Preserving the Fruit of Our Labor: Establishing Digital Preservation Policies and Strategies at the University of Houston Libraries

Santi Thompson, Annie Wu, Drew Krewer, Mary Manning, Rob Spragg

University of Houston
4333 University Drive
Houston, TX 77204-2000
1-713-743-9678

sathompson3@uh.edu; awu@uh.edu; ajkrewer@uh.edu; mmmanning@uh.edu; rspragg@uh.edu

ABSTRACT
To develop a comprehensive digital preservation program for maintaining long-term access to the Libraries' digital assets and align our practices with national standards and guidelines, the University of Houston (UH) Libraries formed the Digital Preservation Task Force (DPTF) to assess previous digital preservation practices and make recommendations on future efforts. This paper outlines the methodology used, including the task force's use of existing models and evaluation criteria, to successfully generate new policies and select Archivematica as our system to process and preserve our digital assets. It concludes with recommended strategies for the implementation of the policies and preservation operations.

General Terms
Institutional opportunities and challenges; Preservation strategies and workflows

Keywords
Digital preservation policy; System evaluation; Archivematica

1. INTRODUCTION
Creating, acquiring, preserving, and making accessible digitized and born digital content has been a major initiative of the University of Houston (UH) Libraries since the founding of the UH Digital Library in 2009. By the summer of 2014, UH Libraries had accumulated ten terabytes of digitized and born-digital content from UH Special Collections and the UH Digital Library.

UH Libraries established many of its digital preservation strategies and techniques for digitized materials within a year of creating the UH Digital Library in 2009. In their 2011 paper, "Implementing METS, MIX, and DC for Sustaining Digital Preservation at the University of Houston Libraries," Mingyu Chen and Michele Reilly outlined the original approach to digital preservation. The authors described a process that relied on a series of tools, including CONTENTdm export functions, JHOVE, and 7train, to generate descriptive and technical metadata in a METS wrapper [1]. Additionally, the article mentioned how UH Libraries was experimenting with the Texas Digital Library (TDL) to create additional storage locations for digital objects

through a cooperative model including the Texas Advanced Computing Center (TACC) [1].

Over time, limitations to this model emerged. While it was critical to capture technical metadata, focusing exclusively on MIX metadata prevented the capture of technical information for other popular file formats, including audio, video, and datasets. The assembled tools also had no way of actively recording and tracking preservation events, through PREMIS metadata or any other mechanism. Perhaps the most important limitation, the existing tools and infrastructure had no formal digital preservation policy guiding current or future practices.

In response to these limitations, as well as to inconsistent practices around digital preservation, UH Libraries formed the DPTF in May 2014. The libraries charged the group with establishing a digital preservation policy and identifying strategies, actions, and tools needed to sustain long-term access to digital objects maintained by the libraries. Along the way, the DPTF combined existing research and evaluation criteria on digital preservation in new ways to generate robust policies and identify a new system that will enact these policies.

2. METHODOLOGY
The DPTF constructed its activities around its charge, which called on the group to:

- Define the policy's scope and levels of preservation
- Articulate digital preservation priorities by outlining current practices, identifying preservation gaps and areas for improvement, and establishing goals to address gaps
- Determine the tools, infrastructure, and other resources needed to address unmet needs and to sustain preservation activities in the future
- Align priorities with digital preservation standards, best practices, and TDL storage services
- Recommended roles, responsibilities, and next steps for implementing the strategy and policy

The DPTF launched parallel actions to fulfill its charge: policy development and system evaluation.

2.1 Policy Development
One activity focused on creating digital preservation policies for content entering into repositories. Policy creation required the group to study formal preservation frameworks, models, and strategies; compare UH Libraries' current practices with best practices from other libraries and archives; and craft new digital

preservation policies that accounted for current technology and future resources.

The Action Plan for Developing a Digital Preservation Program (a toolkit distributed to participants in Cornell University's and The Massachusetts Institute of Technology (MIT)'s Digital Preservation Management: Implementing Short-term Strategies for Long-term Problems workshop) served as the primary tool used for policy creation. Conforming to the Open Archival Information System (OAIS) Reference Model and the Trusted Digital Repository guidelines, the *Action Plan* walks institutions through the process of establishing a high-level framework, creating policies and procedures, building technological infrastructure, and addressing resources needed to sustain a digital preservation program for the long term. The document also includes policies and procedures from other institutions (some of which have used the *Action Plan*) [2].

