Title: Beyond the Repository: Integrating Local Preservation Systems with National Distribution Services
Type: National Leadership Project Grant (2 year/$233,000)
Lead Organization: Northwestern University Libraries (PI: Evviva Weinraub)

Abstract
Many cultural heritage institutions have established repository systems to preserve the important digital content either generated or collected by their constituencies. An essential aspect of digital preservation is risk management. To aid in mitigating risk factors associated with lack of geographic diversity, lack of technological diversity, and loss of data related to human activities and systems failures, several Distributed Digital Preservation (DDP) services, such as Academic Preservation Trust (APTrust), Chronopolis, The Digital Preservation Network (DPN), LOCKSS, and MetaArchive have emerged. As these services have matured, the problem of integrating data from a local repository with a distributed service has not been resolved. In 2016, Northwestern University and the University of California, San Diego were awarded an IMLS National Leadership Grant, LG-72-16-0135-16, to identify technical solutions to bridge the gap between local repository systems and dark, distributed preservation systems. This was a planning grant intended to articulate the questions many institutions are grappling with surrounding the integration of local services with distributed preservation networks, identify broadly-applicable approaches, and propose high-level technical solutions.

The grant team and the advisory board coalesced around three recommendations based on the research conducted during the planning grant (which included a survey that had 170 complete responses and twelve in-depth interviews from selected survey respondents). The first recommendation was to develop a decision-making toolkit for choosing materials to send to Distributed Digital Preservation systems. The second was to determine a shared BagIt profile which would increase interoperability between DDP systems and repositories. The third recommendation was to create the specifications for a shared API that can be used by service providers to create dashboards with data extracted from the DDP BagIt profile, allowing consumers of the DDP service the ability to track distributed resources. All of this work is in the spirit of furthering the sentiment of the [Digital Preservation Declaration of Shared Values](#) and builds on the IMLS National Digital Platform aims to "bridge gaps between disparate pieces of the existing digital infrastructure for increased efficiencies, cost savings, access, and services."

Activities
- In collaboration with the National Digital Stewardship Alliance (NDSA), form a working group to create a decision-making toolkit for choosing materials to send to DDP systems. (Month 1 – 12)
- Determine and implement a shared BagIt profile that will increase interoperability between DDP systems. (Month 1 – 24)
- Develop an API specification that will allow institutions to create a dashboard with data extracted from appropriate elements in the DDP BagIt profile, for local tracking. (Month 12 – 24)
- Travel (Various Months)
  - Travel for representatives from identified DDP systems and the user community to meet three times in Evanston, IL
  - Travel for members of the NDSA working group to meet twice in Evanston, IL
  - Travel for appropriate representative to present work at relevant preservation conferences

Deliverables
- A toolkit for institutions to use to characterize and identify objects to send to DDP systems.
- The agreement upon and implementation of a shared BagIt profile specification to be adopted by all identified DDP systems in the United States.

- Development and release of an API specification for use by local institutions in tracking information deposited into U.S. based DDP systems.

Project Resources: Personnel, Time, & Budget

Two teams will be dedicated to this project over the course of the two-year grant. One team will be dedicated to creating the toolkit (NDSA participants) and the second will work on the shared BagIt profile and API. The NDSA team will meet twice in person in Evanston, IL at Northwestern University Libraries to create the toolkit, and they will also meet regularly virtually in order to meet their goals. We expect their work will take approximately one year to collect and write up appropriately. Stakeholders from the identified DDP projects, as well as three representatives from the user community who have been doing work in this area, will meet three times in Evanston, IL at Northwestern University Libraries to agree upon a shared BagIt specification. Funding from this grant will be used to pay for local development work within the various DDP systems to implement the API specification so institutions can track their content. Two institutions will test the specifications by sending bags to their DDP service providers. Funds from this grant will also be used towards national travel to attend conferences and meetings that are directly relevant to the dissemination of this work such as PASIG, SAA, Open Repositories, DLF, DigiPres, NDSA, iPres, and ICDP, among others.

We are requesting funding ($94,000) for travel for approximately nineteen members of the two teams to attend face-to-face meetings hosted at Northwestern University. These funds will cover airfare, hotel, ground transportation, and per diem for all members of the teams. Funding ($75,000) for development work at the local DDP systems to create the new specifications. Travel funds ($16,000) to attend relevant conferences to promote the work and share the results with other DDP services worldwide. The travel and conference funding will be a subcontract with Educopia for an estimated total of $122,000. Total cost requested for the project is estimated at $233,000 (direct cost: $197,000; indirect costs: $36,000).

Personnel

- Evviva Weinraub (Primary Investigator) AUL for Digital Strategies, Northwestern University
- Bag-IT/API Team
  - Sibyl Schaefer - Digital Preservation Analyst for Research Data Curation at UCSD, Chronopolis Program Manager
  - Carolyn Caizzi, Head of Repository & Digital Curation, Northwestern University
  - Andrew Woods - Fedora Technical Lead, DuraSpace
  - Bill Branan – Services Technical Director, DuraSpace
  - Dave Pcolar – CTO, Digital Preservation Network (DPN)
  - Michael Ritter - Developer Chronopolis & DPN (University of Maryland)
  - Rosalyn Metz - Director of Library Technology and Digital Strategies, Emory University
  - Sam Meister - Preservation Communities Manager, Educopia Institute
  - Katherine Skinner – Executive Director, Educopia Institute
  - Justin Simpson – Director of Archivematica Technical Services
  - Bertram Lyons – Senior Consultant, AVPreserve
  - Nicholas Taylor – Program Manager, LOCKSS
  - John Kunze – Identifier Systems Architect, California Digital Library
  - Andrew Diamond – Senior Software Engineer, AP Trust
  - Tim Dilauro – Data Archivist, Johns Hopkins University
- Curation Toolkit Team
  - Laura Alagna - Digital Preservation Librarian, Northwestern University
  - Lauren Work - Digital Preservation Librarian, University of Virginia and NDSA Content Working Group co-chair
  - NDSA Working Group – 4 people; to be selected