

KEYS TO DIGITAL PRESERVATION: MANAGEMENT, TECHNOLOGY, AND CONTENT

ACRL Choice Webinar series

April 11, 2017

Edward M. Corrado

University of Alabama Libraries, Tuscaloosa, Alabama, USA

Heather Moulaison Sandy

iSchool at the University of Missouri, Columbia, Missouri, USA

AGENDA

- What is digital preservation?
- Management aspects of digital preservation
- Technical aspects of digital preservation
- Content aspects of digital preservation
- What can you do now if you don't already have a preservation system?

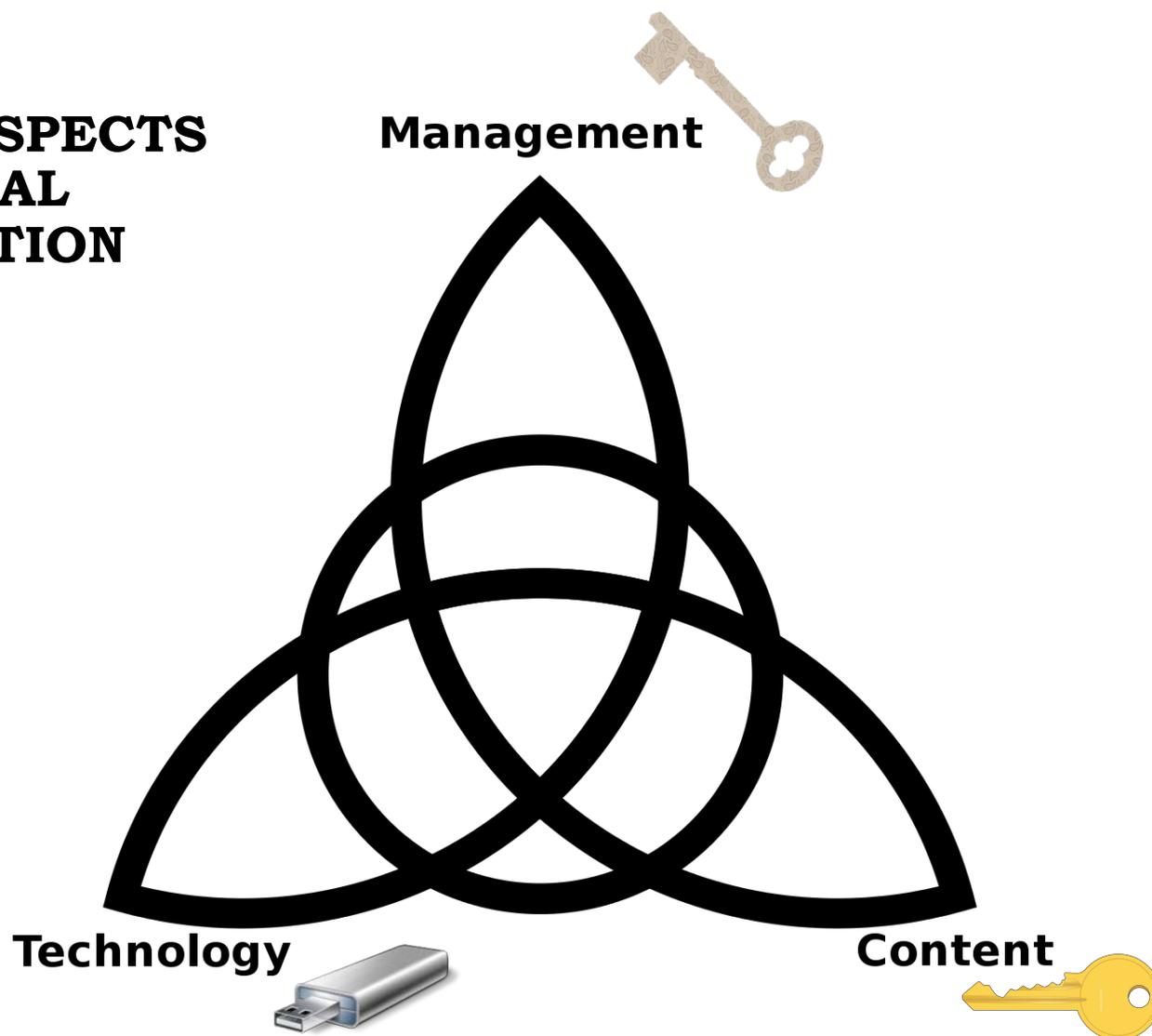
WHAT IS DIGITAL PRESERVATION?