To create the formal policies that inform UH Libraries digital preservation practices, the group performed the following activities:

- Selected and studied *Action Plan for Developing a Digital Preservation Program* to construct digital preservation policies
- Drafted high-level policy framework
- Outlined roles and responsibilities for internal and external stakeholders
- Defined digital assets including digitization quality and metadata specifications; collection selection, acquisition policies, and procedures; and access and use policies
- Identified and described key functional entities for the digital preservation system, including ingest, archival storage, preservation planning and administration, and access
- Drafted potential start-up and ongoing costs for digital preservation at UH Libraries

2.2 System Evaluation

Complementing policy creation, the task force also focused on evaluating software that UH Libraries will operate to fulfill the requirements of the digital preservation policy. The group reviewed research conducted by the Preserving (Digital) Objects with Restricted Resources (POWRR) project, including their Tool Grid and white paper "From Theory to Action: 'Good Enough' Digital Preservation Solutions for Under-Resourced Cultural Heritage Institutions." These resources provided valuable information on the capabilities and functionalities of over 60 tools and systems used for digital preservation activities [3] [4]. The DPTF used POWRR data to narrow potential tools to three for testing: Archivematica, Preservica, and Rosetta. After participating in demos on all three tools, the group elected to test Archivematica based on existing human and financial resources and in-house technological expertise. They chose not to test the other options because the task force found their costs to be prohibitive. Additionally, the proprietary software and structural metadata associated with the other two platforms were not based on open standards. To evaluate Archivematica, the task force adapted criteria developed by the National Library of Medicine (NLM) and made available in their 2008 report "Recommendations on NLM Digital Repository Software" [5]. Evaluation creation focused on the system's ability to support key services and functions, including:

- File types, including legacy formats

- Versioning control
- Virus and fixity checks
- Specified metadata formats
- Audit trail functionality
- Error reporting
- Archival Information Package (AIP) creation
- Dissemination Information Package (DIP) creation
- AIP storage

Finally, as the group tested Archivematica, it generated a list of questions and presented them to Courtney Mumma, the U.S. and International Community Development officer for Artefactual, Inc. during an onsite consultation with UH Libraries.

3. RESULTS

3.1 UH Libraries' Digital Preservation Policy

Principles outlined in UH Libraries' Digital Preservation Policy include collaboration, partnerships, and technological innovation, all of which are rooted in UH Libraries core values as articulated in both the 2013-2016 Strategic Directions document and our institutional mission. The University of Houston supports scholarship, teaching, and learning. As more resources and services associated with these functions become digital, our responsibilities must expand to include the identification, stewardship, and preservation of designated digital content. Additionally, UH has legal, contractual, and consortial obligations to preserve digital content of local and national significance.

The UH Libraries Digital Preservation Policy consists of three main sections: Policy Framework, Policies and Procedures, and Technological Infrastructure.

3.1.1 Policy Framework

The Digital Preservation Policy Framework supports the missions of UH and is the highest level digital preservation policy document at UH Libraries. It makes explicit UH Libraries' commitment to preserving the digital assets in its collections through the development and evolution of a comprehensive digital preservation program. The framework reflects the goals defined in our institutional mission and contains references to other relevant UH Libraries policies and procedures. The audience for the framework includes librarians and staff of UH Libraries, digital content donors/depositors, funders, and users [2] [8]. Sections in the Digital Preservation Policy Framework address:

- Purpose
- Objectives
- Mandate
- Scope
- Challenges
- Principles
- Roles and Responsibilities
- Collaboration
- Selection and Acquisition
- Access and Use

While it is outside the scope of this short paper to address all of these sections in the policy framework, key sections are described. The Purpose section explains the function of the policy framework and how it relates to more granular policies and procedures developed for UH Libraries [6]. The objective section articulates that UH Libraries defines the primary goal of digital preservation activities as maintaining the ability to meaningfully

access digital collection content overtime. The primary concern is preserving the ability to access the archival digital object from which derivative files may be created or re-created over time [6] [7] [9]. Mandates for digital preservation at UH Libraries are dictated by fulfilling organizational commitments, including complying with the charge of the DPTF, supporting scholarship through long-term preservation of resources, maintaining institutional memory through preserving institutional records, and meeting any outstanding legal, contractual, or consortial obligations [6] [9]. The Scope section broadly outlines which assets will be retained and managed by UH Libraries. These assets include:

