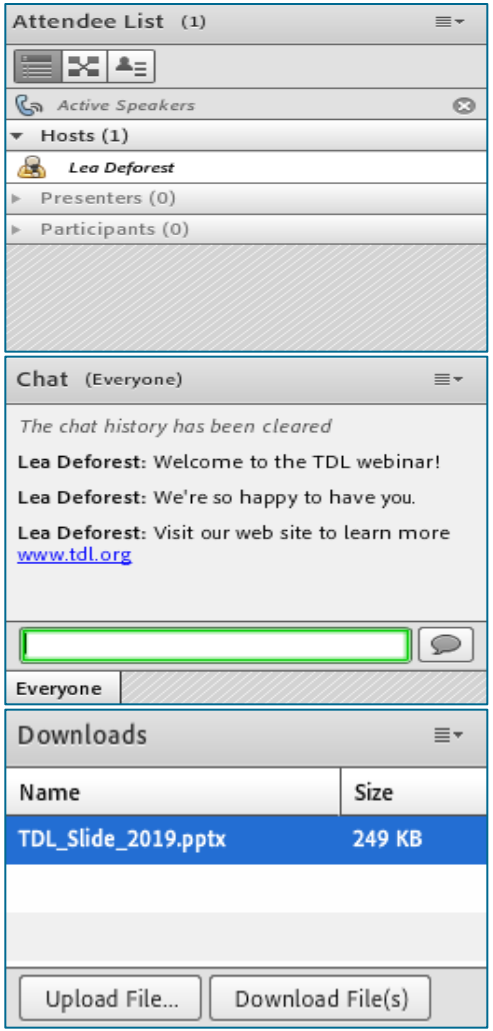
archivematica camp texas

campfire stories

Texas Digital Library



Using Adobe Connect

TOP LEFT | Attendees Pod Hosts, Presenters, Participants

MIDDLE LEFT | Chat Pod Questions, comments, links in “everyone”; direct message individuals

BOTTOM LEFT | Download Pod Files shared by TDL and/or our presenters

campfire stories

featuring lightning talks from:

- Sean Buckner, Texas A&M University
- Julianna Barrera-Gomez, University of Texas San Antonio
- Bethany Scott, University of Houston
- Lauren Goodley, Texas State University

presentation by:

- Sarah Romkey, Archivematica / Artefactual

moderated by Courtney Mumma, Texas Digital Library

Archivematica, the Cornerstone

Groundbreaking, Foundational, Constructive

Groundbreaking, 2015

Digital Preservation program established

Survey and assessment

Archivematica chosen

Gathered tools and resources

Preparations and initial plans made

Credit Wikimedia Commons

LIBRARIES
TEXAS A&M UNIVERSITY

Foundation, 2016

Conceptualized the Work

Documentation – Blueprints

Coordination and communication

Archivematica deployed and tested

Foundation laid

Credit Wikimedia Commons

LIBRARIES
TEXAS A&M UNIVERSITY

Cornerstone, 2017

Archivematica in Production

“Started preserving!”