- “Digital preservation is the active management of digital content over time to ensure ongoing access”
<http://www.digitalpreservation.gov/about/>
- “digital objects are mediated by technology. For the objects to be used, a user must have access to the right combination of hardware and software to enable the object to be re-created. Certain nontrivial consequences flow from this unavoidable technological dependence of digital objects. The major consequence is that **it is not possible to leave the digital object alone and expect it to survive**” (Wilson, 2007).

WHERE ARE YOU?

- We've prepared a quick survey to get a feel for where audience members on this call are in terms of their digital preservation initiatives.
- Please answer the poll and I'll summarize your responses when you've finished.

THREE KEY ASPECTS OF DIGITAL PRESERVATION



FIRST KEY: MANAGEMENT ASPECTS



POLICY AND PLANNING

- Digital preservation, by definition, is a long-term activity. To support it requires policies that promote success.
- Policies are high-level documents reflecting the mission of the institution.
 - Digital preservation requires the development of “an institutional approach to digital preservation and establishing a policy of commitment to the long-term maintenance of digital objects and collections” (Xie & Matusiak, 2016, p. 9).
- *Plans*, unlike policies, are directly actionable and can be case or collection specific.

ORGANIZATIONAL GOALS

- Digital preservation initiatives and their policies need to align with an organization's goals and mission
 - If they do not, the digital preservation initiative might not be sustainable
 - “Digital preservation is a long-term project, so service provision must also be long-term” (Billenness, 2008).

ADVOCATING FOR RESOURCES

- Securing resources (money, personnel, technology, content, etc.) is sometimes tough since, by definition, digital preservation is not an immediate return on investment (ROI)
 - “A 1993 British Library strategic review noted that the Library did both access and preservation, access for today’s users and preservation for tomorrow’s users. Only today’s users, however, helped pay the bills. A preservation plan must balance priorities over time” (Lesk, 2017, p. xix).

HR CONSIDERATIONS: REPOSITORY PERSONNEL

- OpenDOAR repository administrators in the US were surveyed to understand how to improve metadata quality (Moulaison Sandy & Dykas, 2016).
 - An educated repository staff was determined to be key to supporting access

SECOND KEY: TECHNOLOGY

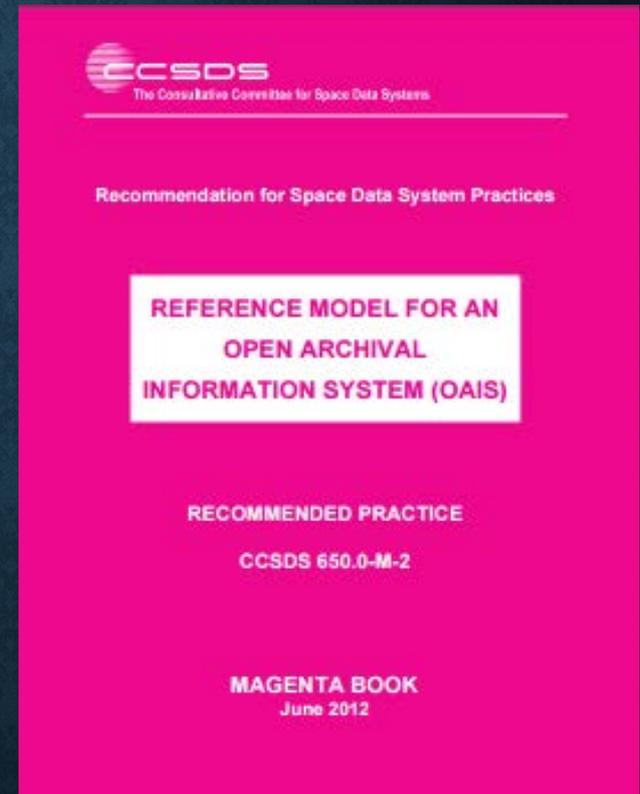


DIGITAL PRESERVATION SYSTEMS

- The Open Archival Information System (OAIS) Reference Model
- Digital Curation Centre (DCC) Curation Lifecycle Model
- A (very) brief discussion of a few digital preservation systems, just to give you some ideas of products that are out there.

