Beyond the Repository: Promoting Interoperability in Distributed Digital Preservation

Laura Alagna
laura.alagna@northwestern.edu
@digitized_laura

LG-72-16-0135-16
Grant overview

• Two-year IMLS National Digital Platform grant

• Based on 2017 IMLS planning grant of the same name
  ◦ Under the planning grant we surveyed digital preservation practitioners to uncover obstacles to integrating local repositories with distributed digital preservation (DDP) services
Previous research

• Research questions:
  ◦ How does one curate objects to ingest into a long-term dark preservation system?
  ◦ How does versioning of objects and metadata play out in a long-term dark preservation systems and how to automate these actions?
  ◦ How can systems that store data differently be made more interoperable?

• Our research revealed:
  ◦ The main obstacles to interoperability are technical and curatorial
  ◦ Final report: https://doi.org/10.21985/N28M2Z
Obstacles to interoperability

• Technical
  ◦ Transferring content from a local repository to a DDP system is complicated
  ◦ Survey respondents indicated that they often accomplished this via workarounds, or cobbling together a number of different tools

• Curatorial
  ◦ In general, survey respondents can only select a subset of preserved content to send to a DDP, but face difficulty in making selections
Grant goals

• To create a shared BagIt profile to increase interoperability between repositories and DDP services
• Local repository systems and DDP services will develop functionality to accommodate the new BagIt profile
  ◦ Fedora and DSpace will implement the shared profile in their BagIt export feature
  ◦ LOCKSS and APTrust will update their systems to be able to accept the shared BagIt profile
  ◦ Users will test the new functionality, provide feedback, and any necessary updates or enhancements will be made based on this feedback
• To create a curation toolkit that will assist cultural heritage organizations in choosing materials to send to distributed digital preservation (DDP) systems
Year 1 summary

- New “BtR” BagIt Profile has been created
  - See it at https://github.com/dpscollaborative/btr_bagit_profile

- Development work to implement the new profile into DSpace, Fedora, Academic Preservation Trust, and LOCKSS has been planned

- The Curation Toolkit is undergoing final revisions, and will be made available online soon
The grant team built on existing work on bags:
- BagIt File Packaging Format or (https://tools.ietf.org/html/rfc8493)

New additions in the BtR bag format:
- Manifests-Allowed
- Tag-Manifests-Allowed
- Bag-Producing-Organization
- Data-Export-Info

Important note: all of this work, like the BagIt File Packaging Format spec and the BagIt Profiles spec, is agnostic to the payload in the bag’s data directory
Curation Toolkit

- Meant to provide users at various stages in their digital preservation programs with resources they need

- Not meant to be a linear flow of steps or an exact workflow; instead it attempts to reflect the reality of digital preservation work and maturity stages at a variety of institution types

- Organized into three modules: Getting Started, Moving Forward, Leveraging Progress
  - Within modules are resources and frameworks for hands-on activities, such as Partnering with IT and Other Providers, Prioritizing Content for DDP, Developing a Digital Preservation Policy, etc.
Milestones & next steps

- All in-person meetings have been completed in 2019
- Software development to implement the BagIt Profile is underway
  - Initial development completed by early June 2020
  - Testing and iteration will take place from July - November 2020
- The Curation Toolkit is undergoing final revisions, and will be made available online soon
Thank you!

Laura Alagna, grant project manager
Digital Preservation Librarian, Northwestern University
laura.alagna@northwestern.edu
@digitized_laura