Online Training Conference

DIGITAL DIRECTIONS

Fundamentals of Creating and Managing Digital Collections

PRESENTED BY NEDCC

Session 7:
Digital Preservation Tools
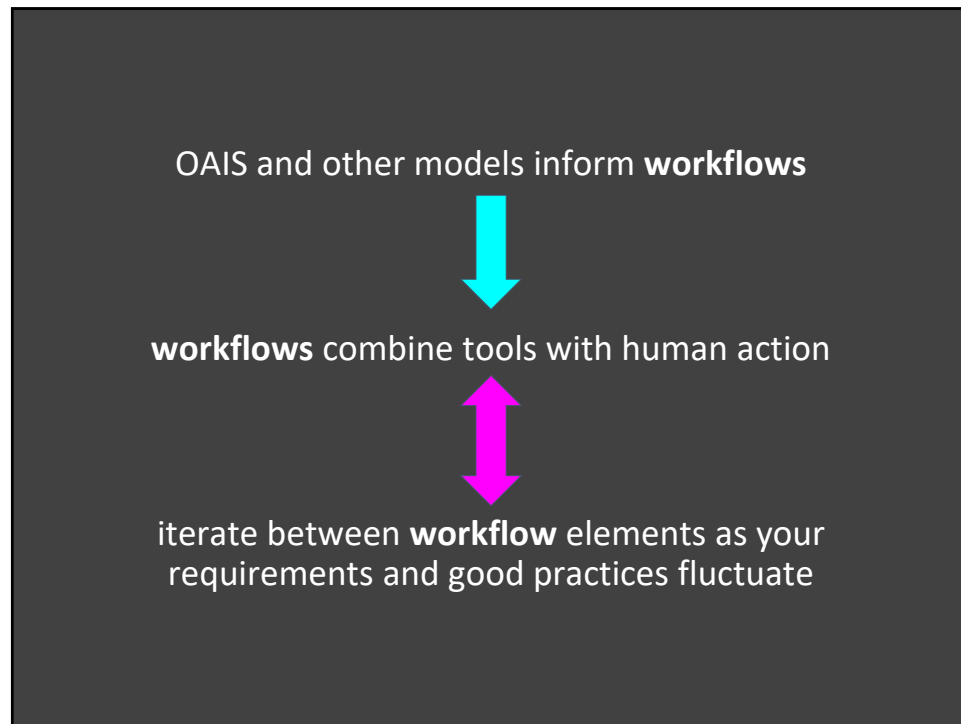
Courtney C. Mumma, MAS/MLIS, Deputy Director

Texas Digital Library (TDL), San Antonio

c.mumma@austin.utexas.edu @lookouthoney

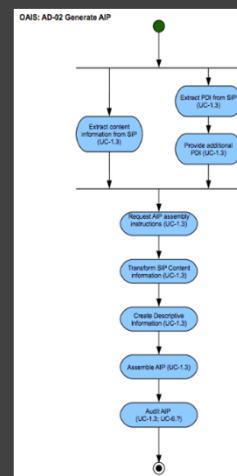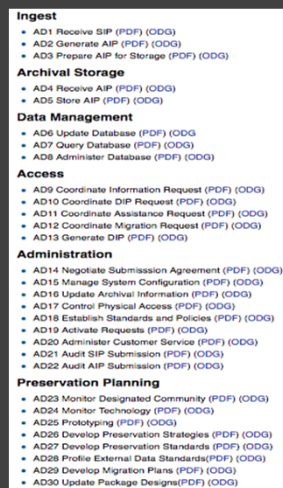www.nedcc.org   Wednesday, 11/17/2021

