backups if you never practice recovery? If you have not tried restoring any of your backups, then assume they will likely fail. It is important to occasionally restore from tape and to check digital materials at random. Relying solely on third party backup systems to do the work for you is NOT digital preservation.

https://libguides.bodleian.ox.ac.uk/digitalpreservation/whatisdp

Digital Preservation Terminology

Digital preservation is a broad field that encompasses everything from project management to technical skills. Not everyone working in digital preservation can possess every skill, but it is the combination of teams with complementary skills that makes a successful digital preservation programme in an organization possible. Having an awareness of the theories behind digital preservation and the risks to digital assets is perhaps the most important universal skill.

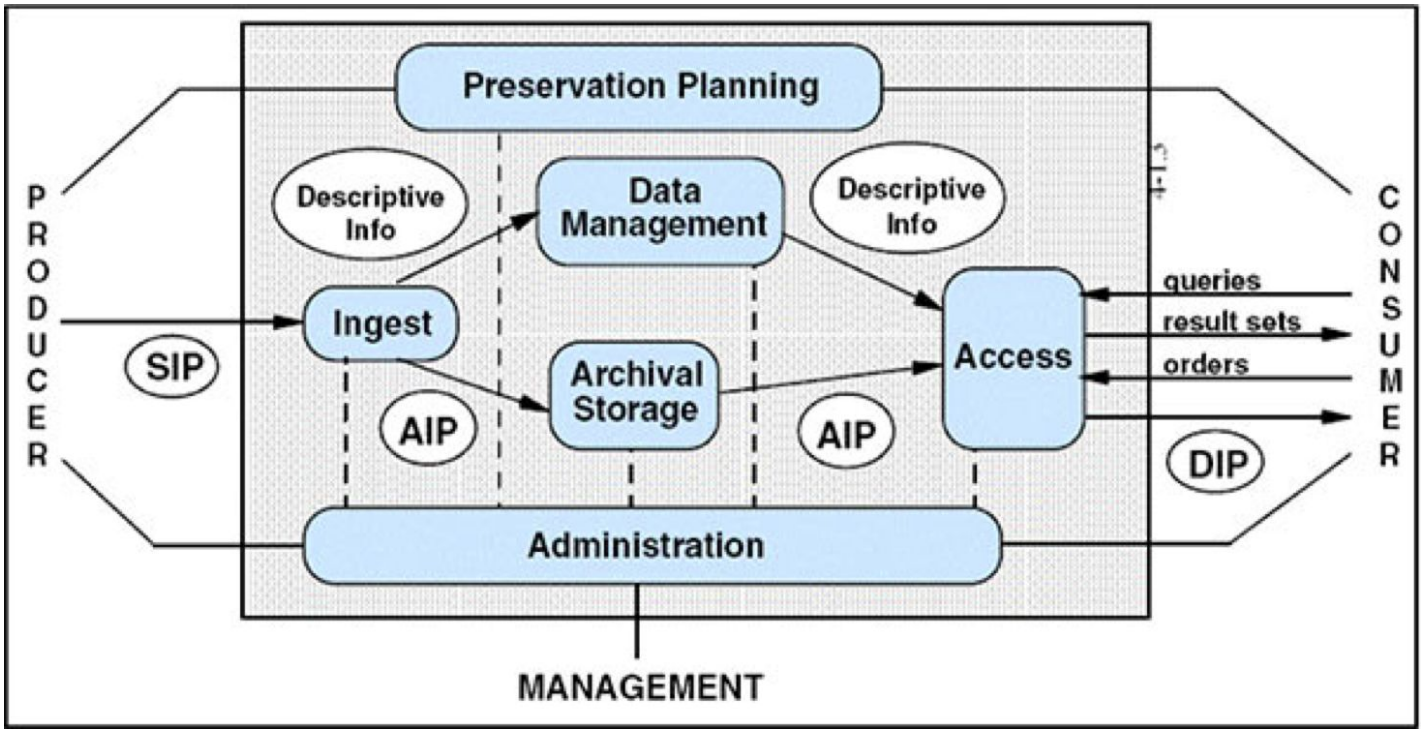
https://libguides.bodleian.ox.ac.uk/digitalpreservation/whatisdp

Digital preservation

- The formal activity of ensuring access to digital information for as long as necessary. It requires policies, planning, resource allocation (funds, time, people) and appropriate technologies and actions to ensure accessibility, accurate rendering and authenticity of digital objects.
- A “lifecycle management” approach to digital preservation is taken, where action is done at regular intervals and future activity is planned. This includes policies and recommendations for appraising and selecting digital information to preserve, acknowledging resources are finite.

https://libguides.bodleian.ox.ac.uk/digitalpreservation/whatisdp

Open Archival Information Systems (OAIS) Reference Model



Two different kinds of digital preservation

- Bit level preservation

- Logical preservation

https://libguides.bodleian.ox.ac.uk/digitalpreservation/whatisdp

Bit Level Preservation

- A term used to denote a very basic level of preservation of the digital object as it was submitted (literally preserving the bits forming a digital object).

- Bit preservation is not digital preservation but it does provide one building block for the more complete set of digital preservation practices and processes that ensure the survival of digital material and also its usability, display, context and interpretation over time.

https://libguides.bodleian.ox.ac.uk/digitalpreservation/whatisdp

Logical Preservation

- The aspect of preservation management that is concerned with ensuring the continued usability of meaningful information content, by ensuring the existence of a usable manifestation the digital object. Sometimes referred to as format preservation or active preservation. It is comprised of three stages:
    - **Characterize**: understanding what digital materials are in the repository
    - **Plan:** decision-making part based on the information gathered from characterization. This will identify threats to continued availability and accessibility, and to plan the actions that will be taken for at risk digital materials
    - **Act:** putting things into action and this should be a mechanistic process as all of the intellectual thinking was done as the plan stage

https://libguides.bodleian.ox.ac.uk/digitalpreservation/whatisdp

Preservation Actions

- The actions that can be taken over time to mitigate the technical challenges of digital over time. These actions include maintaining fixity, migration, emulation and technology preservation. These various technical strategies can help to ensure long-term access to digital objects.
  - Fixity
  - Migration
  - Emulation
  - Technology preservation

https://libguides.bodleian.ox.ac.uk/digitalpreservation/whatisdp

Fixity

Fixity is a term commonly used in digital preservation when talking about digital files and bitstreams. Fixity means the state of being unchanged or permanent. Confirming a digital file's fixity means that it has remained the same over time. Often this process of confirming is called fixity checking or integrity checking. This process will verify that a digital object has not been altered or corrupted.

The most common way to confirm the fixity of a digital object is to create what is known as a checksum or hash for each individual file or in some cases, bitstream (mainly for audiovisual works). A checksum is a string of numbers and letters generated using a mathematical algorithm. A checksum is like a digital fingerprint for a file, because it will be unique for each file.

https://libguides.bodleian.ox.ac.uk/digitalpreservation/whatisdp

Fixity

The most common checksum algorithms used in digital preservation are: MD5, SHA-256 and SHA-1. However, there are others and they go in and out of use over time. It is important to know what algorithm was used to to generate the checksum for a digital file as they are not interoperable.

By monitoring a file's integrity from as early on as possible, any loss or corruption to that file may be detected. However, a checksum has its limits. While a mismatch of checksums during fixity checking may flag that a file's checksum has changed, it cannot diagnose the problem with the file. It can only say there was one. It will be up to you to investigate further.

https://libguides.bodleian.ox.ac.uk/digitalpreservation/whatisdp

Migration

Also known as file format migration or sometimes called file format conversion, migration is different from storage media migration and software refresh. It involves transferring, or migrating, data from an aging or obsolete file format into a new file format, possibly using new applications systems at each stage to interpret the information. Moving from one version of a file format to a later version is a standard practice of migrations. This preservation action is particularly useful when the software used to render the file format type is now obsolete and modern software cannot render it correctly. This is the case with older word processing file formats, such as those created by obsolete software like WordPerfect or WordStar.

https://libguides.bodleian.ox.ac.uk/digitalpreservation/whatisdp

Emulation

An "emulator" is a software which mimics the behaviour of another computer environment. It is used in digital preservation to access software and digital files which require obsolete technological environments to run. For example, an organization could install Windows 3.1 in an emulator and then install Corel WordPerfect version 7.x (1994) under Windows 3.1 to access a WordPerfect file from 1994.

Emulation

Emulation software has been developed by gaming enthusiasts since the early 1990s, but has also sparked debate and interest within the digital preservation community since the early 2000s. While emulation environments were originally seen as complex and time consuming to set up, new developments such as in-browser-emulation has lowered the barrier to use. Today, one of the biggest obstacles to using emulation software is instead around legal concerns. The licensing landscape for obsolete software and Operating Systems required for emulation is still complex.

https://libguides.bodleian.ox.ac.uk/digitalpreservation/whatisdp

Licensing

Licensing, however, constitutes a limited transfer of rights to use an item on stated terms and conditions. Licenses are governed by contract law and, as such, are essentially a private agreement between two parties. That agreement can involve a wide range of terms and conditions . . . and need not incorporate any public policy considerations, beyond some basic limits on what constitutes an enforceable contract

https://www.cni.org/wp-content/uploads/2013/05/Licensing-DArsenault2001Stf.pdf

Emulators

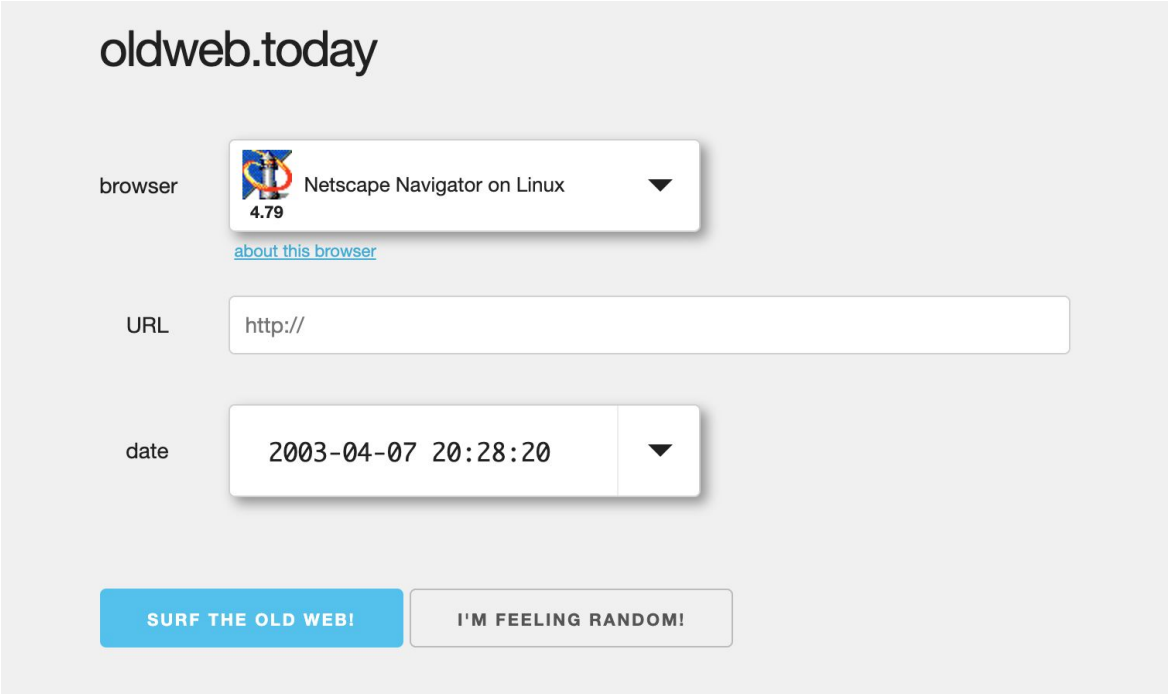
QEMU emulates x86, x86-64 systems, PowerPC, Sparc32, Sparc64, MIPS, ARM, ColdFire, Cris, Microblaze, SH4, Xtensa.

MAME/MESS can currently emulate several thousand different classic arcade video games from the late 1970s through the modern era.

Basilisk II is an Open Source 68k Macintosh emulator. That is, it allows you to run 68k MacOS software on your computer, even if you are using a different operating system.

https://mellon.org/media/filer_public/0c/3e/0c3eee7d-4166-4ba6-a767-6b42e6a1c2a7/rosenthal-emulation-2015.pdf

Emulation - Old web browser



http://oldweb.today/

You may think you don’t have computer software in your collection. Think twice! CDs published long time ago may contain software which require old operating system to run and access the contents.

A patron could not read the content in a CD which contains statistical data published in 1990s about China. It requires Chinese Windows 95 to run.

Emulation as a Service

The Emulation-as-a-Service architecture simplifies access to preserved digital assets allowing end users to interact with the original environments running on different emulators.

Ready-made emulation components provide a flexible web service API allowing for development of individual and tailored digital preservation workflows.