- Digital versions of resources owned and reformatted by UH Libraries and that fall under the parameters of UH Libraries' Digital Collection Development Policy
- Unique born-digital resources that are part of UH Libraries' archival/manuscript collections and which are unlikely to be preserved anywhere else
- Any other content acquired or digitized by UH Libraries that falls under the parameters of UH Libraries' Digital Collection Development Policy

3.1.2 Policies and Procedures

This section describes digital preservation policies, procedures, roles, and responsibilities in greater detail than the policy framework. This section outlines requirements around digital assets, including recommended capture specifications for digital objects, preferred file formats supported by the digital preservation system, and stipulations around the acquisition, transfer, and access of content [2] [6]. Additionally, this section of the policy addresses personnel. It identifies internal and external stakeholders, the roles required by the program, and the specific individuals charged with filling the roles [2] [6] [9].

3.1.3 Technological Infrastructure

UH Libraries' Digital Preservation Policy outlines digital preservation system functions and requirements in greater detail than the policy framework [2] [6]. Specifically, it articulates:

- The rules and requirements for Submission Information Packages (SIPs), Archival Information Packages (AIPs), and Dissemination Information Packages (DIPs)
- The workflow for ingesting, updating, storing, and managing digital objects
- The metadata requirements upon ingest
- The strategic priorities for future digital preservation efforts, including risk management

Functional entities implemented in UH Libraries' digital preservation system, such as pre-ingest, ingest, archival storage, data management, administration, preservation planning, and access are OAIS compliant.

3.2 UH Libraries' Digital Preservation System

The DPTF recommends that UH Libraries adopt Archivematica as its digital preservation system. In addition to this local storage solution, the task force also recommends storing digital objects in the cloud through DuraCloud services provided by TDL.

Rooted in digital preservation best practices, Archivematica combines numerous digital preservation tools to facilitate the acquisition, processing, and storage of digital objects. As an open

source digital preservation system, Archivematica is designed to be extensible; the growing and active developer community continues to expand the tools and functionality of the system. It is also being developed to interoperate with other important digital access and preservation platforms, including DuraCloud and ArchivesSpace.

Using the modified NLM evaluation criteria, the task force identified advantages and disadvantages of Archivematica as a system and its implementation at UH Libraries.

3.2.1 Advantages of Archivematica

- Complies with OAIS reference model
- Uses open source solutions to perform digital preservation activities.
- Supports the ingest of a wide array of file formats
- Automates digital preservation policies, such as format choices when normalizing
- Offers active user development community
- Supports versioning through the adoption of the Archival Information Collection (AIC)
- Records digital preservation events and places this information into METS record as PREMIS metadata
- Offers an intuitive user interface that makes it easy for administrators to customize rules, settings, and workflows as well as to track workflow in a transparent way
- Supports complex archival workflows with multiple users having access, if desired
- Integrates with other digital asset management systems, including CONTENTdm, ATOM, and DuraSpace
- Provides a no-cost system solution with a pay structure for software support and/or customized features

3.2.2 Disadvantages of Archivematica

- Challenges IT staff due to its modular microservices architecture; it is built out of individual tools and is not a "set it and forget it" platform.
- Restricts the ingest of descriptive metadata to CSV file or manual input
- Stores objects in one specified location
- Lacks functionality to self-heal corrupted and/or damaged objects
- Limits the roles for users and administrators
- Lacks robust reporting and notification to assist with digital curation tasks

Despite the disadvantages (which could change over time because the system is actively developed), the task force believes that Archivematica offers a good balance of system functionality, future expansion, and ongoing sustainable costs. The task force will evaluate the disadvantages, prioritize them, and find partners to co-fund development solutions. Additionally, other groups, like DuraSpace, could address some of the identified deficiencies in the future.

To complement Archivematica, the DPTF recommends that UH Libraries' store copies of its content with DuraCloud, a cloud-based digital preservation solution. The task force selected DuraCloud because it can be synced directly with Archivematica, allowing for an automated delivery process. Additionally, DuraCloud is fully supported by TDL, which provides other critical services to UH Libraries.