Developed procedures

DuraCloud@TDL

Credit Wikimedia Commons

LIBRARIES
TEXAS A&M UNIVERSITY

Masonry, 2018

Archivematica Camp - Houston

One AIP at a time

Problems DPNed

Integrated into our infrastructure

Bogdanhoda/Getty Images

LIBRARIES
TEXAS A&M UNIVERSITY

Building Up, 2019

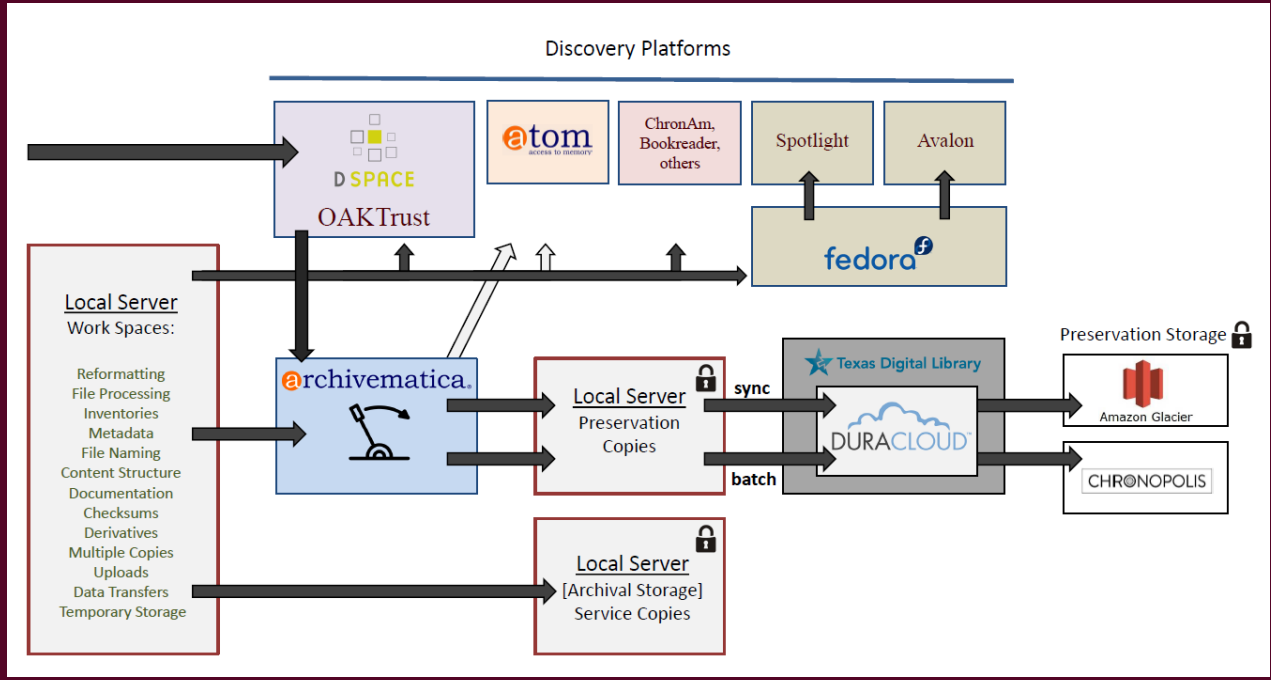
Production

Upgraded

Robust

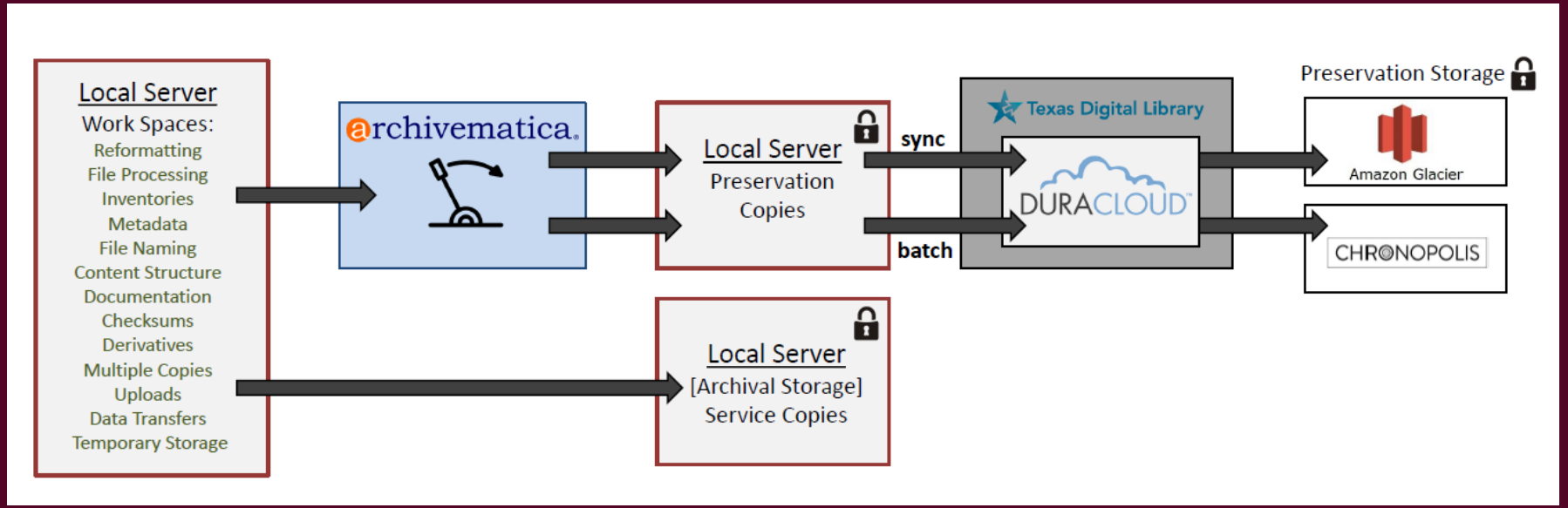
Tiered

Workflows



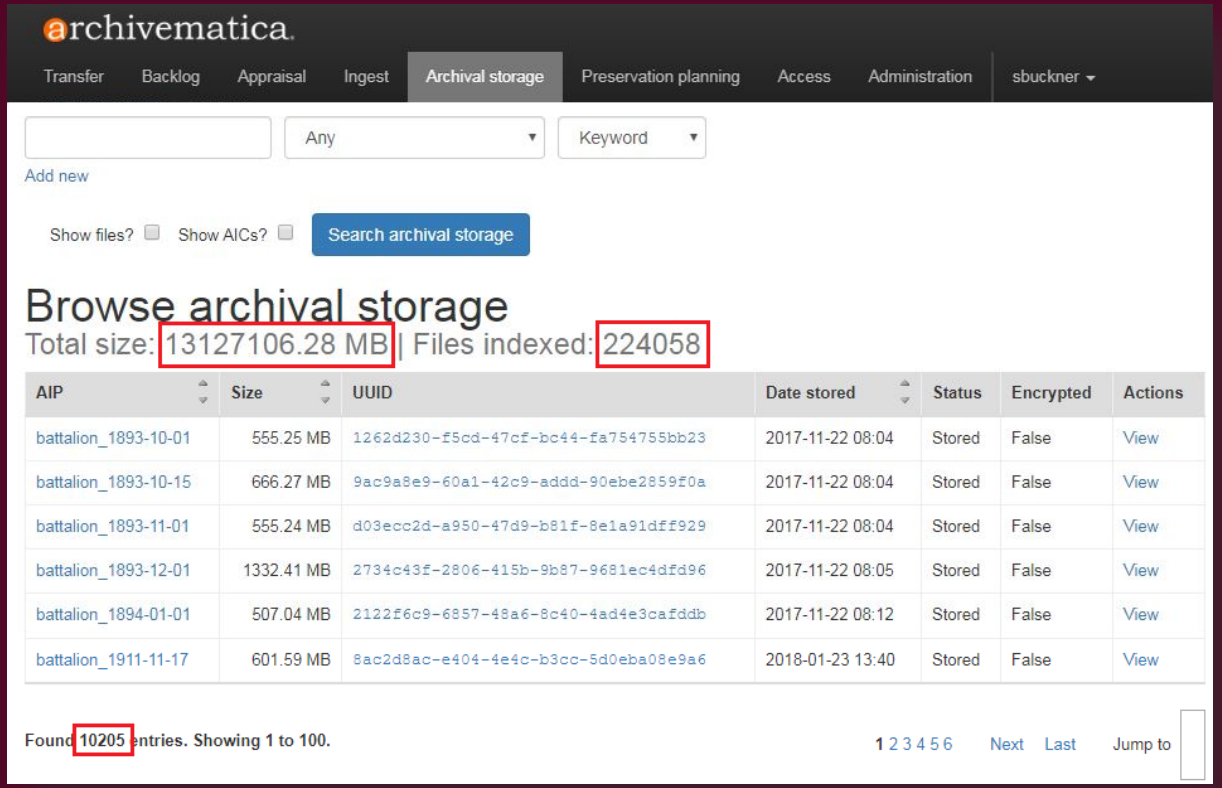