No one tool is the solution to your digital preservation problems

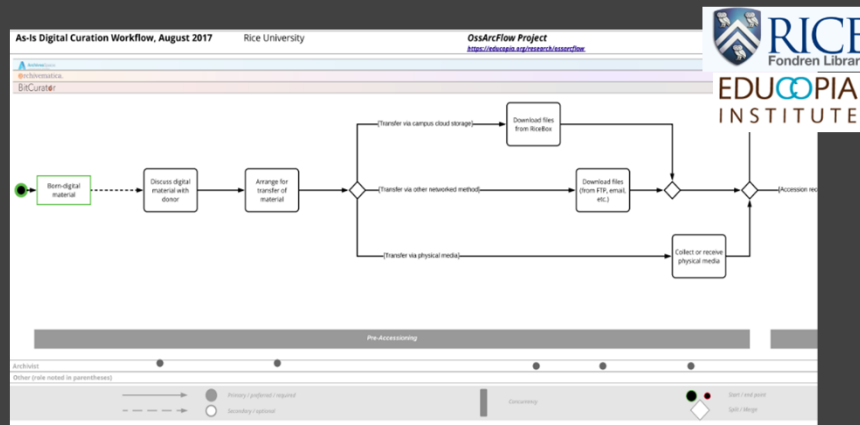Digital preservation is technology-enabled & human-driven

OAIS and other models inform **workflows**

workflows combine tools with human action

iterate between **workflow** elements as your
requirements and good practices fluctuate

Workflows: proto-archivematica

https://wiki.archivematica.org/OAIS_Activity_Diagrams

Workflows: COW

Workflows: OSSArcFlow

The Use of Preservation Tools among Dutch Heritage Organizations

20 May 2020
By Ania Molenda

dutch digital
heritage
network

https://www.netwerkdigitaalerfgoed.nl/wp-content/uploads/2020/06/NDE-The-Use-of-Preservation-Tools-among-Dutch-Heritage-Organizations-juni2020-2.pdf

Your own workflows will be documentary & aspirational....

aka "As-Is" & "To-Be"

share your workflow experience
(doesn't have to be digital preservation)

Integrating tools into your workflows

- **Start small.** Basic tools to accomplish individual tasks are an improvement, but only a step away from your manual processes
- **Refine your tools.** Automate and/or use bundled tools/systems wherever possible and reasonable
- **Update your workflows.** Define sequence of tool application and actions
- **Refine your workflows.** String tools together

Assess Your Requirements

Once you've recorded your As-Is workflows, you can begin identifying tasks that you would like to accomplish with a tool or set of tools.

--Where do you want more automation? More metadata? Where are you happy with current tools? What are your pain points?

ex. An archivist might use dd in the command line to copy disk images in a current workflow. In order to make the workflow more inclusive and gather more metadata, they might want to move to something like Forensics Toolkit to accomplish imaging tasks in a friendly

Resources for Information about Tools

- Community Owned digital Preservation Tool (COPTR) Registry http://coptr.digipres.org/Main_Page
- Preserving (Digital) Objects with Restricted Resources (POWRR) http://digitalpowrr.niu.edu/
- PRONOM Information Resources http://www.nationalarchives.gov.uk/aboutapps/pronom/tools.htm

PAR - Preservation Action Registries

- Artefactual, Arkivum, Preservica, Open Preservation Foundation have completed the initial work phase (funded by Jisc's Research data shared service initiative
- About PAR
  - https://openpreservation.org/about/projects/par/
  - Project pages: http://parcore.org/
  - Slides and videos from Digital Preservation Coalition workshop: https://www.dpconline.org/events/past-events/preservation-action-registry
  - https://github.com/JiscRDSS/rdss-par

Select Tools

What can my organization manage realistically?

(policies, procedures, mandate, support, etc)

What technology could we operate and support?

(which tools fit best with existing infrastructure? can IT help?)

What resources can we dedicate to this work?

(humans, $$$$$, time)

Select Tools

- Interoperability
  - Will it work with other tools you have or will add?
- Right-sized
  - Is there a bunch of extra functionality that you don't need?
    - Look for the appropriate granularity for problem and setting
  - Is there enough useful documentation and training for you effectively implement the tool?
- Maintenance
  - Can you maintain the tool over inevitable changes in staff, roles?
  - How often does the tool update/change?
  - Will you have the bandwidth to keep up with new or changed functions?