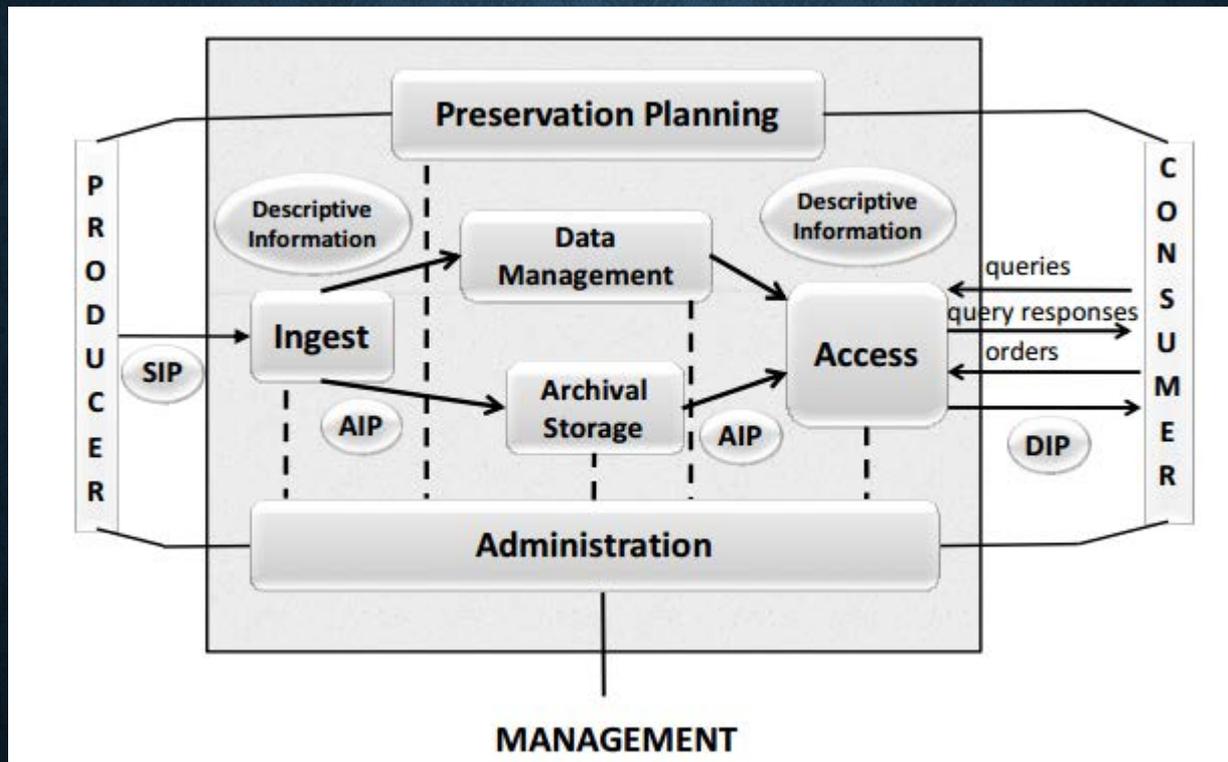
REFERENCE MODEL FOR AN OPEN ARCHIVAL INFORMATION SYSTEM (OAIS)

- A magenta book (**Magenta Books: Recommended Practices**)
- The OAIS Reference Model helps us define a lot of complex concepts through the use of a single, unified vocabulary.
- It also helps us understand what our systems need to be able to do.
 - What DOES mark the difference between a standard digital library software package and one that supports preservation?