Emulation as a Service Infrastructure-EaaSI

Yale University Library received grants from The Andrew W. Mellon Foundation and the Alfred P. Sloan Foundation

Deployment of EaaSI nodes in at least ten partner institutions

3,000 software environments will be available from Yale University Library

Click here for Demo

Technology Preservation (museum)

- Saving everything: files, software and hardware and keep them alive
- Maintenance almost impossible
- Unworkable for larger quantities

Storage

- Storage is often the most thought about thing in digital preservation. While it is foundational to a digital preservation programme, it is only one component of it.
- When it comes to storage, you ideally want to follow these main principles, though there is no one solution for all organisations
    - 2 online copies and 2 tape (nearline and/or offline copies)
    - A minimum of 2 geographically distributed locations
    - If you are going to use cloud storage, do your research carefully (what happens if the third party suppliers goes out of business?)
    - Use different technologies (diversify your hardware and software)

https://libguides.bodleian.ox.ac.uk/digitalpreservation/whatisdp

Information Security (Access control)

Digital material selected for long-term preservation may contain embargo terms and restricted access to users according to terms in donor agreements.

Identify who has read, wrote, move, and delete authorization to individual files.

Maintain logs of who performed what actions on files, including deletion an preservation actions.

Information security methods such as encryption add to the complexity of the preservation process and should be avoided if possible for archival copies.

Descriptive / Preservation Metadata

- Content description
- Specific preservation information:
  - Provenance
  - Rights
  - Technical metadata
- File format information

https://libguides.bodleian.ox.ac.uk/digitalpreservation/whatisdp

How good / bad are your preservation repository?

Levels of Digital Preservation are a tiered set of recommendations by the National Digital Stewardship Alliance in USA for how organizations should begin to build or enhance their digital preservation activities.

Table 1: Version 1 of the Levels of Digital Preservation

	Level 1 (Protect your data)	Level 2 (Know your data)	Level 3 (Monitor your data)	Level 4 (Repair your data)
Storage and Geographic Location	- Two complete copies that are not collocated - For data on heterogeneous media (optical discs, hard drives, etc.) get the content off the medium and into your storage system	- At least three complete copies - At least one copy in a different geographic location - Document your storage system(s) and storage media and what you need to use them	- At least one copy in a geographic location with a different disaster threat - Obsolescence monitoring process for your storage system(s) and media	- At least three copies in geographic locations with different disaster threats - Have a comprehensive plan in place that will keep files and metadata on currently accessible media or systems
File Fixity and Data Integrity	- Check file fixity on ingest if it has been provided with the content - Create fixity info if it wasn't provided with the content	- Check fixity on all ingests - Use write-blockers when working with original media - Virus-check high risk content	- Check fixity of content at fixed intervals - Maintain logs of fixity info; supply audit on demand - Ability to detect corrupt data - Virus-check all content	- Check fixity of all content in response to specific events or activities - Ability to replace/repair corrupted data - Ensure no one person has write access to all copies
Information Security	- Identify who has read, write, move and delete authorization to individual files - Restrict who has those authorizations to individual files	- Document access restrictions for content	- Maintain logs of who performed what actions on files, including deletions and preservation actions	- Perform audit of logs
Metadata	- Inventory of content and its storage location - Ensure backup and non-collocation of inventory	- Store administrative metadata - Store transformative metadata and log events	- Store standard technical and descriptive metadata	- Store standard preservation metadata
File Formats	- When you can give input into the creation of digital files encourage use of a limited set of known open formats and codecs	- Inventory of file formats in use	- Monitor file format obsolescence issues	- Perform format migrations, emulation and similar activities as needed

Is that enough?

What if the preservation program is funded for 3 years? Future funding is unclear!

What if the preservation program is run by people with necessary knowledge but there is no documentation on what is being done?

Trusted Repositories Producer & Consumer

CoreTrustSeal

Certification organization supported by the World Data System of the International Science Council (WDC) and the Data Seal of Approval (DSA)

A legal entity under Dutch Law governed by Standards and Certification Board.

A community based non-profit organization promoting sustainable and trustworthy data infrastructures.

The CoreTrustSeal certification is envisioned as the first step in a global framework for repository certification which includes the extended level certification (nestor-Seal DIN 31644) and the formal level certification (ISO 16363).

Requirements - Core Trust Seal

Organizational Infrastructure

I. Mission/Scope

II. Licenses

III. Continuity of access

IV. Confidentiality/Ethics

V. Organizational infrastructure

VI. Expert guidance

Digital Object Management

VII. Data integrity and authenticity

VIII. Appraisal

IX. Documented storage procedures

X. Preservation plan

XI. Data quality

XII. Workflows

Requirements - Core Trust Seal

Digital Object Management

XIII. Data discovery and identification

XIV. Data reuse

Technology

XV. Technical infrastructure

XVI. Security

Compliance level

0 – Not applicable

1 – The repository has not considered this yet

2 – The repository has a theoretical concept

3 – The repository is in the implementation phase

4 – The guideline has been fully implemented in the repository

III. Continuity of access

R3. The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.

Guidance:

For this Requirement, please describe:

….For example, what will happen in the case of cessation of funding, which could be through an unexpected withdrawal of funding, a planned ending of funding for a time-limited project repository, or a shift of host institution interests?.....

IV. Confidentiality/Ethics

R4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.

Adherence to ethical norms is critical to responsible science. Disclosure risk—for example, the risk that an individual who participated in a survey can be identified or that the precise location of an endangered species can be pinpointed—is a concern that many repositories must address. Evidence sought is concerned with not only having good practices for data with disclosure risks, but also the necessity to maintain the trust of those agreeing to have personal/sensitive data stored in the repository.

V. Organizational infrastructure

R5. The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.

For this Requirement, responses should include evidence related to the following:

● The repository is hosted by a recognized institution (ensuring long-term stability and sustainability) appropriate to its Designated Community.

● The repository has sufficient funding, including staff resources, IT resources, and a budget for attending meetings when necessary. Ideally this should be for a three- to five year period……..

VII. Data integrity and authenticity

R7. The repository guarantees the integrity and authenticity of the data.

The repository should provide evidence to show that it operates a data and metadata management system suitable for ensuring integrity and authenticity during the processes of ingest, archival storage, and data access.

Integrity ensures that changes to data and metadata are documented and can be traced to the rationale and originator of the change.

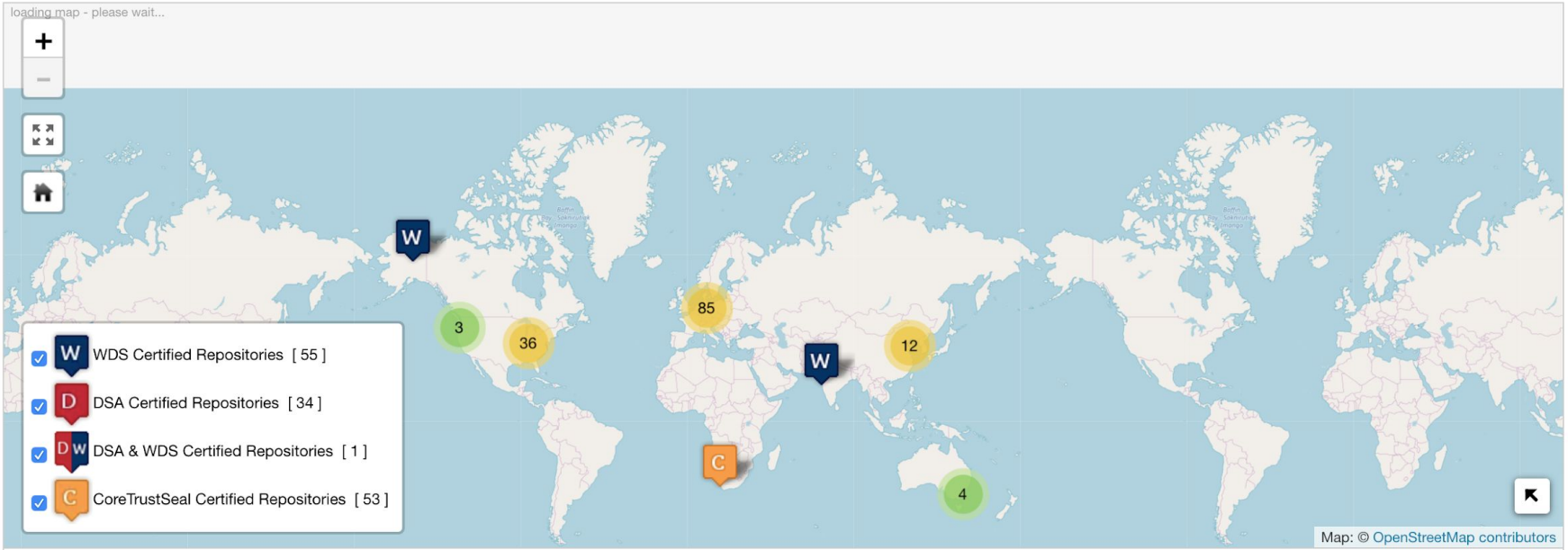
Authenticity covers the degree of reliability of the original deposited data and its provenance, including the relationship between the original data and that disseminated, and whether or not existing relationships between datasets and/or metadata are maintained.

XIII. Data discovery and identification

R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.

Effective data discovery is key to data sharing, and most repositories provide searchable catalogues describing their holdings such that potential users can evaluate data to see if they meet their needs. Once discovered, datasets should be referenceable through full citations to the data, including persistent identifiers to ensure that data can be accessed into the future. Citations also provide credit and attribution to individuals who contributed to the creation of the dataset.

Core Certified Repositories



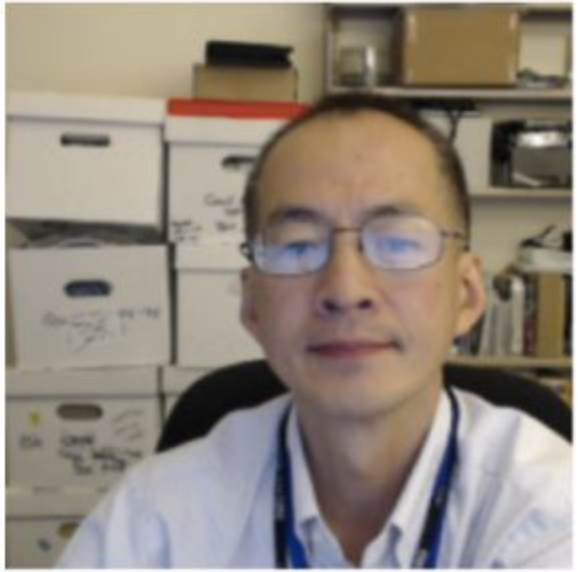
https://www.coretrustseal.org/why-certification/certified-repositories/

Digital Archivist

What Does it Take to Be a Well-rounded Digital Archivist?

October 7, 2014 by Butch Lazorchak

The following is a guest post from Peter Chan, a Digital Archivist at the Stanford University Libraries.



Peter Chan

I am a digital archivist at Stanford University. A couple of years ago, Stanford was involved in the AIMS project, which jump-started Stanford's thinking about the role of a "digital archivist." The project ended in 2011 and I am the only digital archivist hired as part of the project that is still on the job on a full-time basis. I recently had discussions with my supervisors about the roles and responsibilities of a digital archivist. This inspired me to take a look at job postings for "digital archivists" and what skills and qualifications organizations were currently looking for.

I looked at eight job advertisements for digital archivists that were published in the past 12 months. The responsibilities and qualifications required of digital archivists were very diverse in these organizations. However, all of them required formal training in archival theory and practice. Some institutions placed more emphasis on computer skills and prefer applicants to have programming skills such as PERL, XSLT, Ruby, HTML and experience working with SQL databases and repositories such as DSpace and Fedora. Others required knowledge on a variety of metadata standards. A few even desired knowledge in computer forensic tools such as FTK Imager, AccessData Forensic Toolkits and writeblockers. Most of these tools are at least somewhat familiar to digital archivists/librarians.

The Top 10 Blog Posts of 2014 on The Signal

December 24, 2014 by Erin Engle

We're fans of lists here at the Library of Congress and there is no better way to close out the year on *The Signal* than taking a look back at our popular blog posts of the year.

Our most viewed post of the year, and our second most viewed post of all time since our blog launched in 2011, was the post about the discovery of unreleased Duke Nukem video game code. It generated quite a lot of buzz and was picked up by the gaming and technical news sites, including: Polygon, Engadget, Eurogamer, The Verge, Gamasutra, and CNET.

Happy new year. Print by Currier & Ives. c1876. http://hdl.loc.gov/loc.pnp/cph.3b50424

Here's the entire list of top 10 posts of 2014 (out of 189 total posts), ranked by page views based on data as of December 22:

1. Duke's Legacy: Video Game Source Disc Preservation at the Library of Congress
2. Personal Digital Archiving: The Basics of Scanning
3. What Do you Mean by Archive? Genres of Usage for Digital Preservers
4. Research is Magic: An Interview with Ethnographers Jason Nguyen & Kurt Baer
5. Exhibiting .gifs: An Interview with curator Jason Eppink
6. New NDSA Report: The Benefits and Risks of the PDF/A-3 File Format for Archival Institutions
7. We're All Digital Archivists Now: An Interview with Sibyl Schaefer
8. The PDF's Place in a History of Paper Knowledge: An Interview with Lisa Gitelman
9. What Does it Take to Be a Well-rounded Digital Archivist?
10. Digital Archiving: Making It Personal at the Public Library

What Does it Take to Be a Well-rounded Digital Archivist?