LIBRARIES
TEXAS A&M UNIVERSITY

Archivematica @ TAMU Libraries



LIBRARIES
TEXAS A&M UNIVERSITY

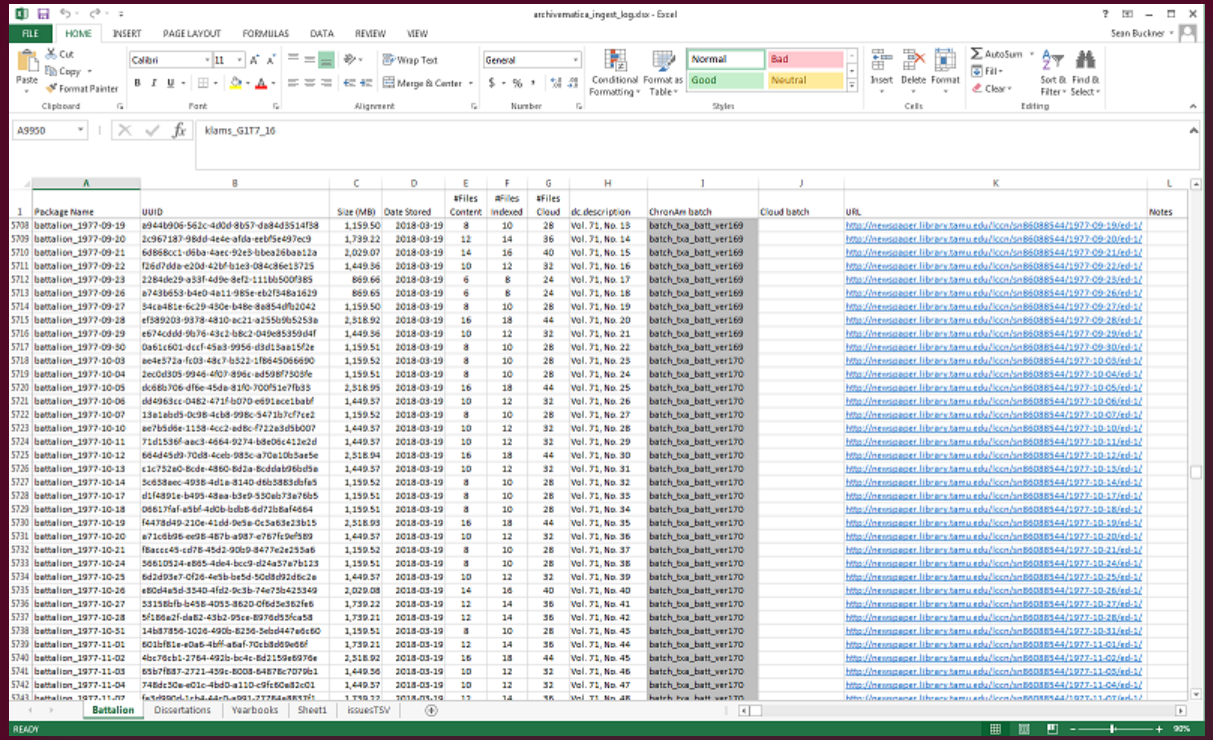
Archivematica @ TAMU Libraries



- Over 13 TB total size (x3)
- Quarter of a million files
- Over 10K AIPs
- Double that in queue
- Triple that in storage

LIBRARIES
TEXAS A&M UNIVERSITY

Archivematica @ TAMU Libraries



External log/inventory

Much preparatory work

No DIPs, more external

Need adjustments

Not enough time!