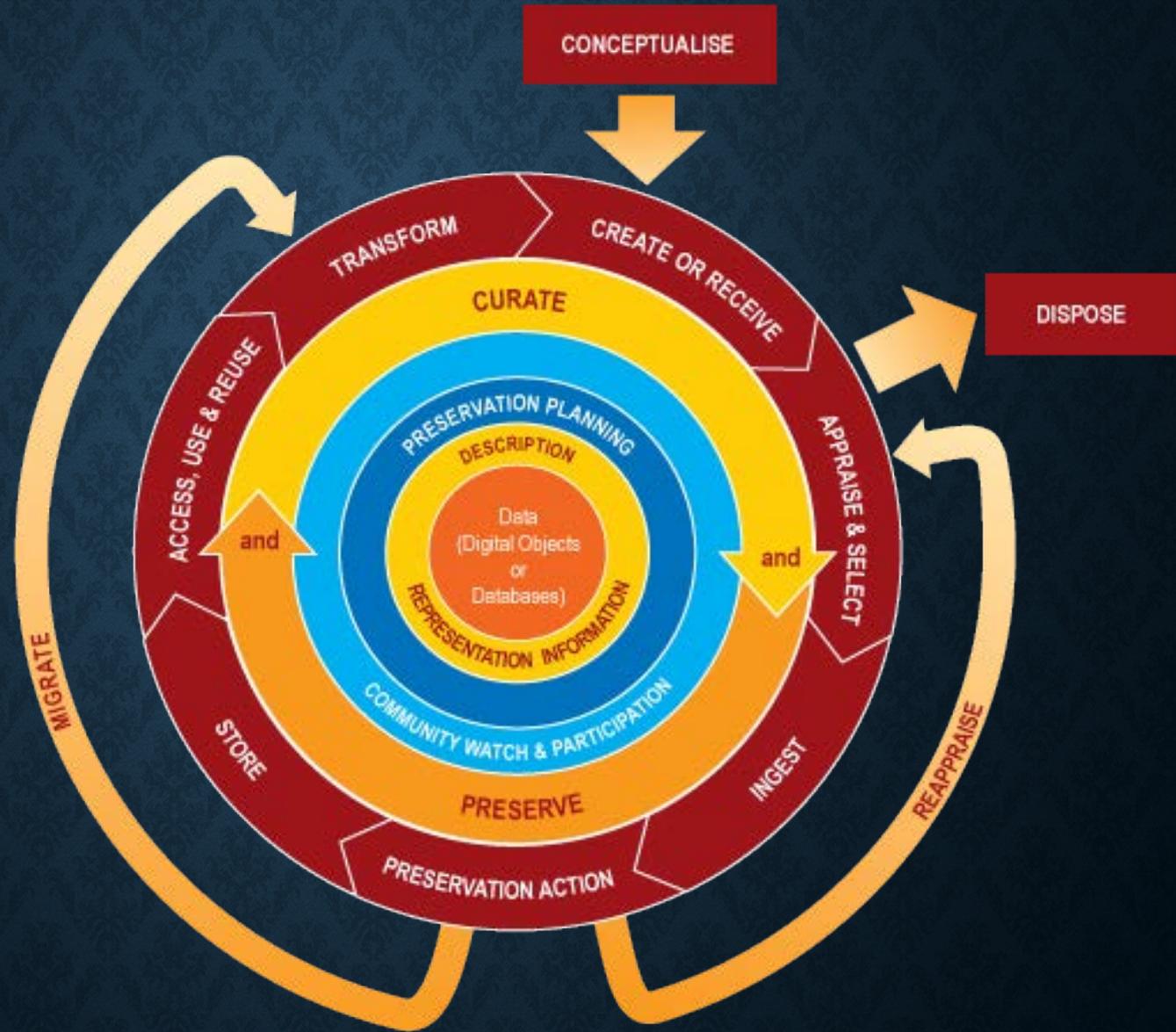
[http://public.ccsds.org/
publications/archive/
650x0m2.pdf](http://public.ccsds.org/publications/archive/650x0m2.pdf)

OAIS REFERENCE MODEL'S FUNCTIONAL ENTITIES



<https://public.ccsds.org/pubs/650x0m2.pdf>

DIGITAL CURATION CENTRE (DCC) CURATION LIFECYCLE MODEL



DURASPACE'S DURACLOUD

- DuraSpace's DuraCloud (<http://www.duracloud.org/>) is an open source platform that uses the cloud for storage (Kimpton & Morris, 2014) and that offers a hosted version that systematically uses two cloud vendors (Han, 2015).
- Fully hosted
- Used by a wide range of institutions worldwide
 - Large and small academic libraries
 - Organizations
 - Public libraries

ARCHIVEMATICA

- Format policies indicate that “preservation formats must all be open standards” (“Technical Architecture,” n.d., Documentation (Archivemata 1.5 (beta))).
- Archivemata is currently in 1.6.
 - “Archivemata is compatible with METS, PREMIS, Dublin Core and other best-practice metadata standards” (Schumacher et al., 2014).