- Responsibilities of Digital Archivists
- Knowledge / Skills / Software / Tools Needed on
  - Collection Development
  - Accessioning
  - Arrangement and Description
  - Discovery and Access
  - Preservation

https://blogs.loc.gov/thesignal/2014/10/what-does-it-take-to-be-a-well-rounded-digital-archivist/

Collection Development

Collection Development

Responsibilities of Digital Archivists

Get overall knowledge (computing habits of donors, varieties of digital material, hardware/software used, etc.) of the digital component of a collection.

Knowledge/Skills/Software/Tools Needed

In-depth knowledge of computing habits, varieties of digital material, hardware/software.

Background: AIMS Born-Digital Material Survey.

https://docs.google.com/document/d/1t-mPmTrUGm7QfljCks4wzn3HINWmya8gFGYFELDESho/edit

Collection Development

Responsibilities of Digital Archivists

Explain to donors / creators / curators the importance of digital preservation when necessary.

Knowledge/Skills/Software/Tools Needed

General knowledge of digital preservation.

Background: “Ensuring the Longevity of Digital Information” by Jeff Rothenberg, January 1995 edition of Scientific American (Vol. 272, Number 1, pp. 42-7).

Collection Development

Responsibilities of Digital Archivists

Explain to donors / creators / curators the difference between “bit preservation” and “logical preservation” when necessary.

Knowledge/Skills/Software/Tools Needed

See Introduction to Digital Preservation: What is Digital Preservation?

https://libguides.bodleian.ox.ac.uk/digitalpreservation/whatisdp

Collection Development

Responsibilities of Digital Archivists

Explain to donors / creators / curators preservation strategies such as migration, emulation and technology preservation when necessary.

Knowledge/Skills/Software/Tools Needed

Knowledge of preferred file formats for digital preservation as recommended by the Library of Congress.

Knowledge of emulation / virtualization tools and platform such as QEMU, MAME / JSMESS / MESS, SheepShaver, bwFLA, etc.

Collection Development

Responsibilities of Digital Archivists

Explain to donors / creators / curators the use of forensic software to accession and process born-digital collections when necessary.

Knowledge/Skills/Software/Tools Needed

Special knowledge of making use of forensic software in an archival context.

AccessData FTK Forensic Toolkit, FTK Imager

Collection Development

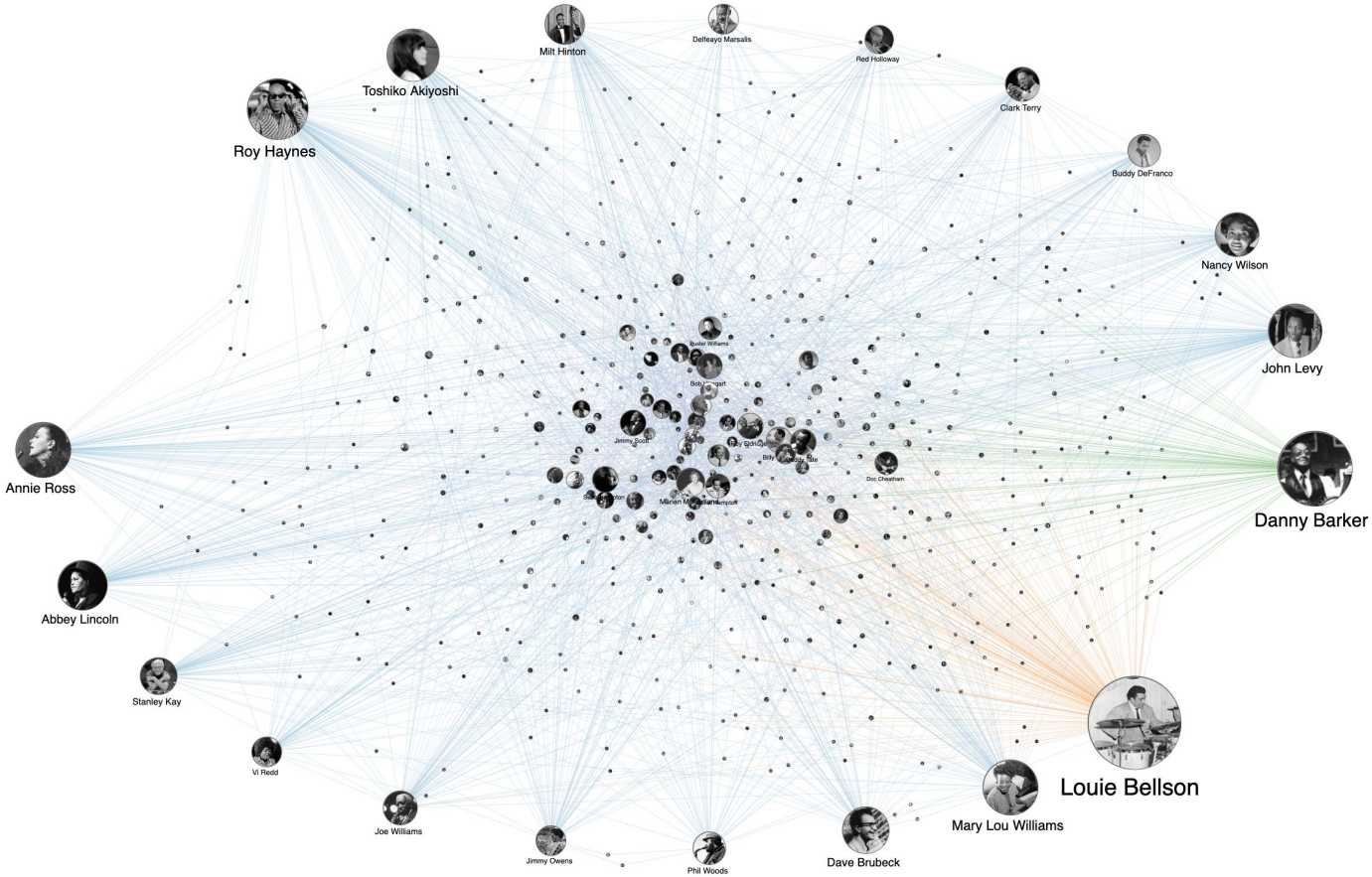
Responsibilities of Digital Archivists

Explain to donors / creators / curators the use of entity extraction / social network analysis / visualization to process and delivery born-digital collections when necessary.

Knowledge/Skills/Software/Tools Needed

General knowledge of tools used in processing and delivering born-digital archives such as entity extraction, social network analysis and visualization software.

Linked Jazz


Milt Hinton
Defleayo Marsalis
Red Holloway
Toshiko Akiyoshi
Clark Terry
Roy Haynes
Buddy DeFranco
Nancy Wilson
John Levy
Annie Ross
Danny Barker
Abbey Lincoln
Stanley Kay
Louie Bellson
Vi Redd
Mary Lou Williams
Joe Williams
Jimmy Owens
Phil Woods
Dave Brubeck

Type: Person

Show 10 entries Search:

Entity	Score	Messages
Elizabeth Thompson	1	1708
Buckminster Fuller	1	1026
David Haley	1	677
Josh Harrison	1	646
Eve Andree Laramee	1	405
Al Franken	1	383
Linda Weintraub	1	361
Noah Wardrip-Fruin	1	268
James Brady	1	248
Roger Malina	1	231

Type: Museum

Show 10 entries Search:

Entity	Score	Messages
Nevada Museum of Art	1	816
Museum of Modern Art	1	171
Art Institute of Chicago	1	137
SITE Santa Fe	1	99
MoMA PS1	1	85
Los Angeles County Museum of Art	1	83
Walker Art Center	1	80
San Francisco Museum of Modern Art	1	80
Tate Modern	1	73
Redpath Museum	1	64

Collection Development

Responsibilities of Digital Archivists

Explain to donors / creators / curators on publishing born-digital collections in linked open data vs. EAD finding aids / other HTML based web publishing method when necessary.

Knowledge/Skills/Software/Tools Needed

General knowledge of linked open data / EAD finding aids / HTML based web publishing method.

Collection Guide

∞ https://oac.cdlib.org/findaid/ark:/13030/tf7b69n911/

Collection Title: Creeley (Robert) papers, 1950-1997

Collection Number: M0662

Get Items: Online items available

Contact Stanford University::Manuscripts Division

View entire collection guide

PDF (0.85 Mb) HTML [400 hits]

Search this collection

robert creeley go

Entire Collection Guide Online Items

Collection Overview

Description The Robert Creeley Papers document the life work of a leading American poet of the 20th century, one of the core members of the "Black Mountain School." They also document several important movements in American poetics in the second half of the century. The papers include Creeley's personal and professional correspondence, journals, business records, personal mementos, clippings, artwork, and other documents generated and collected by him from 1950 to 1997.

Background Recognized as a seminal figure of American letters in the second half of the 20th century, Robert White Creeley was born in Arlington, Massachusetts, on May 21, 1926, attended the Holderness School and then Harvard College. He received degrees from The Black Mountain College (B.A., 1956) and the University of New Mexico (M.A., 1960).

Extent 443.5 Linear Feet (612 boxes, 3 cartons, 32 flat boxes, 1 oversize box, 2 cassette boxes; 55 audiocassettes, 16 reel to reel cassettes, 16 sound discs, 2 videocassettes, 4 film reels.)

Restrictions Property rights reside with the repository. Copyright resides with the creators of the documents or their heirs and assigns. To obtain permission to publish or reproduce, please contact the Public Services Librarian of the Dept. of Special Collections.

Availability The collection is open for research except that all medical records for Robert Creeley and his family have been restricted, as have student recommendations and certain financial documents. Audiovisual materials must be reformatted before use (Series 12) and Born-Digital materials in Series 13. Sub-series: Mixed

Table of contents

400 Search hits Clear search hits

The Bob Fitch Photography Archive

Movements for Change

Martin Luther King, Jr. Gallery

Trained to be a Protestant minister, Bob Fitch's career as a photojournalist began in 1965 when he joined the Rev. Martin Luther King Jr.'s organization, The Southern Christian Leadership Conference, as a staff photographer. As Fitch notes, "I worked for two intense years as the volunteer photographer for Dr. King and the SCLC, crisscrossing "Black Belt" states to document his people-to-people speaking tours promoting get-out-the-vote campaigns"

Fitch's work with the SCLC in 1965 and 1966 produced powerful images of Dr. King's speaking and leadership, as well as of the courageous efforts of marchers in events such as the 1966 Meredith March Against Fear. [View gallery]






Collection Development

Responsibilities of Digital Archivists

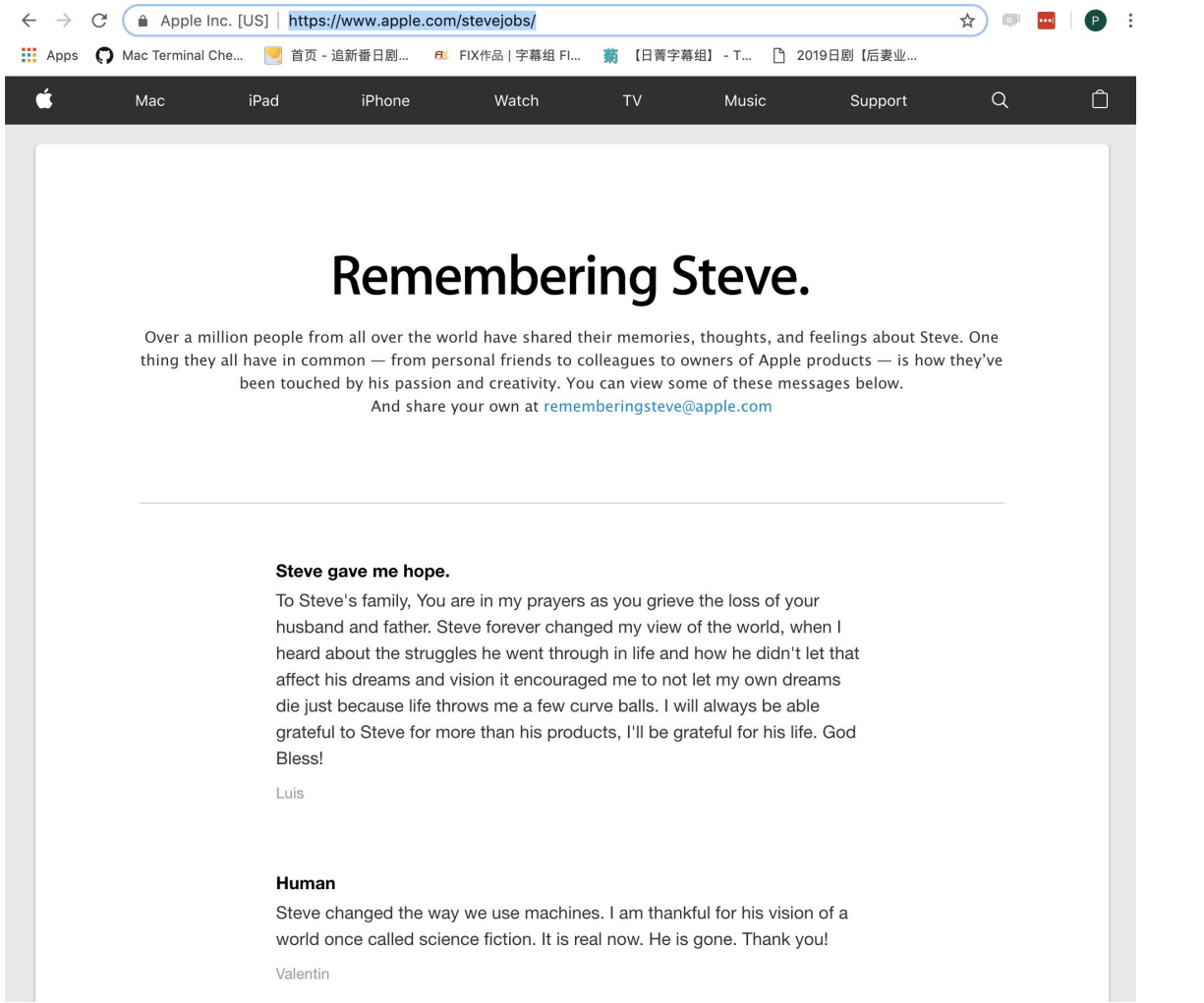
Explain web archiving to donors / creators / curators.