LIBRARIES
TEXAS A&M UNIVERSITY

Moving Forward, 2020-

Tackle large, legacy collections

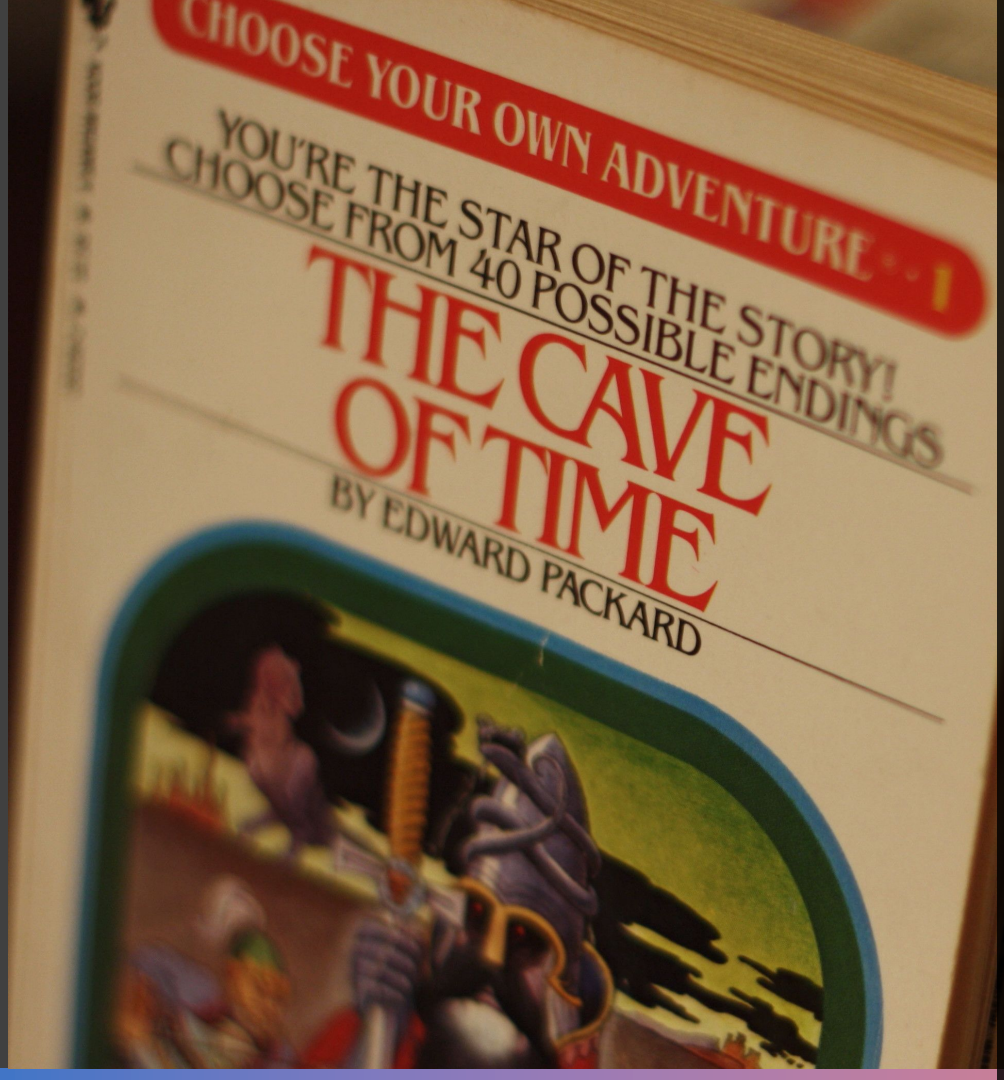
Interoperability, integrated workflows

Fully utilize Cloud options

Codify documentation

Sean Buckner
Clinical Associate Professor
Coordinator of Digital Preservation & Digitization
sbuckner@library.tamu.edu

LIBRARIES
TEXAS A&M UNIVERSITY



Navigating Archivematica

Choose Your Own SIP Adventure!

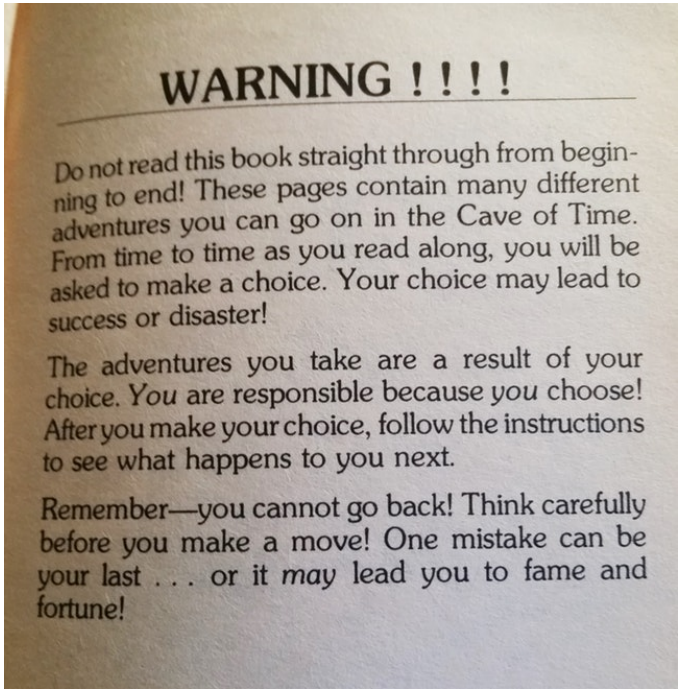
(We totally did this-and had some fun- and you can too!)

Julianna Barrera-Gomez, Digital Archivist

UTSA LIBRARIES
SPECIAL COLLECTIONS

Choices you face . . .

- System setup
  - Out of the box, self-hosted?
  - Turn key?
- What are you processing?
  - Archival collections / collection level?
  - Library items / item level?
- What kind of metadata & documentation do you have?
  - Administrative metadata only?
  - Descriptive metadata? To what level?
- What's your timeline?



WARNING ! ! ! !

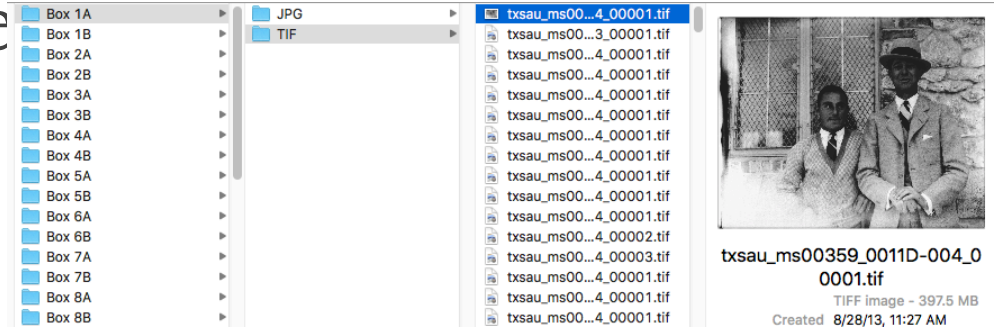
Do not read this book straight through from beginning to end! These pages contain many different adventures you can go on in the Cave of Time. From time to time as you read along, you will be asked to make a choice. Your choice may lead to success or disaster!

The adventures you take are a result of your choice. *You* are responsible because *you* choose! After you make your choice, follow the instructions to see what happens to you next.