LIBNOVA'S LIBSAFE

- Libnova (<http://www.libnova.com/en/>) is a company that provides digital preservation solutions to the cultural heritage sector. The company started in Spain in 2009 and expanded internationally in 2014.
- One of Libnova's digital preservation solutions is Libsafe, an OAIS and ISO 14,721 compliant digital preservation system that can be hosted in the cloud (or locally).
- It is designed to make “difficult tasks easy” (“About Us,” n.d., Technology Changes, Information Prevails).
- Libsafe is a flexible system that can preserve any file format, identifying and validating over 1,400 of them, and allows for both standard and custom metadata schema (“Libsafe Digital Preservation Software” n.d.).

PRESERVING (DIGITAL) OBJECTS WITH RESTRICTED RESOURCES (POWRR) GRID

- The table evaluating features of repository software is freely available:
<http://digitalpowrr.niu.edu/tool-grid/>
 - Originally produced of an IMLS grant from a number of years ago
 - POWRR Grid v2 is based off of the Community Owned digital Preservation Tool Registry (COPTR) wiki
- Community members still update the wiki from time to time and it's worth a look as you get started thinking about what kinds of capacities your system should have.
 - The Ubergrid, for example, is helpful in looking up systems that support content by file type: <http://www.digipres.org/tools/ubergrid/>
- Even if not entirely up to date, it's good to think of all the options available and then see which products might now meet your needs

<http://www.digipres.org/tools/ubergrid/>

Access

Annotation

Discovery

Redaction

Data capture and Deposit

Disk Imaging

File Copy

Access, Use and Reuse

Create or Reuse

Audio

Audio/Video to WAV Converter 

BWF MetaEdit 

CDRDAO (CDR Disk At Once) 

Clipper  ✓

DBpoweramp Music Converter 

Easy CD-DA Extractor 

Exact Audio Copy 

ExifTool 

FFmpeg 

GetID3() 

IsoBuster  ✓



FILES: PROMOTING LONG-TERM ACCESS

- “The greatest danger to digital materials is that we forget the meaning of them. Preservation depends on our knowledge: we may have bits but be unable to interpret them“ (Lesk, 2017, p. xviii).
- Decisions about file types to preserve
 - Compression
- Storage decisions
 - Best practices dictate we store files in at least two separate places, ideally geographically diverse
 - Might seem cut and dried, but recent discussions on an email distribution list remind us that there is one one-size fits all solution
- Ensuring the integrity of the files
 - Bit rot can destroy files over time
 - Through the use of specialized systems and software, we are able to ensure the integrity of field

WHICH FILE TYPES ARE BEST?

- No surprise that there is no one-size-fits-all solution...
- Generally, digital preservationists will opt to preserve files that are
 - Open
 - Widely used (and with good community support)
- As management decisions are made and as content is solicited for inclusion, digital preservationists need to be sure they have the rights to transform files in the future into the latest file formats (e.g., Word 2025, anyone?)

METADATA IS NECESSARY FOR DIGITAL PRESERVATION

- There are standards that support storage, access, and retrieval
 - Those are necessary for any kind of digital library initiative, since otherwise digital files are completely invisible
- There are other standards that support digital preservation in particular

PREMIS AS PRESERVATION METADATA

Preservation metadata supports the long-term access and use of content. The primary preservation metadata schema is PREMIS (PREservation Metadata: Implementation Strategies), now in version 3.0.

- The newest major revision of the Data Dictionary was released in June 2015 and was last updated in November 2015 (<http://www.loc.gov/standards/premis/v3/index.html>).