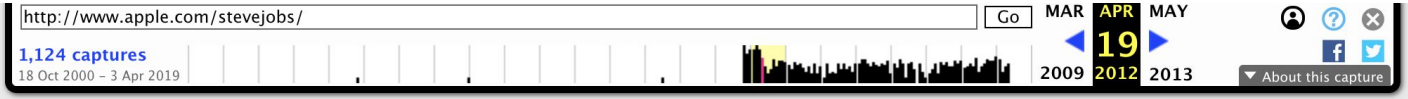
Knowledge/Skills/Software/Tools Needed

General knowledge of web archiving, cataloging, delivery and preservation of web sites.

Knowledge of web archiving software such as Heritrix and HTTrack.

Knowledge of Wayback Machine from Internet Archive.


Apple Inc. [US] | https://www.apple.com/stevejobs/
Apps
Mac Terminal Che...
首页 - 追新番日剧...
FIX作品 | 字幕组 FI...
【日菁字幕组】 - T...
2019日剧【后妻业...
Mac
iPad
iPhone
Watch
TV
Music
Support
Remembering Steve.
Over a million people from all over the world have shared their memories, thoughts, and feelings about Steve. One thing they all have in common — from personal friends to colleagues to owners of Apple products — is how they've been touched by his passion and creativity. You can view some of these messages below.
And share your own at rememberingsteve@apple.com
Steve gave me hope.
To Steve's family, You are in my prayers as you grieve the loss of your husband and father. Steve forever changed my view of the world, when I heard about the struggles he went through in life and how he didn't let that affect his dreams and vision it encouraged me to not let my own dreams die just because life throws me a few curve balls. I will always be able grateful to Steve for more than his products, I'll be grateful for his life. God Bless!
Luis
Human
Steve changed the way we use machines. I am thankful for his vision of a world once called science fiction. It is real now. He is gone. Thank you!
Valentin



Store Mac iPod iPhone iPad iTunes Support

Remembering Steve.

Over a million people from all over the world have shared their memories, thoughts, and feelings about Steve. One thing they all have in common — from personal friends to colleagues to owners of Apple products — is how they've been touched by his passion and creativity. You can view some of these messages below.
And share your own at rememberingsteve@apple.com

Remembering Steve.

Over a million people from all over the world have shared their memories, thoughts, and feelings about Steve. One thing they all have in common — from personal friends to colleagues to owners of Apple products — is how they've been touched by his passion and creativity. You can view some of these messages below.
And share your own at rememberingsteve@apple.com

:(

Such sad news. I pray that he found the Lord before his death. In life he was a man of infinite genius and the world is definitely a lot better because of his existence. Thank you Steve for all you have done

Benjamin

Amazing human being

R.I.P. Steve Jobs! I thank you for all the passion you have showed us and the amazing products that everyone can use today! You will be missed.

Anonymous

Collection Development

Responsibilities of Digital Archivists

Explain to donors / creators / curators about the archives profession in general.

Knowledge/Skills/Software/Tools Needed

Knowledge of establishing and maintaining control, arranging and describing the born-digital archival materials in accordance with accepted standards and practices to ensure the long-term preservation of collections.

Accessioning

Accessioning

Responsibilities of Digital Archivists

Copy files contained in storage media including obsolete formats such as 5.25 inch floppy disks, computer punch cards.

Knowledge/Skills/Software/Tools Needed

Knowledge of onboard 5.25 inch. floppy disk controller and hardware interface, IDE, SCSI, Firewire, SATA, FC5025, KryoFlux, Catweasel, Zip drive, computer tapes, etc.

Knowledge of file systems such as FAT, NTFS, HFS, etc.

Accessioning

Responsibilities of Digital Archivists

Ensure source data in storage media will not be erased / changed accidentally during accessioning.

Knowledge/Skills/Software/Tools Needed

Knowledge of write-protect notch / slide switch in floppy disks and hardware write blocker.

Accessioning

Responsibilities of Digital Archivists

Ensure proper audit trail in copying files from storage media.

Knowledge/Skills/Software/Tools Needed

Knowledge of forensic software (e.g. BitCurator, FTK Imager for PC and Command FTK Imager for Mac).

Accessioning

Responsibilities of Digital Archivists

Get file count, file size and file category of collections.

Knowledge/Skills/Software/Tools Needed

Knowledge of forensic software (e.g. AccessData FTK, EnCase Forensic, etc.).

Knowledge of JHOVE, DROID, Pronom, etc.

File Items

Evidence Items: 576
Checked Items: 0
Unchecked Items: 313911

File Category

Archives: 1004
Databases: 5
Documents: 31580
Email: 214
Executable: 6034
Folders: 14293
Graphics: 128485
Internet/Chat Files: 0
Mobile Phone: 0
Multimedia: 62820
OS/File System Files: 1298
Other Encryption Files: 68
Other Known Types: 20
Presentations: 0
Slack/Free Space: 44
Spreadsheets: 52
Unknown Types: 67994
User Types: 0

Accessioning

Responsibilities of Digital Archivists

Ensure computer viruses, if they exist in collection materials, are under control during accessioning.

Knowledge/Skills/Software/Tools Needed

Knowledge of the unique nature of archival materials (no replacement, etc.), behavior of viruses stored in file containers and special procedures in using antivirus software for archival materials.

Accessioning

Responsibilities of Digital Archivists

Accession email archives.

Knowledge/Skills/Software/Tools Needed

Knowledge of Internet protocol (POP, IMAP) and email format (Outlook, mbox).

Knowledge of commercial and open source software package to archive and reformat email such as Emailchemy, Mailstore, ePADD (Email: Process, Accession, Discover and Deliver).

Method used by Royal Library of Copenhagen

Accessioning

Responsibilities of Digital Archivists

Archive web sites.

Knowledge/Skills/Software/Tools Needed

Knowledge of web archiving software such as Heritrix and HTTrack.

Knowledge of legal issues in archiving web sites.

Knowledge of web archiving services such as Archive-It.

Accessioning

Responsibilities of Digital Archivists

Create accession records for born-digital archives.

Knowledge/Skills/Software/Tools Needed

Knowledge of archival data management systems such as ArchiveSpace.

Arrangement and Description

Arrangement and Description

Responsibilities of Digital Archivists

Screen out restricted, personal, classified and sensitive information such as social security numbers, credit card numbers, classified data, medical records, etc. in archives.

Knowledge/Skills/Software/Tools Needed

Knowledge of the sensitivity of personal identifiable information (PII) and tools to locate PII (e.g. AccessData FTK, Identity Finder, ePADD).

Knowledge of legal restrictions on access to data (DMCA, FERPA, etc.)

Arrangement and Description

Responsibilities of Digital Archivists

Classify text elements in born-digital materials into predefined categories such as the names of persons, organizations, locations, other entities when appropriate.

Knowledge/Skills/Software/Tools Needed

Knowledge of entity extraction software and tools to perform entity extraction (such as Open Calais, Xementa, Stanford Named Entity Recognizer, Open NLP, ePADD).

Arrangement and Description

Responsibilities of Digital Archivists

Show the network relationship of people in collections when appropriate.

Knowledge/Skills/Software/Tools Needed

Knowledge of network graph and tools such as Gephi, NodeXL.

Arrangement and Description

Responsibilities of Digital Archivists

Create controlled vocabulary to facilitate arrangement and description when appropriate.

Knowledge/Skills/Software/Tools Needed

Knowledge of the controlled vocabulary concept.

Knowledge of W3C standard for publishing controlled vocabulary (SKOS).

Knowledge of services for publishing SKOS such as Open Metadata Registry.

Knowledge of Linked Open Data publishing platform such as Wikidata.

Arrangement and Description

Responsibilities of Digital Archivists

Enhanced understanding of archival objects (e.g. video game) in additional to author, title, subject heading offered in library catalog.

E.g. provide relationship among video game titles, series, fanchise, adaptation, add-on to enhance understanding of archival objects.

Knowledge/Skills/Software/Tools Needed

Knowledge of Linked Open Data publishing platform such as Wikidata and WIkibase.

Arrangement and Description

Responsibilities of Digital Archivists

Describe files with special formats (e.g. born digital photographic images).

Knowledge/Skills/Software/Tools Needed

Knowledge of image metadata schema standard (IPTC, EXIF) and software to create/modify such metadata (Adobe Bridge, Photo Mechanic, etc.).

Arrangement and Description

Responsibilities of Digital Archivists

Describe image/video files by names of persons or objects in images with the help of software when appropriate.

Knowledge/Skills/Software/Tools Needed

Knowledge of facial / object recognition functions in software such as Picasa, Photoshop Elements and web service such as Google Vision.

Arrangement and Description

Responsibilities of Digital Archivists

Transcribe audio files with the help of software when appropriate.

Knowledge/Skills/Software/Tools Needed

Knowledge of speech-to-text functions from web service such as Google Cloud Speech-to-Text.

Arrangement and Description

Responsibilities of Digital Archivists

Create EAD finding aids.

Knowledge/Skills/Software/Tools Needed

Knowledge of accepted standards and practices in creating finding aids.

Knowledge of XML editor or other software (such as Archivists' Toolkit) to create EAD finding aids.

Discovery and Access

Discovery and Access

Responsibilities of Digital Archivists

Deliver born-digital archives in reading room computers.

Knowledge/Skills/Software/Tools Needed

Knowledge of security measures required for workstations in reading room to prevent unintentional transfer of collection materials such as disabling Internet access and USB ports.

Knowledge of software to deliver images in collections such as Adobe Bridge.

Knowledge of software to read files with obsolete file formats such as QuickView Plus.

Discovery and Access

Responsibilities of Digital Archivists

Deliver born-digital archives using Institutions' catalog system.

Knowledge/Skills/Software/Tools Needed

Knowledge of the interface required by the Institutions' catalog system to make the delivery.

Discovery and Access

Responsibilities of Digital Archivists

Deliver born-digital archives using Institution repository systems.

Knowledge/Skills/Software/Tools Needed

Knowledge of DSpace, Fedora, Hydra and their interfaces developed to facilitate such delivery.

Discovery and Access

Responsibilities of Digital Archivists

Publish born-digital archives using linked data platform.

Knowledge/Skills/Software/Tools Needed

Knowledge of linked data publishing platform such as Wikidata and Wikibase.

Discovery and Access

Responsibilities of Digital Archivists

Deliver born-digital archives using exhibition software.

Knowledge/Skills/Software/Tools Needed

Knowledge of open source exhibition software such as Omeka, Stanford Spotlight.

Discovery and Access

Responsibilities of Digital Archivists

Deliver archived web sites.

Knowledge/Skills/Software/Tools Needed

Knowledge of delivery options available in Web Archiving Services such as Archive-It.

Discovery and Access

Responsibilities of Digital Archivists

Deliver email archives.

Knowledge/Skills/Software/Tools Needed

Knowledge of commercial software such as Mailstore.

Knowledge of open source software such as ePADD (Email: Process, Accession, Discover and Deliver).

Discovery and Access

Responsibilities of Digital Archivists

Deliver software collections using emulation.

Knowledge/Skills/Software/Tools Needed

Knowledge of emulation platform such as EaasI.

Discovery and Access

Responsibilities of Digital Archivists

Deliver finding aids of born-digital archives using union catalogs such as OAC.

Knowledge/Skills/Software/Tools Needed

Knowledge of uploading procedures to respective union catalogs such as OAC.