Remember—you cannot go back! Think carefully before you make a move! One mistake can be your last . . . or it *may* lead you to fame and fortune!

https://www.bustle.com/articles/142370-why-those-choose-your-own-adventure-books-defined-my-childhood

Our First Adventure

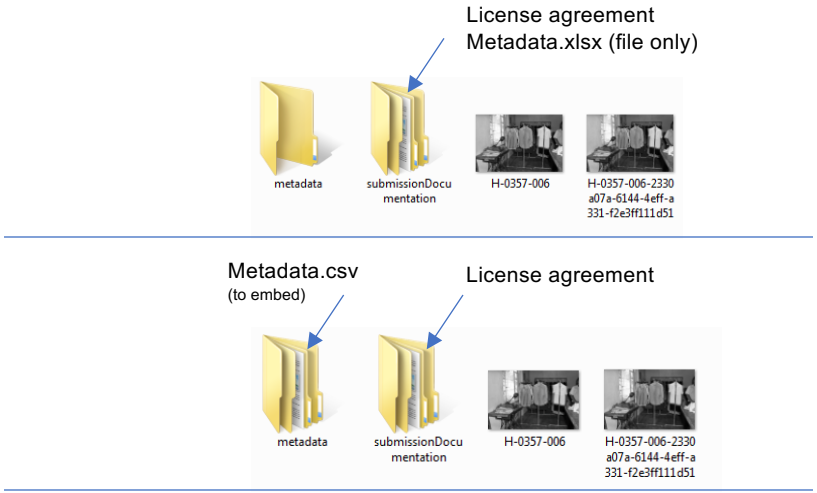


- System setup archivesDIRECT
  - Turn key: ArchivesDirect (hosted) & DuraCloud
- What we've been processing
  - Archival collections / collection level
  - Standard directories (no bags)
- What kind of metadata & subDoc we have
  - Basic descriptive metadata for digitized photographs, at item level, in a big spreadsheet
  - Digitized copies of deeds, other rights documents
- Timeline
  - ASAP (irreplaceable files, network woes) and 1 digital archivist

MS 359: L-3514-A: Lydia Mendoza with guitar, 1948
http://digital.utsa.edu/cdm/ref/collection/p9020coll2/id/375

Batching the SIPs & Adding the Metadata

- Arranging the SIP directory structure
  - What dependencies impact this?
    - Pre-Archivematica SIP workflow & packaging
    - Your CMS?
    - Your DIP structure?
- If you decide to include the metadata…
  - Is it complete & in the right format?
  - Embed it into files?
  - Apply it to directory?
  - Go even more meta & link it as an AIP

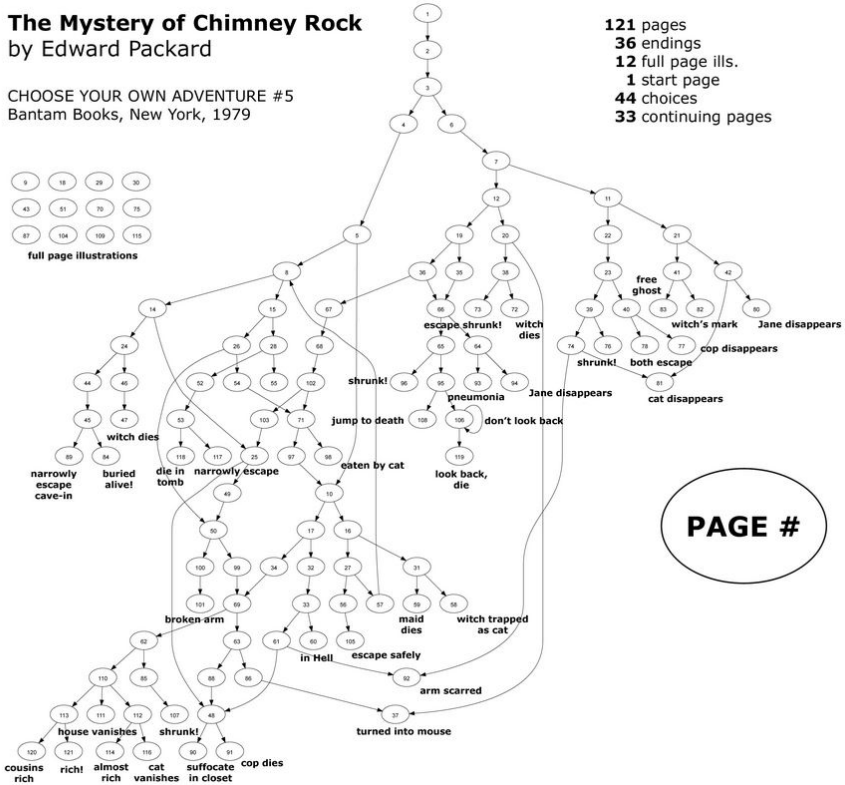


Found 114 entries. Showing 111 to 114.

AIP	Size	UUID	AIC	Files	Date stored
NEDCC_MS353_Belgin_Box26A_b (view raw)	30044.61 MB	6e6334bd-8b14-4e0a-9a86-9bdeaef001ab	Part of AIC#NEDCC	227 files	2017-12-20 09:31
NEDCC_MS353_Belgin_Box26A_a (view raw)	31298.96 MB	49c1b4d7-df44-476e-a24a-b70ceb8b9ba5	Part of AIC#NEDCC	228 files	2019-01-18 15:08
NEDCC_MS353_Belgin_Box26A_a (view raw)	31299.71 MB	acc638f0-1c3f-4352-9a17-972f44eb9cdd	Part of AIC#NEDCC	228 files	2017-12-19 10:57
NEDCC_CollectionsInfoandMetadata (view raw)	4.01 MB	a07a3f97-8d99-457d-be59-3b2a17abdd28	Part of AIC#NEDCC	10 files	2019-01-30 15:30