PREMIS, CONTINUED

- PREMIS continues to be recommended across the board, even if its inclusion is not always straightforward in digital libraries and its creation is labor-intensive
- A new book on PREMIS has just been published: *Digital Preservation Metadata for Practitioners: Implementing PREMIS*, Dappert, Angela, Guenther, Rebecca Squire, Peyrard, Sébastien (Eds.)

RIGHTSSTATEMENTS.ORG

- *Rights metadata* helps users make use of content. Digital Public Library of America and Europeana announced in mid-April, 2016, the launch of RightsStatements.org.
 - As a response to a white paper from 2015
 - In conjunction with Creative Commons
 - Proposes 11 standardized rights statements for online cultural heritage that are human- and machine-readable

THIRD KEY: CONTENT



WORKING WITH CREATORS/OWNERS

- Importance of getting creators on board
 - They can describe their content
 - They know the field and what will be interesting in the future (in theory)
 - They also will be users, potentially

THINKING OF USERS

- What do your users expect?
- What expectations are linked:
 - To the system and retrieval?
 - To the file and use?

SPECIALTY CONTENT

- Text isn't the only kind of file worth preserving for the future...
 - Everything we described before is relevant to non-text resources, too
 - Specialty content like (AV files, etc.) has additional considerations that are inherent
 - Additional management issues, additional technology considerations, etc.

AUDIOVISUAL CONTENT

- Audiovisual file formats can be complicated
 - For example: to compress or not to compress?

WEB SITES

- Web archiving is the process of harvesting web pages and other content on the Web and “storing it, ensuring the data is preserved in an archive, and making the collected data available for future research” (Niu, 2012, para. 1).
- The Internet Archive
 - The Wayback Machine
- Web archiving services
 - Archive-It (from Internet Archive)
 - ArchiveThe.Net

THREE DIMENSIONAL (3D) STILL IMAGE AND VIDEO

In 2013 Autodesk and the Smithsonian Institute teamed up to create x3D Explorer, which “will allow the Smithsonian to digitally preserve its extensive collection as interactive, 3D models” (Quirk, 2013, para. 1).

- This project also makes 3D files available to download in a variety of file formats.
- Some of them are in long-standing 3D file formats such as OBJ and STL.
- However there are also newly emerging 3D file formats Many of these are for use in 3D computer graphics applications.
- In many cases, gaming applications are leading the way. Three examples are:
 - Open Game Engine Exchange (OpenGEX) format
 - X3D, an XML-based file format which is designed to represent computer graphics in 3D, and
 - Alembic, an interchange format for computer graphics that has been widely adopted by visual effects and animation professionals.

PRESERVING AND EMULATING DIGITAL ART OBJECTS (CORNELL UNIVERSITY)

- **The Preserving and Emulating Digital Art Objects White Paper** describes the project's findings, discoveries, and challenges. The ultimate goal of the project team has been the creation of a preservation and access practice grounded in thorough and practical understanding of the characteristics of digital objects and their access requirements, seen from the perspectives of collection curators and users alike. Equally important has been the establishment of service frameworks and policies that are sustainable, realistic, and cost-efficient. So all through the project, one of our principles has been moving the experience gained through research into practice. This white paper aims to contribute to better and more practical understanding, management, and curation of digital assets. Although the initiative focused on new media art, we hope that our methodologies and findings will inform other types of complex born-digital collections as well.

<https://confluence.cornell.edu/display/pafdao/Home>

WHAT CAN YOU DO NOW IF YOU DON'T ALREADY HAVE A DIGITAL PRESERVATION SYSTEM?

- Education / Training (Webinars, workshops, conferences, classes, books)
- Identify content / Create a Digital Asset Registry (Digital Preservation Coalition, 2017)
- Organize Your Digital Content
- Make Copies (Backups)
 - Multiple copies in multiple locations
- Manage Content Over Time
- Risk Assessment
- Document Your Processes / Decisions (Throughout)

CONCLUSION

- Digital preservation is a new frontier that is both challenging and exciting
- Three key interrelated aspects of digital preservation
 - Management
 - Technology
 - Content
- Any kind of undertaking like this is in many ways a *management exercise*

THANK YOU!

Questions?