Discovery and Access

Responsibilities of Digital Archivists

Deliver finding aids of born-digital archives using union catalogs such as OAC (Online Archive of California) .

Knowledge/Skills/Software/Tools Needed

Knowledge of uploading procedures to respective union catalogs such as OAC.

Preservation

Preservation

Responsibilities of Digital Archivists

Prepare the technical metadata (checksum, creation, modification and last access dates, file format, file size) of files in archives for transfer to preservation repository.

Knowledge/Skills/Software/Tools Needed

Knowledge of forensic software such as AccessData FTK and BitCurator, etc.

Programming skill in XSLT to extract the information when appropriate from reports generated by the software.

Preservation

Responsibilities of Digital Archivists

Use emulation to preserve software collections.

Knowledge/Skills/Software/Tools Needed

Knowledge of emulation tools and platform such as QEMU, JMESS/MESS, Basilisk II and bwFLA.

Preservation

Responsibilities of Digital Archivists

Use migration to preserve digital objects.

Knowledge/Skills/Software/Tools Needed

Knowledge of migration tools such as Xena, Adobe Acrobat Professional, etc.

Preservation

Responsibilities of Digital Archivists

Submit items to preservation repository.

Knowledge/Skills/Software/Tools Needed

Knowledge of preservation system such as Archivematica, LOCKSS and preservation services such as Portico, Tessella.

Preservation

Responsibilities of Digital Archivists

Preserve archived web sites.

Knowledge/Skills/Software/Tools Needed

Knowledge of preservation options available in Web Archiving Services such as Archive-It.

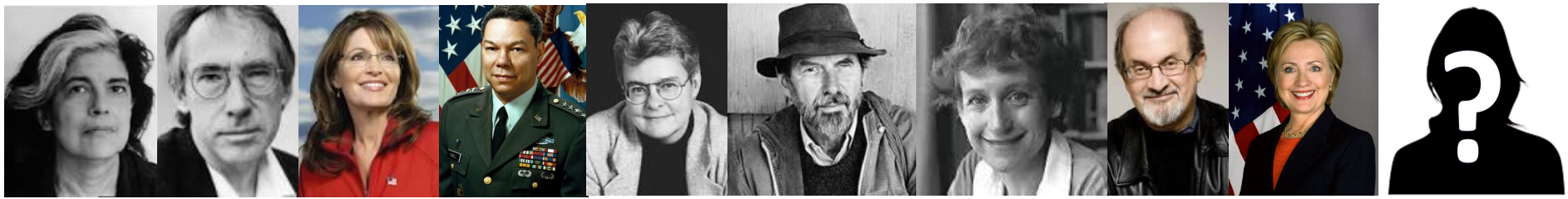
Knowledge of preserving web sites in preservation repository.

ePADD

Appraising, Processing and Providing Access to Historical Email Archives with ePADD

Peter Chan, Josh Schneider

@e_padd



Context and Overview

Email is important, ubiquitous, and ripe for research

(But…)

Email presents major challenges to donors, memory institutions, and researchers

- Can include information which donors and institutions will not or cannot share
- The scale of collections can compound these issues, and also make it difficult for researchers to find & use these materials

Ch-ch-ch-ch-changes

- Administration at libraries, archival institutions, and museums are aware of these challenges
- Fortunately, grant and funding agencies and other advocates are as well

Requirements on Email Archival Software/Process

I. Transfer Emails to Institutions

II. Normalize email to MBOX or EML or XML

III. Handle Attachments / Embedded links

IV. Detect Abnormal Email Messages

V. Identify and Handle Sensitive Information

VI. Search Functions

VII. Classify / Group Contents Automatically

VIII. Label / Annotate Functions

IX. Access to Email Archives Online and in a Reading Room

X. Integrate with Digital Preservation Repositories

XI. Work with Other Systems

ARTS | BOOKS

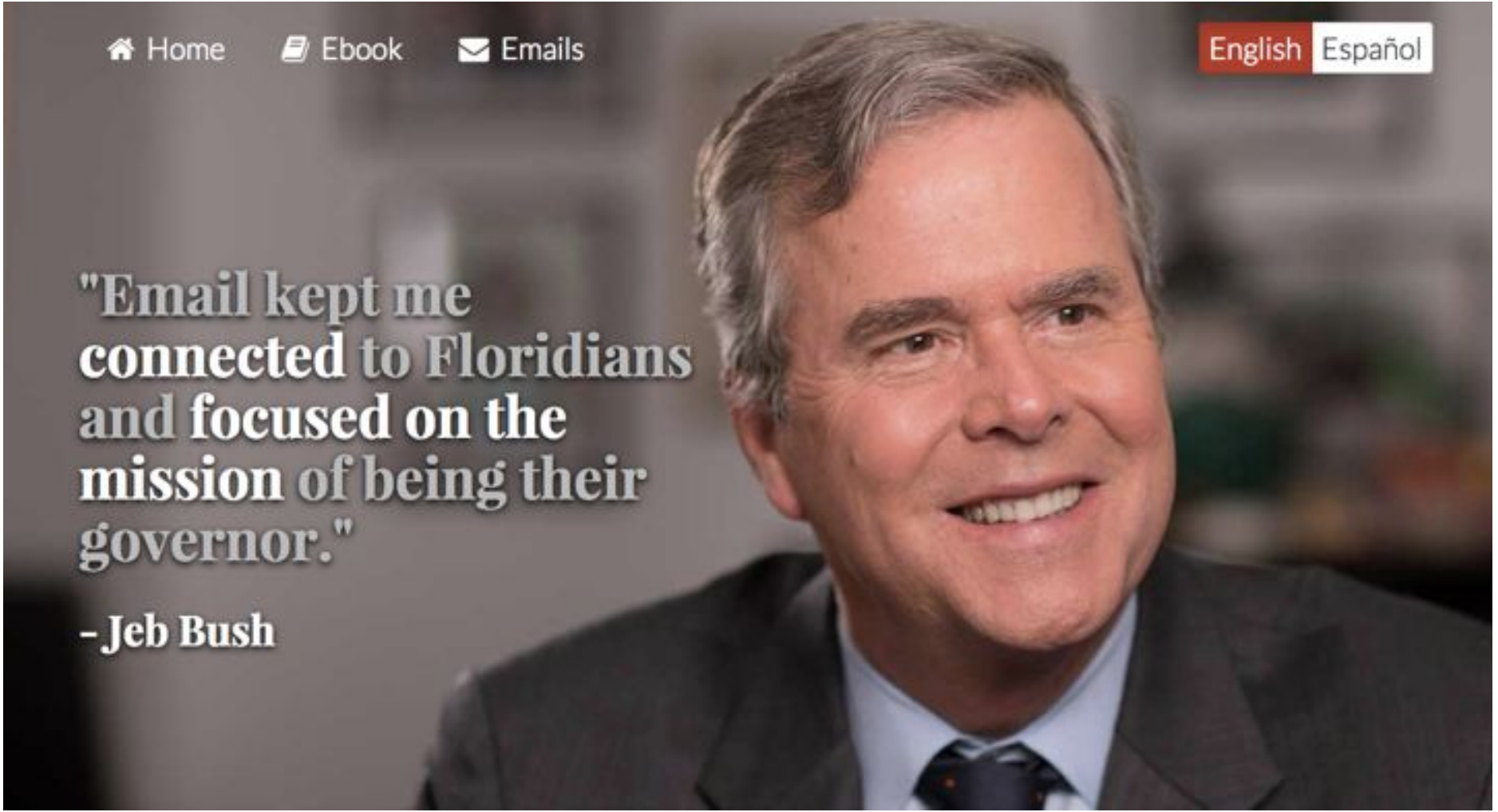
New Email Archive Tool to Sift Literary Legacies

New software developed at Stanford is allowing digital archivists to sort through thousands of emails

Diaries in India inspired software to organize digital archives. *PHOTO: CHAITANYA SHAREEF KULKARNI*

ePADD helps cultural memory institutions meet the challenges of collecting, appraising, processing, and providing access to email

- Screening email for sensitive, restricted, or legally protected info
- Providing browsing access to the intellectual content of a collection
- Enabling research and scholarship


Home
Ebook
Emails
English
Español
"Email kept me connected to Floridians and focused on the mission of being their governor."
- Jeb Bush

In the spirit of transparency,

I am posting the emails of my governorship here. Some are funny; some are serious; some I wrote in frustration. But they're all here so you can read them and make up your own mind.

Take me there

Download the original Outlook (.pst) files here:

JebEmails.part01.rar

JebEmails.part02.rar

JebEmails.part03.rar

JebEmails.part04.rar

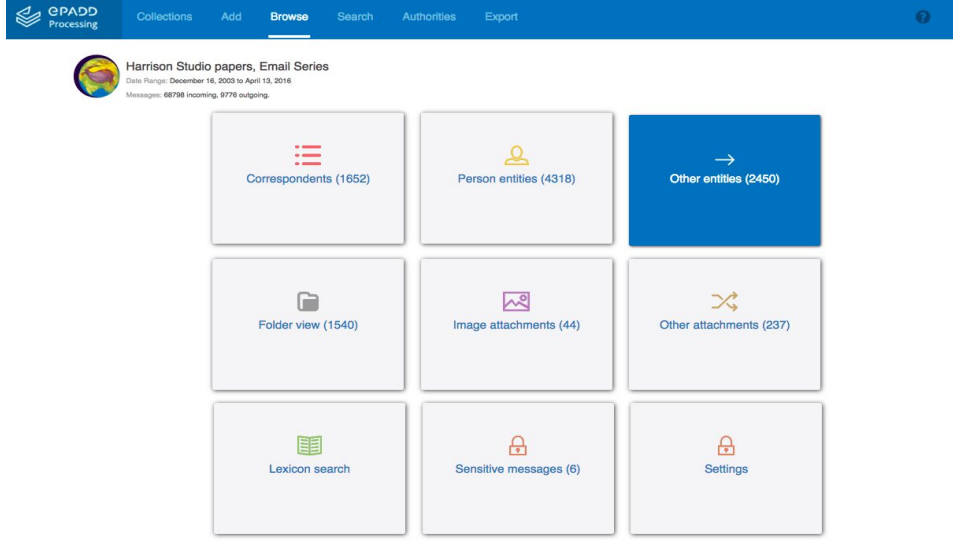
JebEmails.part05.rar

JebEmails.part06.rar

This page previously included raw .PST data files provided by the Florida Department of State. We were informed that some personal information was available in the raw data so we removed these files. Please contact the Florida Department of State with any questions or public records request. You may still read these emails on the email calendar link, where we have redacted personal information we have been able to locate.

Free & Open Source Software

- Incorporates machine learning, automated metadata extraction, and natural language processing
- Supports review, discovery, and access for email



Leads

PROJECT DIRECTOR

Glynn Edwards

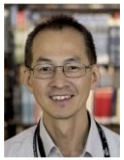
Assistant Director, Department of Special Collections

Digital collections

Manuscripts and archives

Rare books

PROJECT MANAGER



Peter Chan

Digital Archivist

Manuscripts and archives

COMMUNITY MANAGER

Josh Schneider

Assistant University Archivist

Manuscripts and archives

Stanford history

Developers

TECHNICAL ADVISOR

Sudheendra Hangal

Associate Professor of Practice in Computer Science

Ashoka University

sudheendra.hangal@ashoka.edu.in

SOFTWARE ENGINEER

Chinmay Narayanan

Assistant Professor of Computer Science

Ashoka University

chinamy.narayan@ashoka.edu.in

Partners

Harvard University

METRO

University of California, Irvine

University of Illinois - UC

Funders

INSTITUTE of Museum and Library SERVICES

NATIONAL ARCHIVES

NATIONAL HISTORICAL PUBLICATIONS & RECORDS COMMISSION

STANFORD UNIVERSITY LIBRARIES

User Community

British Library

Brown University

California Inst. of Technology

Canadian Centre for Architecture

Center for Jewish History

Columbia University

Duke University

Emory University

Fordham University

Getty Research Institute

Harvard University

Indiana University - PUI

Mass. Inst. of Technology

Museum of Modern Art

National Library of NZ

New York Philharmonic

New York University

Princeton University

Rockefeller Archive Center

Royal Lib of Copenhagen

Smith College

Stanford University

Tufts University

University of California, Berkeley

University of California, Irvine

University of California, LA

University of Copenhagen

University of Illinois – UC

University of Minnesota

University of Southern California

University of Texas at Austin

University of Virginia

University of Warwick

Wildlife Conservation Society

ePADD



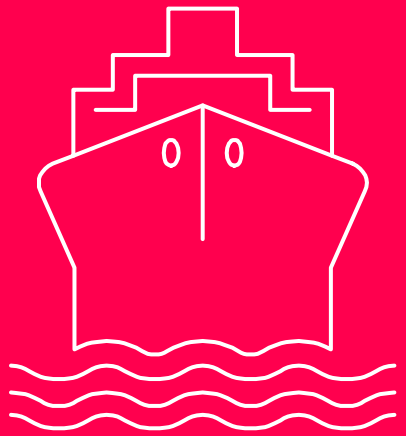




Software Architecture

ePADD Demo

What’s Next



I. Transfer Emails to Institutions

- Download messages from to IMAP/MAPI/Other servers.
- Copy messages from Outlook accounts by donors to designated Outlook accounts set up by collecting institutions.
- Journal email messages.
- Export messages from email clients.
- Archive web pages referred in URLs included in messages.
- Capture files specified in links to cloud storage included in messages to local storage.
- Identify and list encrypted messages for donors to decrypt before transfer.
- Encrypt and transfer the files containing email messages with checksums.