Running the Microservices & Generating the AIP

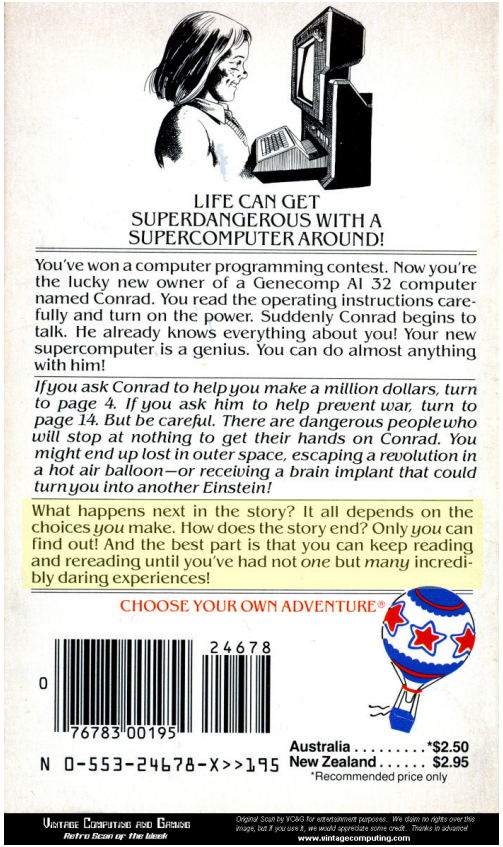
- Lots of decision points
  - Processing configurations
  - Tools to use
  - Storage options
- Find your balance
  - Remember MPLP
  - Remember you have the ability to reprocess
  - …or even start over!



https://www.seanmichaelragan.com/html/[2008-03-07]_Choose_Your_Own_Adventure_book_as_directed_graph.shtml

Success !!!!

- AIPs are stored!
  - 121 AIPs
  - 2.86 TB in DuraCloud
- Bugs have been found!
  - AIC error
- Next adventures await!
  - A/V collections
  - Structured AIPs for born-digital/hybrid collections
- Community has been formed!
  - You aren't alone
  - Any choice is a success
  - You *can* go back and re-adventure!



http://www.vintagecomputing.com/index.php/archives/448/retro-scan-of-the-week-choose-your-own-adventure

Archivematica Campfire Stories

...from University of Houston Libraries!

500000

Bethany Scott, Digital Projects Coordinator

bscott3@uh.edu

UNIVERSITY of HOUSTON | LIBRARIES

Last year at UH...

Preservation workflow now in production for digitized content

Digital objects packaged and transferred individually

~1200 objects (26 TB) in preservation storage

archivematica®

Future work:

- Developing a similar workflow for born-digital archives
- Interface improvements for searching and browsing preservation storage

As of November 2019

Preservation workflow now in production for digitized content

Digital objects packaged and transferred individually

~~~1200 objects (26 TB) in preservation storage~~~

1959 objects (28.77 TB) in preservation storage, 20386 files indexed

archivematica®

Future work:

- Developing a similar workflow for born-digital archives
- Interface improvements for searching and browsing preservation storage

UNIVERSITY of HOUSTON | LIBRARIES
~~~

As of November 2019

Preservation workflow now in production for digitized content

Digital objects packaged and transferred individually

~~~1200 objects (26 TB) in preservation storage~~~

1959 objects (28.77 TB) in preservation storage, 20386 files indexed

Future work:

- ~~Developing a similar workflow for born-digital archives~~
- ~~Interface improvements for searching and browsing preservation storage~~

MIGRATION!

archivematica®

UNIVERSITY of HOUSTON | LIBRARIES
~~~

Migration impacts on preservation

Changing filenaming conventions

Collections with download restrictions

Digital objects in ASpace

AIPs	AIP Name/Collection	Number of Objects
1-116	10604411_southern	116
117-916	1997_006a_kuhtav	800
917-989	2010_009_hiphop	73
990-1101	2010_020_tioh	112
1102-1137	2012_006_samplified	36
1138-1247	2015_008_beste	110
1248-1253	2016_008_eidenav	6
1254-1260	2017_034_bebegow	7
1261	2018_015_lukasek	1
1262-1291	2018_016_nwc	30
1292-1687	22329406_gcam	396
1688-1691	9582873_domestic	4
1692-1722	987443698_hetag	31
1723-1728	aa_201804	6
1729-1738	aa_201809	10
1739-1753	aa_201812	15
1754-1906	gcam_201801	153
1907-1959	gcam_201908	53

Archival collections, possible file renaming

Collections with download restrictions: add service files, possible file renaming

Collections that need to have digital objects combined in ASpace

Takeaways and next steps

Consider all use cases

...and be flexible when rework is needed

Nothing is set in stone!

Up next: completing rework tasks, continue adding migration collections to AM, and audit preservation storage in 2020

ARCHIVEMATICA @TXST

MICROSERVICES AND MICROSTEPS

BACKGROUND

- Digital Preservation Working Group 2014-current
  - 3 (now 4) departments: UA, Wittliff, Digital & Web Services, Acquisitions
  - Digital & Web Services: Department head, Digital Media Specialist, Programmer
  - Stakeholders
- 1. Created Digital Preservation Policy Document
- 2. Request for distributed storage

WHY ARCHIVEMATICA?

- Other options
  - By hand (every microservice)
    - Checksums
    - Excel/csv metadata
    - Extract(ed) metadata
    - Folders
  - Preservica
    - Expense
    - Opacity
  - Spanish company (?)
- Pros
  - Automated
  - Transparent
  - Opt in or out of microservices
- Cons
  - Ubuntu
  - Labor: decisions and prep, different for each institution

SET UP: DECISION POINTS

WHAT IS AN ITEM?

For still image and audio – a collection

For video-a file

METADATA

How much at item level

How much at collection level

UA / WITTLIFF / IR

Different!