Open Source Mythbusting

- Free
- Unsupported
- Only used by small or under-resourced institutions
- Security is more of a problem than in proprietary tools
- Incompatible with proprietary tools

https://www.kisspng.com/png-pusheen-cat-unicorn-tenor-pusheen-cat-2977904/" target="_blank">Pusheen Cat Unicorn Tenor - pusheen cat @kisspng

Ask for the same things from open source and proprietary tools

openness, transparency, interoperability, security, and support

Individual tools exist for every step in the digital preservation process. We'll focus on commonly used tools with bundled functionality.
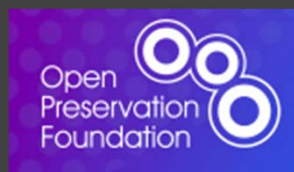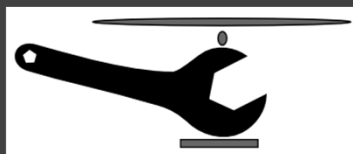
Tools for figuring out what you have

The Big 4 In Digital Preservation Processing

Web Archiving Tools

Email Tools

- Emailchemy: http://www.weirdkid.com/products/emailchemy/
- EPADD: https://library.stanford.edu/projects/epadd/about

- CLIR report - The Future of Email Archives (Aug 2018) https://www.clir.org/pubs/reports/pub175/
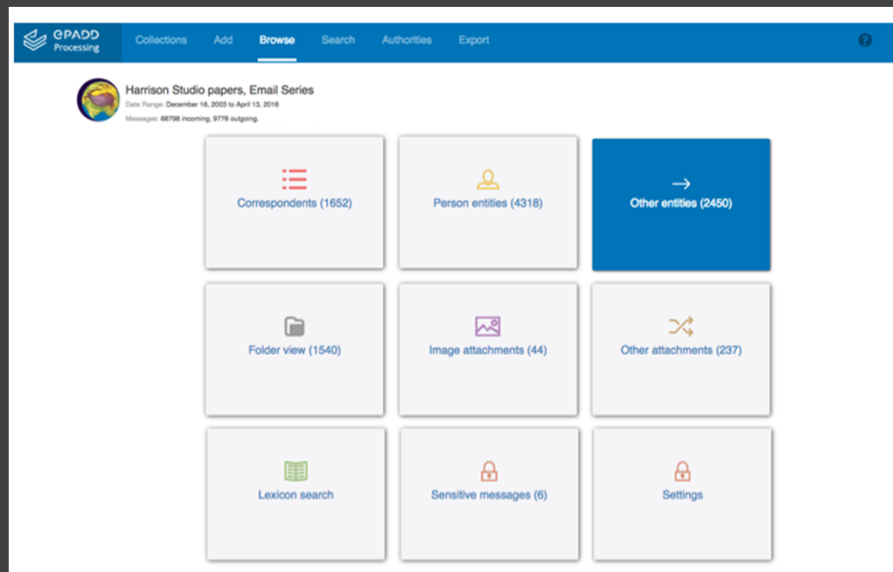
Harrison Studio papers — Email Series, Stanford University, ePADD Processing module, 2017 (Version 3.0).
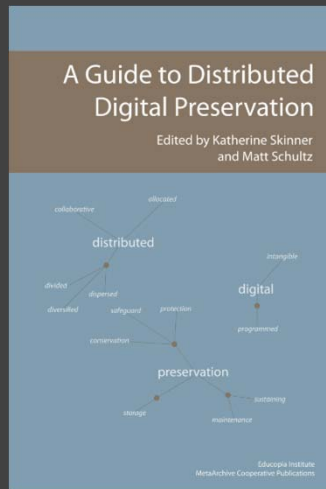
23
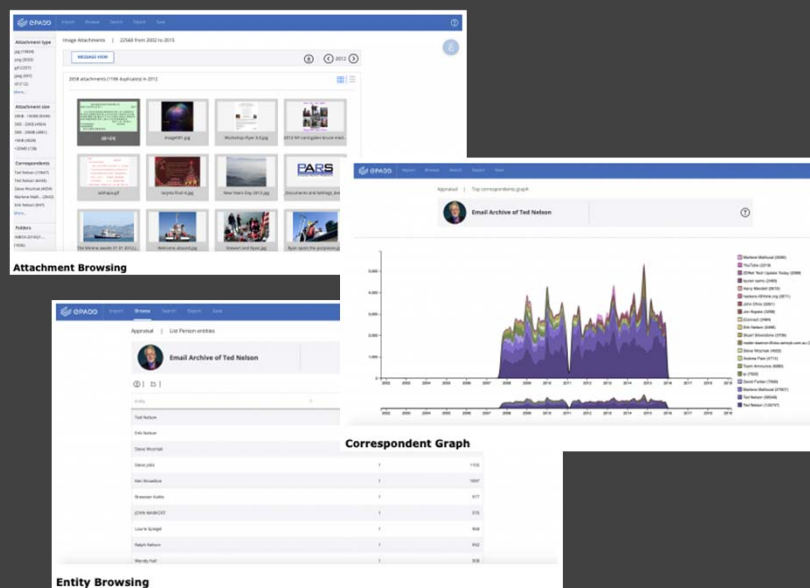
Browsing via the Image Attachment filter, viewing the correspondent graph and entity browser for Ted Nelson's emails at Stanford.