4. IMPLEMENTATION STRATEGIES

With the completion of policy creation and system selection, implementing the newly established program will be the next step. The DPTF suggests that a new group, which we refer to as the Digital Preservation Team (DPT), be formed to create specific workflows that maximize Archivematica's ability to execute digital preservation policies. The creation of this team will allow engaged stakeholders to leverage their diverse knowledge and growing expertise in digital preservation in order to establish day-to-day workflows and procedures. DPT members should resolve several short-term priorities including:

- Training team members on the features and functionality of Archivematica
- Establishing workflows for digitized and born digital content that meet specifications for SIP, DIP, and AIC creation, dissemination, and storage as outlined in the UH Libraries Digital Preservation Policy
- Configuring system settings in Archivematica to automate aspects of the digital preservation policy

Additionally, the team should plan for long-term objectives, including:

- Collaborating with libraries stakeholders to identify and integrate areas where the digital preservation system and the new digital asset management system interoperate
- Advising the libraries on digital preservation policies related to materials that have no existing guidelines, such as electronic serials that are produced by the University and require preservation.
- Assessing and adapting workflows over time to increase efficiency and ensure compliance with policies

5. CONCLUSION

To date, the work of the DPTF has created a model that can inform the larger profession and has benefitted our local institution. The task force combined existing digital preservation research and evaluation criteria in new ways to generate robust policies and to identify a system to sustain these policies. DPTF members believe that sharing this information with external institutions could offer them an evaluation technique to draw upon when beginning the process of establishing digital preservation policies. Locally, the task force linked digital preservation issues with the mission of UH Libraries and aligned the libraries practices with national standards and guidelines, specifically OAIS and the requirements outlined by the Trusted Digital Repository model. By ensuring continued access to the libraries valuable and unique resources, we are protecting substantial institutional investments and supporting the University's goal to establish itself as a preeminent public research university in the 21st century.

6. REFERENCES

[1] Chen, M. and Reilly, M. 2011. Implementing METS, MIX, and DC for sustaining preservation at the University of Houston Libraries. *Journal of Library Metadata* 11 (2): 83-99. DOI= http://dx.doi.org/10.1080/19386389.2011.570662

[2] Kenney, A. and McGovern, N. (n.d.) *Action plan for developing a digital preservation program.* Cornell University and the Massachusetts Institute of Technology.

[3] Preserving (Digital) Objects with Restricted Resources (POWRR). Tool grid. Retrieved on 2015 April 16 from http://digitalpowrr.niu.edu/tool-grid/

[4] Schumacher, J., Thomas, L.M., VandeCreek, D., Erdman, S, Hancks, J., Haykal, A., Miner, M. Prud'homme, P., Spalenka, D. 2014. From theory to action: "Good Enough" digital preservation solutions for under-resourced cultural heritage institutions. A digital POWRR white paper. Retrieved on 2015 April 16 from http://commons.lib.niu.edu/bitstream/handle/10843/13610/FromTheoryToAction_POWRR_WhitePaper.pdf;jsessionid=314748CEF2C065FBE290B0B99A9E04AB?sequence=1

[5] National Library of Medicine (NLM) Digital Repository Evaluation Selection Working Group. 2008. Recommendations on NLM digital repository software. Retrieved on 2015 April 16 from http://www.nlm.nih.gov/digitalrepository/DRESWG-Report.pdf

[6] University of Houston Libraries Digital Preservation Task Force. 2015. University of Houston Libraries Digital Preservation Policy. University of Houston Libraries. Retrieved on 2015 November 15 from http://info.lib.uh.edu/sites/default/files/docs/DigPresPolForWeb_20150917.pdf

[7] National Library of Australia (NLA). 2013. Digital preservation policy, 4th edition. Retrieved on 2015 April 20 from https://www.nla.gov.au/policy-and-planning/digital-preservation-policy

[8] The Inter-University Consortium for Political and Social Research (ICPSR). 2012. ICPSR digital preservation policy framework. Retrieved on 2015 April 20 from http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/preservation/policies/dpp-framework.html

[9] The Ohio State University (OSU) Libraries. 2013 August. Digital preservation policy framework. Retrieved on 2015 April 20 from http://library.osu.edu/documents/SDIWG/Digital_Preservation_Policy_Framework.pdf