II. Normalize email files for preservation

- Mbox –
- EML –
- XML (compliant with EXAS schema) –

III. Handle Attachments / Embedded links

- List all attachments in one place for easy browsing of files.
- Unzip all zip, 7z, rar files.
- Export attachments for use in other software such as QuickView Plus.
- Enable bulk browsing and preview of image attachments.
- Allow plain text preview of document attachments.
- Automatically search Wayback Machine when users attempt to access dead urls in email messages.
- Run facial recognition on image attachments, and provide an interface to assign names to the faces.
- Run object detection on image attachments, to classify objects into categories (dog, table, etc.).
- Run automated audio transcription on audio attachments.
- Run object detection and automated audio transcription on video attachments.

IV. Detect and Manage Irregular Email Messages

- Duplicated messages.
- Missing headers (to/from/date) in messages.
- Invalid domain name in email addresses.)
- Missing attachment(s) in the message.
- Corrupted attachments.
- Encrypted messages.
- Other issues.

V. Identify and Handle Sensitive Information

- Supply predefined regular expressions related to sensitive information.
- Provide metadata and full text search on headers, body, and attachments.
- Supply predefined keywords to search for potentially sensitive messages.
- Supply predefined entities from common knowledge base for screening potential sensitive messages.
- Ability to embargo messages until a specific date, or until a fixed period of time has transpired.
- Ability to redact part of a message temporarily / permanently.
- Ability to enforce different access terms on messages and their attachments.
- Ability to apply a restriction to a single message, or a group of messages meeting certain criteria.
- Incorporate classifier trained to recognize email that is sensitive according to predefined criteria.

VI. Search Functions (body and / or attachments)

- Search on header fields - to, from, cc, bcc, date, subject line.
- Simple full text search - body/attachments.
- Advanced full text search - combine with entities, lexicon, labels, etc.
- Lexicon search - search terms according to themes defined by users.
- Entity search - search the email collection for matching entities in a block of text.
- Term search - search the email collection for matching terms in a list.
- Filter Function - ability to separate groups of messages for further browsing and searching.
- Cross-collection metadata search and/or browsing.

VI. Search Functions (continue)

- Control on stemming for search terms.
- Keyword suggestion (similar, broader, and narrower) based on Wordnet.
- Keyword suggestion based on Word2Vec.
- Fuzzy search - users control how many misspelled characters are allowed in search terms.
- Natural language search - search using regular spoken language, such as English.
- Image / video search - person, objects, etc.
- Audio search (audio / video files) - transcripted text.

VII. Classify / Group Contents Automatically

- Resolve email addresses associated with an individual to a single correspondent.
- Resolve variant names (e.g. Bill and William) for all Person Entities.
- Extract common entity types (person, organization, location).
- Extract fine-grained entity types such as events and museums.

VIII. Label / Annotate Functions

- Message-based labels and annotations.
- Role-based labels and annotations.
- Part of a message based labels.
- Ability to easily assign a label to a set of message meeting certain criteria.

IX. Provide Access to Email Online/Reading Room

- Assign accession, collection, and institutional metadata to email collections, according to archival practice.
- EML –
- XML (compliant with EXAS schema) –

X. Integrate with Digital Preservation Repositories

- Export email archives in BagIt format for preparation of AIP (Archival Information Package).
- Import DIP (Dissemination Information Package) which contains the normalized attachments for discovery & access.
- Record and export PREMIS events for preservation purposes.

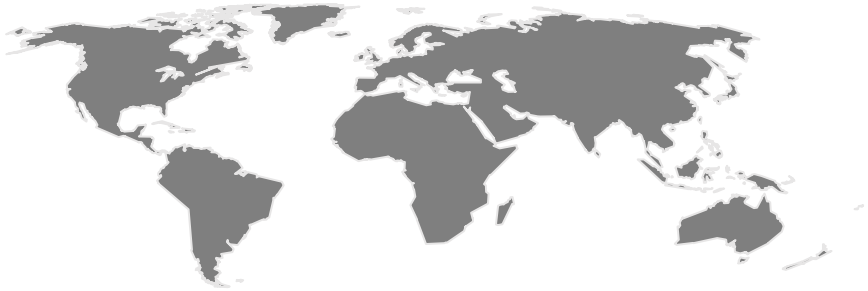
XI. Work with Other Systems

- Provide support for users to confirm correspondents using external authority files (e.g. FAST, VIAF, ISNI).
- Provide support for users to confirm correspondents using local authority files.
- Export headers in CSV format for social network analysis.
- Export correspondents to authority files / Archival Management Systems (e.g. SNAC, ArchiveSpace).
- Export entities in RDF for linked data systems.
- Export metadata to integrated library systems to create catalog record.
- Integrate human resources system with email server to archive email accounts according to institution policy.
- Create Wikidata items to represent the collections and selected correspondents / entities.

How can you participate?

Our User Community

Growing our User Community



British Library
Brown University
California Inst. Of Tech
Center for Jewish History
Columbia University
Duke University
Fordham University
Getty Research Institute
Harvard University
Indiana University - PUI

NY Museum of Modern Art
National Library of NZ
New York Philharmonic
New York University
Princeton University
Rockefeller Archive Center
Royal Lib of Copenhagen
Smith College
Stanford University

University of California, Berkeley
University of California, Irvine
University of California, Los Angeles
University of Copenhagen, Denmark
University of Illinois - UI
University of Minnesota
University of Southern California
University of Virginia
Wildlife Conservation Society

Features Business Explore Marketplace Pricing

ePADD / epadd

Code Issues 46 Pull requests 1 Projects 1 Insigh

Releases Tags

Pinned Tweet

ePADD @e_padd · 22h

ePADD 7.0 released to Github! New support for email collections with multiple owners, improvements to factoring of correspondent names, improved loading times, & more. Thanks to @US_IMLS @mnylc @HarvardArchives @UIarchives @UCI_archives, and @StanfordLibs!

ePADD/epadd

ePADD is a software package developed by Stanford University's Special Collections & University Archives that supports archival processes around the appraisal,

github.com

7 13

Latest release

v7.0

33fed19

ePADD 7.0

Edit

hangal released this 22 hours ago

Assets 6

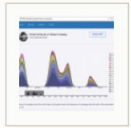
epadd-discovery-standalone.jar	273 MB
epadd-standalone.jar	275 MB
epadd.dmg	348 MB
epadd.exe	276 MB

ePADD

ePADD is a software package developed by Stanford University's Special Collections & University Archives that supports archival processes around the appraisal, ingest, processing, discovery, and delivery of email archives.

Visit the Discovery Module for Stanford University's Special Collections & University Archives to see ePADD in action.



Tweets by @e_padd

ePADD Retweeted

DPC Chatter
@dpc_chat
Getting ready to join us on Thursday for #DPA2018, we have our finalists for the Award for Research and Innovation, sponsored by @SoftwareSaved ...
In no particular order:@StanfordLibs and @e_padd: dpconline.org/events/digital… #WDPD2018

Nov 26, 2018

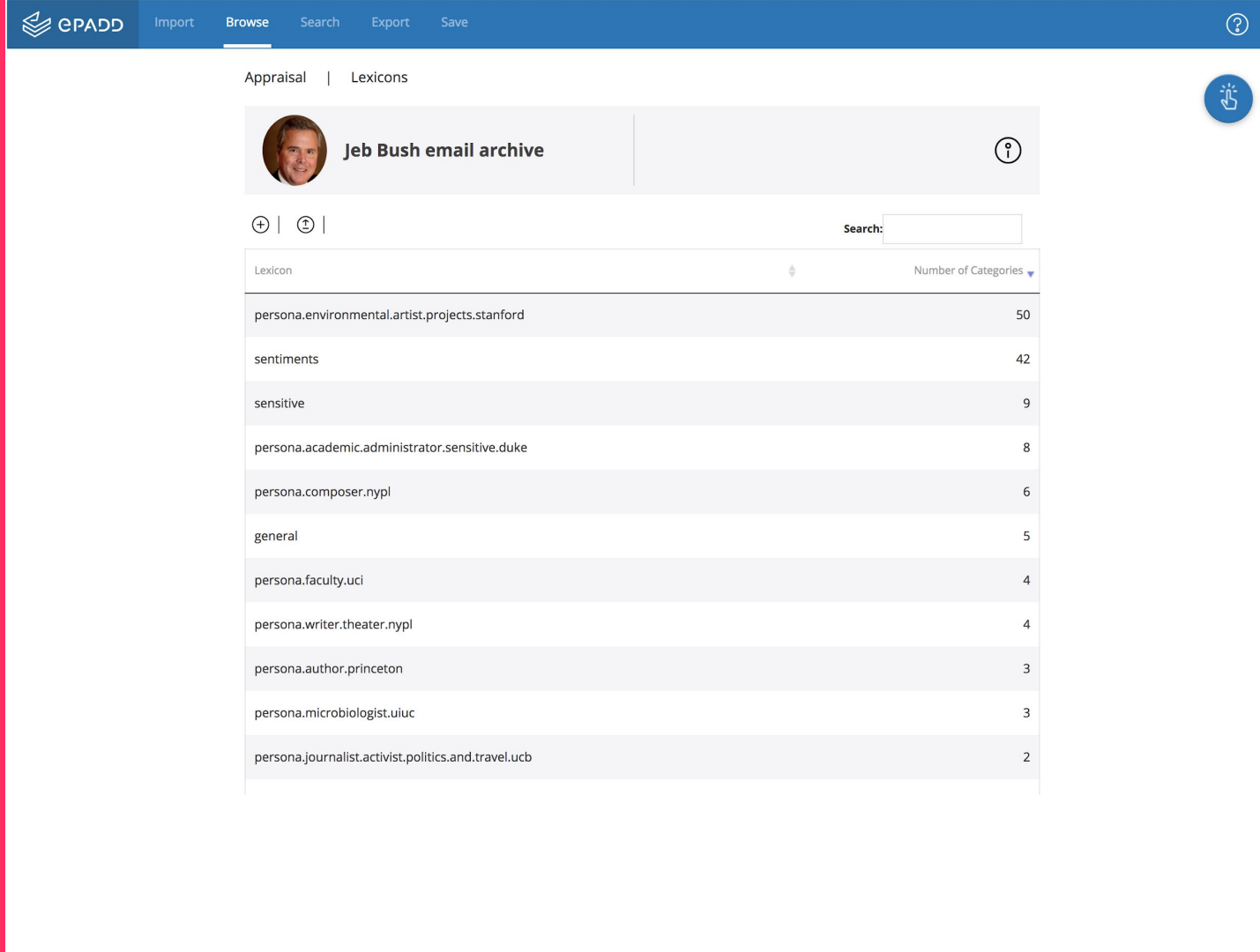
ePADD Retweeted

Netwerk Digitaal Erfgoed
@DigitaalErfgoed
Nog maar een paar plekken beschikbaar voor de @e_padd workshop op 29 nov in @AmsterdamMuseum! Georganiseerd voor #WDPD2018 door #DPC en het Netwerk Digitaal Erfgoed. Neem je laptop mee, je instructeur is #DPAwards18 Finalist Josh Schneider van@StanfordLibsbit.ly/2OxA8J3