24

Storage: Distributed Digital Preservation

Distributed digital preservation methodologies hold that any responsible preservation system must **distribute copies of digital files to geographically dispersed locations**. It also must **preserve, not merely back-up**, the files in these different locations.

MetaArchive Cooperative. A Guide to Distributed Digital Preservation, book, 2010; Atlanta, Georgia. (https://digital.library.unt.edu/ark:/67531/metadc12850/: accessed December 5, 2019).
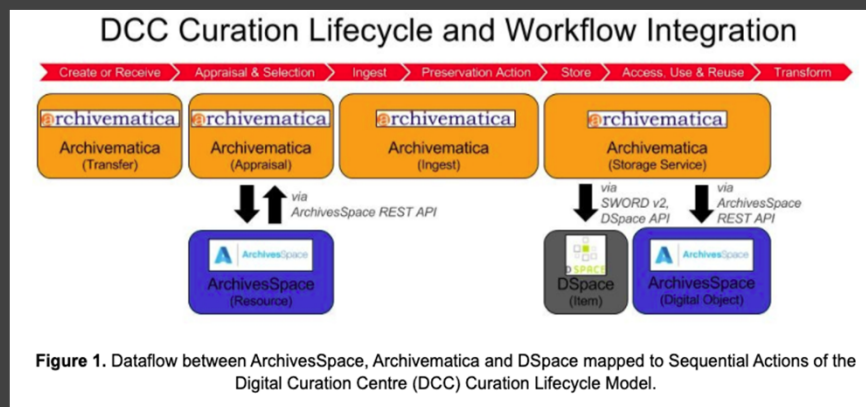
Preservation Storage
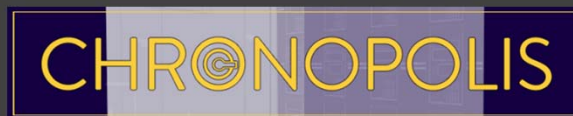
Preservation Storage Criteria

- The Preservation Storage Criteria provides a list of criteria, or design attributes, for storage that supports the work of digital preservation.
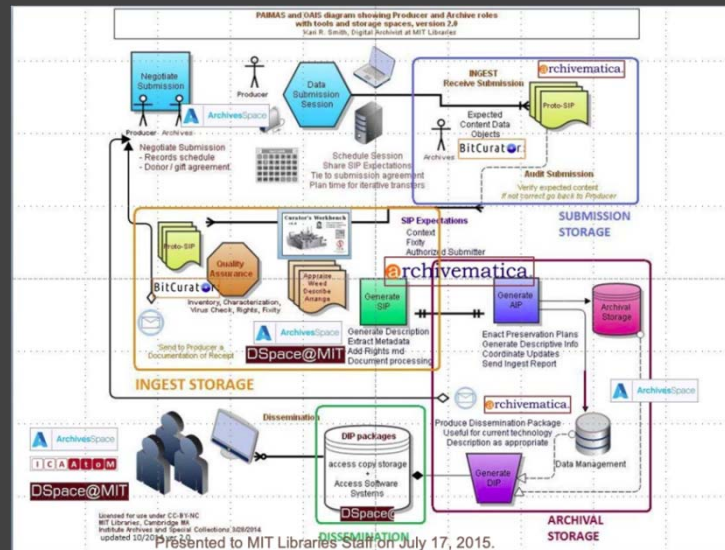 - https://osf.io/sjc6u/