REFERENCES

- About Us. (n.d.). Libnova. Retrieved from <http://www.libnova.com/en/about-us/>.
- Billenness, C. "Building a Sustainable Model for Digital Preservation Services" (paper presented at the third annual We Preserve Conference, Nice, October 28–30, 2008), slide 2, accessed September 29, 2013, <http://www.digitalpreservationeurope.eu/preservation-training-materials/files/WEPRESERVEsustainability.pdf>.
- Canadian Heritage Information Network, "Checklist for Creating a Preservation Policy," <http://canada.pch.gc.ca/eng/1443721644527/1443721642096>.
- Dappert, A. (2015, May 27). "Digital preservation metadata and improvements to PREMIS in Version 3.0," webinar presented at the DCMI/ASIST Joint Webinar. Retrieved from <http://www.loc.gov/standards/premis/v3/tutorial.html>.
- Digital Preservation Coalition, "Digital Preservation Handbook," <http://www.dpconline.org/handbook/getting-started>.
- Han, Y. (2015). Cloud storage for digital preservation: Optimal uses of Amazon S3 and Glacier. *Library Hi Tech*, 33(2), 261–71. doi:10.1108/LHT-12-2014-0118.
- Kaur, H. (2015). Digital preservation of manuscripts: An Indian perspective with special reference to Punjab," in *Emerging Trends and Technologies in Libraries and Information Services (ETTLIS)*, 2015 4th International Symposium on (Jan. 6-8, 2015): 271-274. doi:10.1109/ETTLIS.2015.7048210.
- Kengyel, E. (n.d.). "Open Game Engine Exchange (OpenGEX)." Retrieved from <http://opengex.org/>.
- Kimpton, M., & Morris, C. M. (2014). Managing and archiving research data: Local repository and cloud-based practices. In *Research Data Management: Practical Strategies for Information Professionals*.
- Libsafe Digital Preservation Software. (n.d.) Retrieved from <http://www.digitalpreservationsoftware.com/digital-preservation-solutions/libsafe-digital-preservation-software/>.

- Libsafe. (n.d.). Retrieved from <http://www.digitalpreservationsoftware.com/wp-content/uploads/documents/20140201-libsafe-catalog-eng.pdf>.
- Niu, J. (2012) "An Overview of Web Archiving," D-Lib Magazine 18, no. 3/4 (March/April 2012), para "Introduction," doi: 10.1045/march2012-niu1.
- Norris, S. (2014) Toward an ontology of audio preservation. *Journal of the American Institute for Conservation*, 53(3), 171-181. doi:10.1179/1945233014Y.0000000026
- Panzer-Steindel, B. (2007, April 8). Data Integrity: Draft 1.3. Retrieved from https://indico.cern.ch/event/13797/session/0/contribution/3/attachments/115080/163419/Data_integrity_v3.pdf.
- PREMIS Editorial Committee. (2015). PREMIS data dictionary for preservation metadata: Version 3.0. Retrieved from <http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>.
- Quirk, V. (2013, Nov. 29). Autodesk launches new tool for digital preservation. *ArchDaily*. Retrieved from <http://www.archdaily.com/452845/autodesk-launches-new-tool-for-digital-preservation>.
- Rights statements. (n.d.). Retrieved from <http://rightsstatements.org/en/>
- Schumacher, J., Thomas, L. M., VandeCreek, D., Erdman, S., Hancks, J., Haykal, A. Miner, M., Prud'homme, P.-A., & Spalenka, D. (2014, August). *From theory to action: 'Good enough' digital preservation solutions for under-resourced cultural heritage institutions: A Digital POWRR white paper for the Institute of Museum and Library Services*. Retrieved from <http://commons.lib.niu.edu/handle/10843/13610>
- Sprout, B., & Jordan, M. (2015). Archivemata as a service: COPPUL's shared digital preservation platform, *Canadian Journal of Information & Library Sciences*, 39(2), 235-244.
- Technical Architecture. (n.d.). Archivemata: Open-Source Digital Preservation System. Retrieved from <https://www.archivemata.org/en/docs/archivemata-1.5/user-manual/overview/technical/#technical-arch>.
- Wilson, A. (2007). *Significant properties report*. Version 2. Arts and Humanities Data Service. http://www.significantproperties.org.uk/wp22_significant_properties.pdf