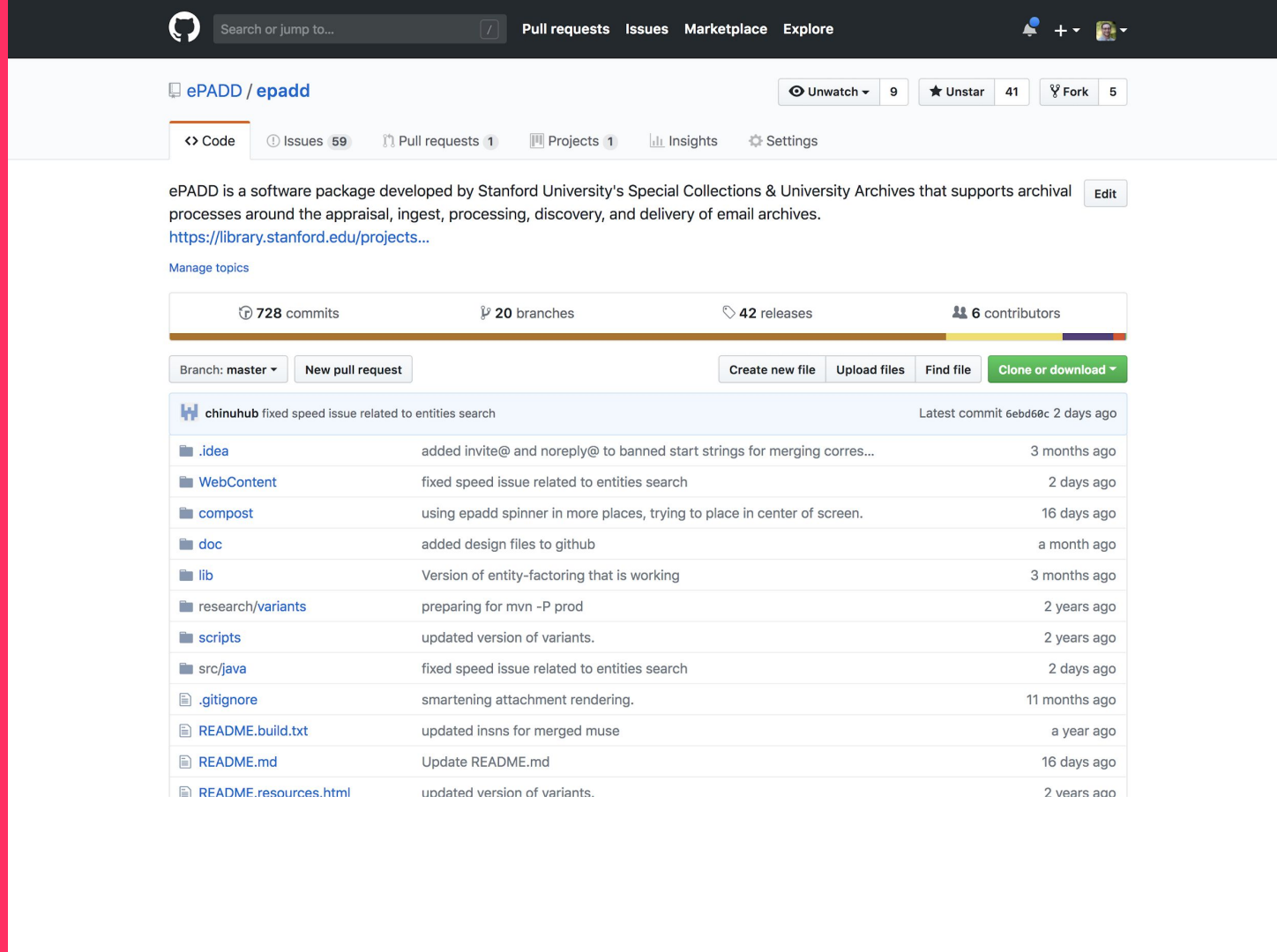
Workshop Email Archiving, Using ePADD...
29-11-2018 @ 10:00 - 12:00 - More celebration on November 29 - World Digital
netwerkdigitaalerfgoed.nl

Embed View on Twitter

Working Groups



Code Contribution



Cross-project Pollination

Thanks!

Visit **library.stanford.edu/projects/epadd**

Follow **@e_padd**

Watch **youtu.be/vu1Oi8TiGiU**

Receive **epadd_list@stanford.lists.edu**

Download / Contribute **github.com/epadd**

Participate **epadd.nimeyo.com**

Reach **epadd_project@stanford.edu**

Wikidata / Wikibase

What is Wikidata?

- Wikidata, launched in 2012, is a collaboratively edited knowledge base hosted by the Wikimedia Foundation.
- It is intended to provide a common source of open data which can be used by Wikimedia projects such as Wikipedia, and by anyone else, under a public domain license.
- Factual claims are stored as statements
  - Subject - predicate - object
  - Item - property - value (e.g. DOOM, video game)

Wikimedia Foundation

- The nonprofit that hosts Wikipedia, Wikidata, and others
- Created MediaWiki which is used in Wikipedia and Wikibase which is used in Wikidata
- Donations and contributions US$98 million for 07/2017 - 06/2018
- 300 staff and contractors
- 200,000 volunteer editors

Wikipedia

- A free encyclopedia, written collaboratively by the people who use it. Anyone can edit almost every page.
- 5th most popular websites in the world (as of May 16, 2018)
- More than 5.7 million English, 2.2 million German and 1.1 million Japanese articles (Dec. 14, 2018)
- Encyclopædia Britannica - 120,000 articles
- “One posts their misinformation, someone corrects it and the first author posts his points right back.”

Wikipedia - Demographics of Hong Kong

Historical population		
Year	Pop.	±% p.a.
1841	7,450	—
1851	32,983	+16.04%
1861	119,320	+13.72%
1871	120,124	+0.07%
1881	160,402	+2.93%
1891	221,441	+3.28%
1901	368,987	+5.24%
1911	456,739	+2.16%
1921	625,166	+3.19%
1931	840,473	+3.00%
1941	1,640,000	+6.91%
1945	600,000	−22.23%
1951	2,070,000	+22.92%
1961	3,128,648	+4.22%
1971	3,936,630	+2.32%
1981	5,183,400	+2.79%
1991	5,752,000	+1.05%
2001	6,714,300	+1.56%
2011	7,071,600	+0.52%
2017	7,409,800	+0.78%

Sources: Census and Statistics Department,[6][7][8][9] CICRED,[10] Office for National Statistics[11]

維基百科 - 香港人口

歷回普查地區別香港人口 (1931-2011年)[10][11][12][13][14]

地區	1931年		1961年		1971年		1981年		1991年		2001年		2011年	
	人數	%	人數	%	人數	%	人數	%	人數	%	人數	%	人數	%
新界	98,157	11.7	409,945	13.1	665,700	16.9	1,303,005	26.1	2,374,818	41.9	3,343,046	49.8	3,691,093	52.2
九龍	263,020	31.3	1,578,026	50.4	2,194,853	55.8	2,450,187	49.1	2,030,683	35.8	2,023,979	30.2	2,108,419	29.8
香港島	409,203	48.7	1,004,875	32.1	996,183	25.3	1,183,621	23.7	1,250,993	22.0	1,335,469	19.9	1,270,876	18.0
水域	70,093	8.3	136,802	4.4	79,894	2.0	49,747	1.0	17,620	0.3	5,895	0.1	1,188	0.0
總計	**840,473**		**3,129,648**		**3,936,630**		**4,986,560**		**5,674,114**		**6,708,389**		**7,071,576**	

Wikidata

- 53 million items (04/30/2018)
- Page views by country in 2017: 8.08M Germany; 5M USA; 4.1M Russia
- 2017: a Wikidata turning point. Wikidata used by
  - Google Knowledge Graph
  - Digital assistants: Siri, Alexa
  - Infoboxes on Wikipedia

美利堅合眾國
United States of America（英語）


國旗


國徽

格言：
「我們相信上帝」
"In God We Trust"[1]

其他格言 顯示▼

國歌：《星條旗》
The Star-Spangled Banner

國家進行曲：《星條旗頌》[2]
The Stars and Stripes Forever


美國於北美洲的位置


美國本土及其海外領地

首都	華盛頓特區 38°53′N 77°01′W
最大城市	紐約市（都市人口） 40°43′N 74°00′W
官方語言	無規定（聯邦政府）[a]
國家語言	英語[b]
族群	77.1% 白人 13.3% 黑人 5.6% 亞洲人 2.6% 混血兒 1.2% 印第安人 0.2% 太平洋島民[3]

United States of America


Flag


Great Seal

Motto:
"In God We Trust"[1][a]
Other traditional mottos [show]

Anthem:
"The Star-Spangled Banner"

March:
"The Stars and Stripes Forever"[2][3]


The contiguous United States with Alaska and Hawaii


The United States including its territories

Capital	Washington, D.C. 38°53′N 77°01′W
Largest city	New York City 40°43′N 74°00′W
Official languages	None at federal level[b]
National language	English[c]
Ethnic groups (2016)[6]	**By race:** 77.1% White 13.3% Black 5.6% Asian 2.6% Other/multiracial 1.2% Native 0.2% Pacific Islander **Ethnicity:** 17.6% Hispanic or Latino 82.4% non-Hispanic or Latino

The rise of Wikidata as a linked data source

Linked data source	2018 Rank	2015 Rank
id.loc.gov	1	3
VIAF (Virtual International Authority File)	2	1
DBpedia	3	2
GeoNames	4	3
Wikidata	5	15

http://hangingtogether.org/?p=6775

Wikidata examples

- Nintendo DS https://www.wikidata.org/wiki/Q170323

- Doom https://www.wikidata.org/wiki/Q189784

Quantitative analysis - SPARQL endpoint

- Create your own query
- Modify example (Number of films by year and genre) to show video game (Q7889) information
- Change from Scatter chart to Table

Wikidata - Users

- National Library of Wales
  - https://blog.wikimedia.org/2016/11/05/wikidata-visiting-scholar-art-dataset/
- The Smithsonian
  - https://confluence.si.edu/display/LODPP/Smithsonian+Open+Data+Pilot
- Europeana
  - https://pro.europeana.eu/page/get-your-vocabularies-in-wikidata
- Yale / BnF / Open Preservation Foundation
  - (https://ipres2017.jp/wp-content/uploads/7.pdf)

#	Item	Label	Label	Label	Label	Label
1	trivia video game Q60617948	de Wissensspiel	ja トリビアゲーム	fr	zh 琐事电子游戏	ko 상식 퀴즈 게임
2	vehicular combat game Q2070892	de Fahrzeugskampfspiel	ja 車両戦闘ゲーム	fr jeu de combat motorisé	zh 车辆格斗电子游戏	ko 차량 전투 게임
3	science fiction video game Q27670585	de Science Fiction Videospiel	ja SFゲーム	fr jeu vidéo de science fiction	zh 科學幻想電子遊戲	ko 공상과학 게임
4	4X Q603555	de Globalstrategiespiel	ja 4X	fr Jeu 4X	zh 4X概念体系	ko 4X
5	adult video game Q3362070	de Videospiele für Erwachsene	ja 成人向けゲーム	fr jeu vidéo pour adultes	zh 成人电子游戏	ko 성인 게임
6	tile-matching video game Q7802107	de Puzzle-Videospiel	ja タイルマッチングゲーム	fr jeu de correspondance de tuile	zh 消除类游戏	ko 타일 매칭 게임
7	computer wargame Q2454898	de Kriegsspiel	ja 戦争ゲーム	fr jeu de guerre sur ordinateur	zh 计算机战棋	ko 컴퓨터 전쟁게임
8	adventure game Q23916	de Adventure	ja アドベンチャーゲーム	fr jeu d'aventure	zh 冒险游戏	ko 어드벤처 게임
9	construction and management Q1036289	de Aufbaustrategiespiel	ja ミニスケープ	fr jeu de gestion	zh 建造与经营模拟游戏	ko 건설경영 시뮬레이션 게임

Issues in Wikidata

- Data model - properties decided by Wikidata
- Ensure properties listed in Wikidata behave according to your expectation - e.g. broad match (Q39894595)
- Data can be edited by anyone
- All data publish as public domain CC0 (public domain)

What is Wikibase

- Wikibase is the software that enables MediaWiki to store structured data or access data that is stored in a structured data repository.

Wikibase

- Address the following issues
  - Control on who can edit information
  - Implement data model best fit for your need (your own interpretation of work, expression, manifestation, etc.)
  - Contribute to LOD - Persistent URL
  - Quantitative analysis - SPARQL endpoint

Wikibase Issues

- If institutions are not using the same Wikibase, how can they synchronize among different incidences of Wikibase hosted by different institutions?
- Resource to host Wikibase instance
- Understand the properties listed in Wikibase
- Know how to install, maintain the software

Wikibase - Users

- OCLC (controlled vocabularies)
  - https://www.oclc.org/research/themes/data-science/linkeddata/linked-data-prototype.html
- Rhizome (modeling for preservation of digital art)
  - https://wikimediafoundation.org/2018/09/06/rhizome-wikibase/
- German National Library (controlled vocabularies)
  - https://wiki.dnb.de/display/GND/Authority+Control+meets+Wikibase

Rhizome - early Wikibase user

- In the digital arts field, we deal with pretty specialized performance information that the world at large is probably not interested in, or the community hasn't come to an agreement how to describe it.
- Licensing restrictions of Wikidata and Commons prevent certain information to be stored there: for instance, reference information about software would in many cases be contained in screenshots, which for Rhizome's purposes is not permitted on Wikidata and Commons.

Federated Wikibase Instances

- In digital art, artists have sometimes deliberately strayed away from standards, or have exploited very specific versions of software and file formats. Here we see a large need for federation [Ed. note: meaning individual but interconnected databases]: many different Wikibases, used by individual organizations, containing specialized data, while all pointing to the same Wikidata items, describing these items from the perspective of their own specialization.