Digital Preservation Storage Criteria game created by Andrea Goethals at iPRES2018

Examples of systems in a workflow

28

Examples of systems in a workflow

29

More more more

- Carleton College has great textual examples of their entire end to end workflow with tools and commands: https://wiki.carleton.edu/pages/viewpage.action?pageId=12246353
- Discrete tools and specifications you will hear about that are wrapped into steps provided by the systems I've talked about include:
  - BAGIT, EXIFTool, FIDO, Sigfried, md5deep, ImageMagick, Tesseract, 7Zip, Sleuthkit, bulkextractor, wayback and open wayback, Clam AV, ElasticSearch, Solr, Ffmpeg, FITS, Inkscape, JHOVE, MediaConch, MediaInfo, and MORE!!!

30

what tools do you use? and how do you string them together?

what tools do you have questions about?

Common tool challenges

- Installation glitches and other bugs
- Development delays
- Insufficient documentation
- Unresponsive support
- Task and data redundancy - overlapping functions
- Incompatibility
- Learning curve / training requirements
- Data dependence - only renderable/understandable using the tool
- Organizational / IT support (access challenges, root and/or admin permissions)
- Multiple dependencies

Pay attention to what the community is doing and share what you're doing. Engage & participate.
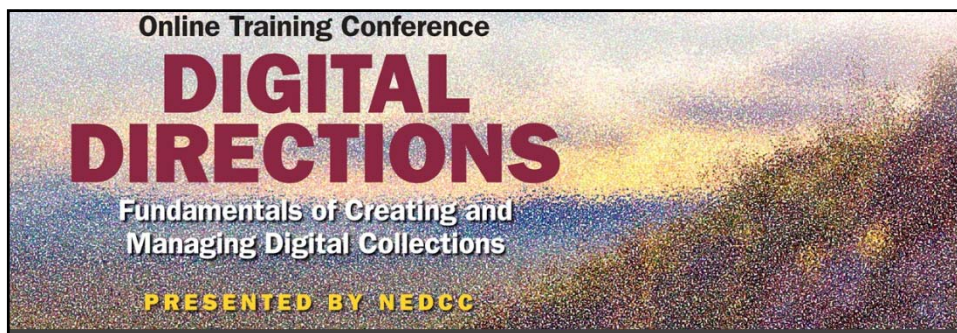
Ex. https://github.com/archivematica/archivematica-case-studies

On the horizon

- Emulation (ex. EaaSI and SPN - Emulation as a Service Infrastructure and the Software Preservation Network http://www.softwarepreservationnetwork.org/eaasi/)
- Free web archiving tools: Webrecorder **OldWeb.today**
- AI/Machine learning (ex. Machine Learning in Libraries Summit https://labs.loc.gov/static/labs/meta/ML-Event-Summary-Final-2020-02-13.pdf, Predictive Coding at U of Illinois https://osf.io/6yp9j/)
- Ethical challenges to processes and tools (ex. https://www.docnow.io/)
- Private & Sensitive Data Storage
- Environmental sustainability

questions?

Thanks to the DPM workshops and the AVP/ workflow curriculum for content which was adapted for part of this session.

Online Training Conference

DIGITAL DIRECTIONS

Fundamentals of Creating and Managing Digital Collections

PRESENTED BY NEDCC

Thank you!

Please complete the online evaluation form.
A link was sent to your email address.
You can edit your responses until you click submit.