-Archivematica instance/server

-AIP storage directory on server

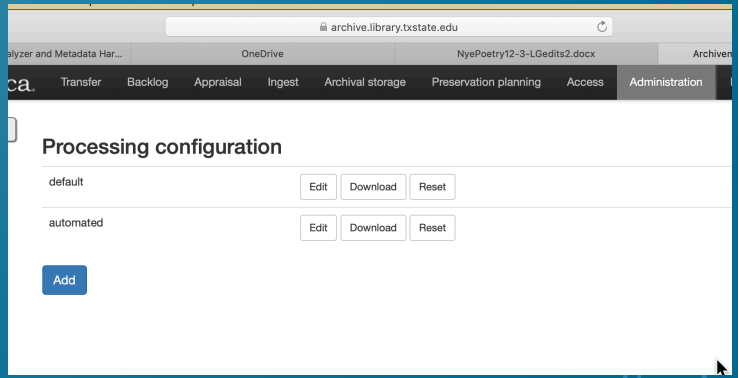
-Workflow and decisions

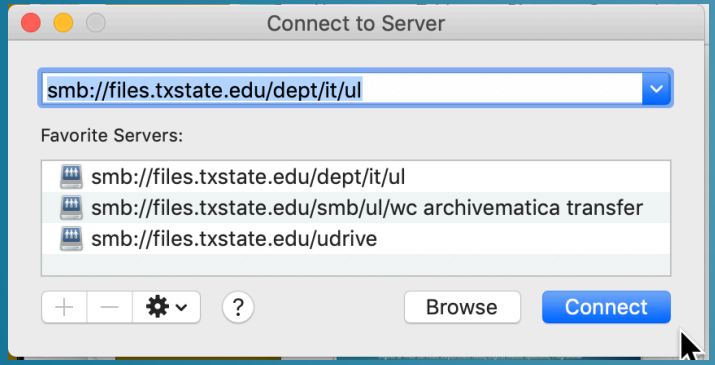
ARCHIVEMATICA DEFAULTS

I edited these.

Normalize

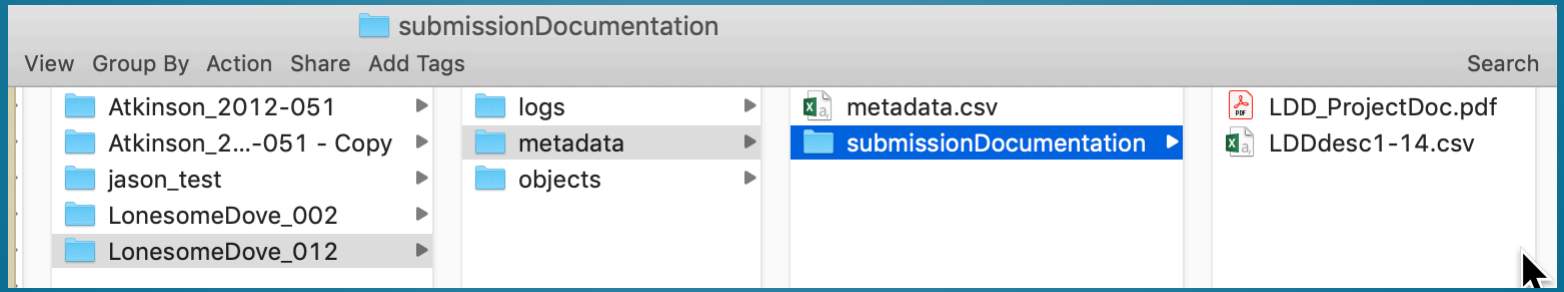
Compression algorithm and level

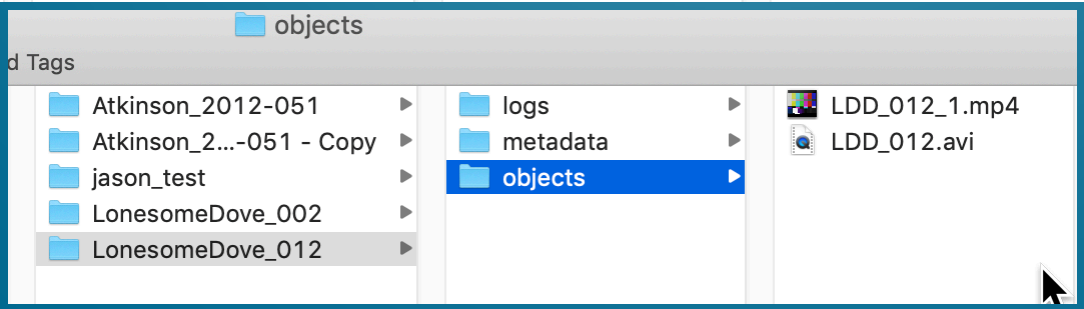




PREP – FOLDERS!

- Automate?
- QC files and metadata





DOCUMENTATION

- Multiple excel sheets!
- Procedure documents become documentation
- -AIPChecklist.docs (what prep is needed for Archivematica processing)
- -AIPChecklist.xlsx Review of current digital materials (accessing AIP checklist status)
- -AIPPriorityList.docx (with file sizes, and labor time guestimate)
- -ArchivematicaTransfer.xlsx (What I Did) (and metadata.csv template)
- -ArchivematicaWorkflow.docx (procedure document)

NEXT STEPS

- Figure out audio/multiple files per item (workflow)
- Rack up some AIPs (Spring intern)
- Ideas?