Robert Creeley Example

Robert Creeley (Q918620)

American poet
Bob Creeley | Creeley | Robert White Creely

edit

In more languages Configure

Language	Label	Description	Also known as
English	Robert Creeley	American poet	Bob Creeley Creeley Robert White Creely
Korean	No label defined	No description defined	
Chinese	罗伯特·克里利	No description defined	
Japanese	No label defined	No description defined	

All entered languages

archives at

Stanford University Libraries — edit

1 reference

Stanford University Libraries, Department of Special Collections & University Archives — edit

0 references

+ add reference

+ add value

Robert Creeley papers, 1950-2005 (Q60549450)

The papers are divided into 15 series. edit

In more languages Configure

Language	Label	Description	Also known as
English	Robert Creeley papers, 1950-2005	The papers are divided into 15 series.	
Korean	No label defined	No description defined	
Chinese	No label defined	No description defined	
Japanese	No label defined	No description defined	

Statements

owned by	Stanford University Libraries, Department of Special Collections & University Archives 1 reference	edit + add value
has part	Robert Creeley Emails 0 references	edit + add reference + add value

Robert Creeley Emails (Q60677840)

email messages of Robert Creeley archived at Stanford University Libraries edit

addressee	Charles Bernstein 0 references + add reference	edit
	Bruce Jackson 0 references + add reference	edit
	Robert Creeley 0 references + add reference	edit
		+ add value

main subject	Saddam Hussein	edit
	1 reference	
	George W. Bush	edit
	0 references	
		+ add reference
	Iraq	edit
	0 references	
		+ add reference
	Brown University	edit
	0 references	
		+ add reference
	al-Qaeda	edit
	0 references	
		+ add reference
	Iraq War	edit
	0 references	
		+ add reference
	Vietnam War	edit
	0 references	
		+ add reference
	World Trade Center	edit
	0 references	
		+ add reference
		+ add value

Publications / Resources

The Personal Archives Accessible in Digital Media (paradigm) project 2005-2007

The Personal Archives Accessible in Digital Media (paradigm) project saw the major research libraries of the Universities of Oxford and Manchester come together to explore the issues involved in preserving digital private papers through gaining practical experience in accessioning and ingesting digital private papers into digital repositories, and processing these in line with archival and digital preservation requirements.

http://www.paradigm.ac.uk/index.html

Digital Forensics and Born-Digital Content in Cultural Heritage Collections 2010

This report introduces the field of digital forensics in the cultural heritage sector and explores some points of convergence between the interests of those charged with collecting and maintaining born-digital cultural heritage materials and those charged with collecting and maintaining legal evidence.

https://www.clir.org/pubs/reports/pub149/

An Inter-Institutional Model for Stewardship (AIMS) 2010-2011

The AIMS project evolved around a common need among the project partners — and most libraries and archives — to identify a methodology or continuous framework for stewarding born-digital archival materials. The AIMS Framework was developed to define good practice in terms of archival tasks and objectives necessary for success. The Framework presents a practical approach but also a recognition that there is no single solution for many of the issues that institutions face when dealing with born-digital collections. Instead, the AIMS project partners developed this framework as a further step towards best practice for the profession.

https://dcs.library.virginia.edu/aims/white-paper/

Born Digital: Guidance for Donors, Dealers, and Archival Repositories 2013

This report offers recommendations to help ensure the physical and intellectual well-being of born-digital materials transferred from donors to archival repositories. The report surveys the primary issues and concerns related to born-digital acquisitions and is intended for a broad audience with varying levels of interest and expertise, including donors, dealers, and repository staff.

https://www.clir.org/pubs/reports/pub159/

Levels of Digital Preservation 2013

The “Levels of Digital Preservation” (PDF; Proceedings of the Archiving (IS&T) Conference, April 2013, Washington, DC) are a tiered set of recommendations for how organizations should begin to build or enhance their digital preservation activities.

https://ndsa.org//activities/levels-of-digital-preservation/

Emulation & Virtualization as Preservation Strategies 2015

In this report commissioned by the Foundation, David Rosenthal describes current technology frameworks for emulation and virtualization, and outlines the issues and challenges in deploying these technologies to preserve both digital artefacts from the past and current digital material that will age into legacy status.

https://mellon.org/resources/news/articles/emulation-virtualization-preservation-strategies/

Digital Preservation Handbook 2015

The Handbook provides an internationally authoritative and practical guide to the subject of managing digital resources over time and the issues in sustaining access to them. It will be of interest to all those involved in the creation and management of digital materials.

https://dpconline.org/handbook

Library of Congress Recommended Formats Statement 2018

Recommended Formats Statement identifies hierarchies of the physical and technical characteristics of creative formats, both analog and digital, which will best meet the needs of all concerned, maximizing the chances for survival and continued accessibility of creative content well into the future.

https://www.loc.gov/preservation/resources/rfs/

The Future of Email Archives 2018

Email is an increasingly important part of the historical record, yet it is particularly difficult to preserve, putting future access to this vast resource at risk. The Future of Email Archives looks at what makes email archiving so complex and describes emerging strategies to meet the challenge.

The report is intended for the archival community, digital preservation professionals, technologists and software developers, commercial vendors, historians and scholars, institutional administrators, and funding agencies and foundations.

https://www.clir.org/pubs/reports/pub175/

Core Trustworthy Data Repositories Extended Guidance 2018

This document consists of the Core Trustworthy Data Repositories Requirements for 2017–2019 with introductory paragraphs on Background & General Guidance, which are set by the CoreTrustSeal Board and remain unchanged for the period 2017–2019. The fixed text is recognizable by the boxes drawn around it.

https://www.coretrustseal.org/wp-content/uploads/2017/01/20180629-CTS-Extended-Guidance-v1.1.pdf

A Preservationist's Guide to the DMCA Exemption for Software Preservation 2018

The Library of Congress recently adopted several exemptions to the Digital Millennium Copyright Act (DMCA) provision prohibiting circumvention of technological measures that control access to copyrighted works. The exemptions went into effect on October 28, 2018 and last until October 28th, 2021. This guide is intended to help preservationists determine whether their activities fall under the new exemption.

http://www.softwarepreservationnetwork.org/1201-exemption-guide-for-software-preservationists/

National Digital Stewardship Alliance

The National Digital Stewardship Alliance is a consortium of more than 220 partnering organizations, including universities, professional associations, businesses, government agencies, and nonprofit organizations, all committed to the long-term preservation of digital information. Members work together to preserve access to our national digital heritage.

https://ndsa.org//

Digital Preservation Coalition

The Digital Preservation Coalition (DPC) is a UK-based non-profit limited company which seeks to secure the preservation of digital resources in the UK and internationally to secure the global digital memory and knowledge base. (Wikipedia)

https://www.dpconline.org/

Digital Archives Specialist (DAS) Curriculum and Certificate Program

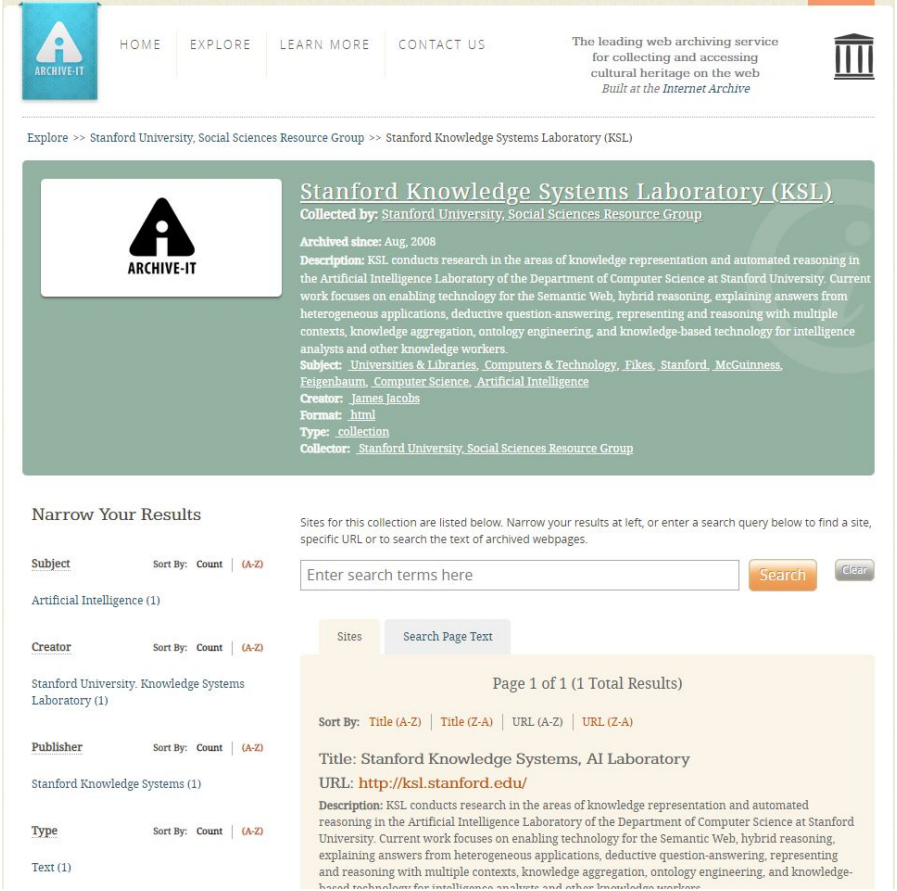
The DAS curriculum is structured in tiers of study that guide you to choose courses based on your specific knowledge, training, and needs. You can choose individual courses—or you can take your learning to the next level by earning a Digital Archives Specialist Certificate from SAA after completing required coursework and passing both course and comprehensive examinations.

https://www2.archivists.org/prof-education/das

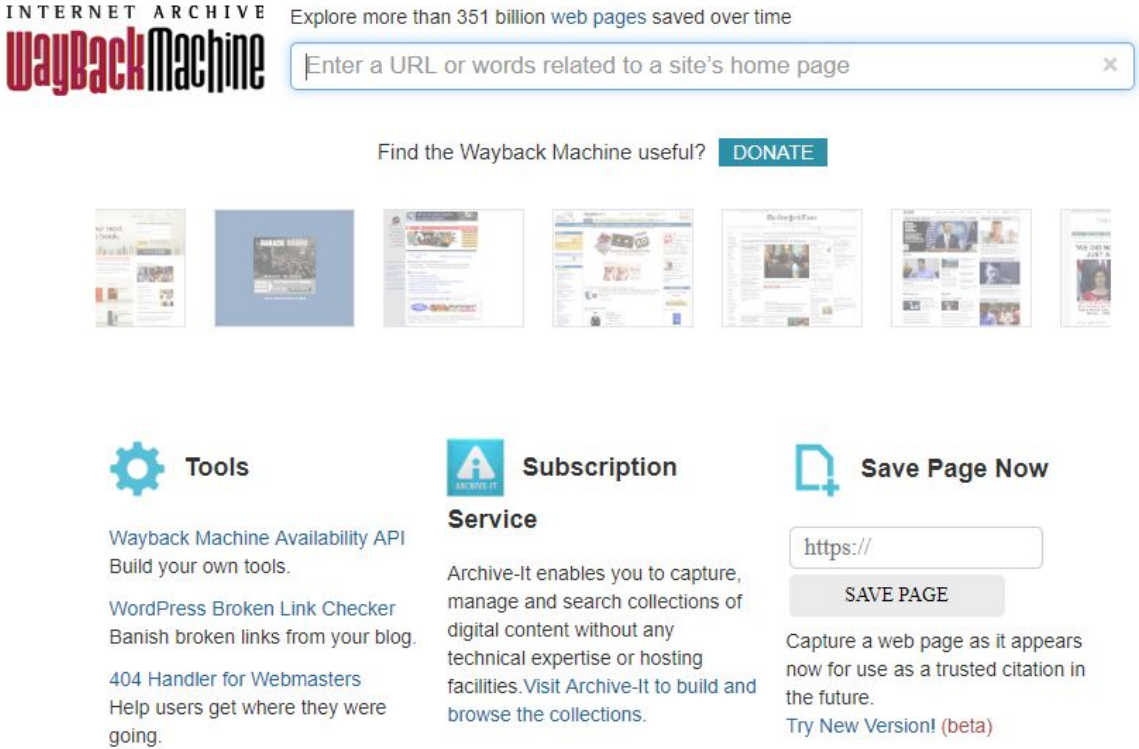
Introduction to Digital Preservation

This guide covers the many faces of digital preservation, including the terms, concepts, models, standards, actions, risks, and tools. Digital preservation is a broad field that encompasses everything from project management to technical skills. ……. This guide contains enough information to provide a general awareness, but it also contains resources for further study.

https://libguides.bodleian.ox.ac.uk/digitalpreservation



Web archive of http://ksl.stanford.edu


INTERNET ARCHIVE
WayBackMachine
Explore more than 351 billion web pages saved over time
Enter a URL or words related to a site's home page
Find the Wayback Machine useful?
DONATE
Tools
Wayback Machine Availability API
Build your own tools.
WordPress Broken Link Checker
Banish broken links from your blog.
404 Handler for Webmasters
Help users get where they were going.
Subscription Service
Archive-It enables you to capture, manage and search collections of digital content without any technical expertise or hosting facilities. Visit Archive-It to build and browse the collections.
Save Page Now
https://
SAVE PAGE
Capture a web page as it appears now for use as a trusted citation in the future.
Try New Version! (beta)

Web archive