- Lauren Goodley
- Archivist, The Wittliff Collections
- Texas State University
- lgoodley@txstate.edu
- twitter @laurenbgood

archivematica camp texas

questions pt. 1

Texas Digital Library

Artefactual Update on Archivematica

- Organisational Changes at Artefactual
  - New Managing Director
  - New hires
- Software Development Roadmap
  - Recently released new features
  - Current Development work, upcoming priorities
- Notable new Artefactual Clients/Archivematica Users
- Organisational Changes in the Archivematica Project
  - Impact of growing and changing user base
  - New services being considered by Artefactual

Towards a New Org Chart

Artefactual is growing and changing

January 2019 - Douglas Cerna hired as developer based in El Salvador

February 2019 - Justin Simpson became Managing Director

May 2019 - Sarah Mason hired as Systems Archivist based in UK

Nov 2019 - Amaya Rodrigo hired as Systems Administrator based in Spain; Jennifer Roberts hired as Systems Archivist based in Yukon, Canada

From January 2018 - Nov 2019, 33% growth in staff (from 21 to 29)

Organisational changes

Establishment of an Artefactual Governance Team (shareholders)

Management team changed from 3 to 6 people

Artefactual Management Team

Managing Director Justin Simpson

Director of Archival and Digital Preservation Services Kelly Stewart

Technical Support & IT Services Manager Darren Craze

Project Delivery Manager Joel Simpson

Archivematica Program Manager Sarah Romkey

AtoM Program Manager Dan Gillean

Artefactual Governance Team

Artefactual Governance Team is made up of the current shareholders

Evelyn McLellan
Justin Simpson
Pieter Van Garderen
David Juhasz

Artefactual follows a model of Steward-Ownership

All shareholders are also employees
Profits are a means to an end, not an end in itself
Governance Team is responsible for setting the principles and mission

Key Business Activities

Artefactual has Five Key Business Activities

Client Programs and Support

Delivering subscription based services to clients (hosting, maintenance, support)

Project Management

Delivering project based services to clients (development, consulting, data migration)

Open Source Projects

Stewarding the Archivematica and AtoM projects

Business Development

Selling and Marketing our services to new and existing clients

Finance

Ensuring the financial stability of the company

Archivematica Roadmap

More Frequent Releases

1.8.1 - 10th January, 2019

1.9.0 - 6th March, 2019

1.9.1 - 11th April 2019

1.9.2 - 28th June 2019

1.10 - 5th Sept 2019

1.10.1 - 23rd Oct 2019

1.11 Planned for Jan-Feb 2020

Archivematica Roadmap

Released in version 1.10

Transfer Backlog - bagit packages (SFU Archives)

Appraisal Tab ‘tag and select’ improvements (SFU Archives)

Import older AIPS (City of Vancouver Archives)

Avalon Integration (Northwestern Univeristy Libraries)

External Identifiers (IISH)

Import existing Structmap (IISH)

SCOPE (Canadian Centre for Architecture)

Automated Testing (Wellcome)

Archivematica Roadmap

Current work planned for 1.11.0

Support Python 3 (part 1 of 2 or more)

Improved Processing Storage Usage Reporting (Artefactual)

Improved Duracloud support (Arterfactual)

Anticipated scalability improvements (Norwegian Health Archives/Piql)

PREMIS event import (Norwegian Health Archives/Piql)

Update to PRONOM v. 95 (Artefactual)

Removal of quarantine (Artefactual)

Archivematica is Growing

Notable new Archivematica users and/or Artefactual Clients in last year

Wellcome Collection (UK)
Norwegian Health Archives (Norway)
Canadian Research Knowledge Network (Canada)
DANS (Netherlands)
AARNet (Australia)
National Library of Norway (Norway)
Washington Research Library Council (USA)
Edmonton Public School Archives (Canada)
Memorial University Newfoundland (Canada)
Dalhousie University (Canada)
University of Washington Libraries (USA)
University of Nevada, Reno (USA)

Archivematica is Growing

Notable new Archivematica users and/or Artefactual Clients in last year

Computer History Museum (USA) - new Artefactual client
Yukon Archives (Canada)
City of Ottawa (Canada)
Rennsselaer Polytechnic Institute (USA)
Nelson Mandela Foundation (South Africa)
AMBEV (Brazil)
Ed Rachal Foundation (Texas)
City of Tilburg (Netherlands)
Picturae (Netherlands)
Brooklyn Historical Society (USA)
National Library Board of Singapore
National Archives of Chile

Archivematica is Evolving

Process changes

- Issues repository
- Architectural Decision Record repository
- Archivematica Automated User Acceptance Tests (AMAUAT) release process
- More public visibility into process and invitation to discussion
  - Python 3 upgrade
  - Normalization discussion
- “Proper” roadmap still a work in progress

Towards New Archivematica Services

Over the last several months, Artefactual has been working with a wide range of organisations, to help define how Archivematica should grow and the best role for Artefactual in that growth.

Wellcome Collection hosted a meeting in February 2019 where some of these ideas were first written down.

We'll have more information and public announcements in the new year.

Towards New Archivematica Services

Product Support Program

Roadmap development and maintenance

Support for third party (non Artefactual) developers

Enterprise Partner Program

Ease transition from Project to Program

Establish Large Scale Institutional level Digital Preservation Programs

Large national level organisations, consortia, service providers

Towards New Archivematica Services

Proof of Concept and Pilot Projects

Fixed length engagements to start new programs or move from project to program

Gap Analysis (what's missing in Archivematica, gaps in organisational capacity)

Assist in 'learn by doing'

Examples of interested orgs: Bodleian Libraries, AARNet

Community Development

3 Archivematica Camps in 2019 (Vancouver, London, Geneva)

Archivematica Con planned for 2020 (New York)

archivematica

camp texas

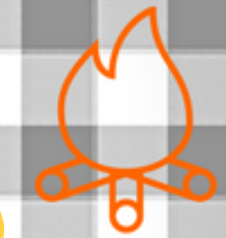
questions pt. 2

Texas Digital Library

archivematica®

camp texas

thank you!



Texas Digital Library