Chronicles in Preservation:
Building Institutional Capacity in Digital Preservation

Matt Schultz
Educopia Institute
matt.schultz@metaarchive.org

Mark Phillips
University of North Texas
mark.phillips@unt.edu

Nick Krabbenhoeft
Educopia Institute
nick@metaarchive.org

Stephen Eisenhauer
University of North Texas
stephen.eisenhauer@unt.edu

Building Capacity

Standards help organizations establish better digital preservation practices. However, fulfilling every best practice and requirement is beyond the resources of most organizations. The Chronicles in Preservation project is evaluating how to improve the preservation readiness of digitzed and born-digital newspapers on a spectrum of essential to optimal levels of conformance to digital preservation standards. The project is publishing the *Guidelines for Digital Newspaper Preservation Readiness* (available for public review at http://publishing.educopia.org/chronicles), developing and testing a set of Interoperability Tools for improving existing collections, and evaluating a series of test exchanges with three Distributed Digital Preservation (DDP) systems in the *Comparative Analysis of DDP Frameworks.*

Identifying Solutions

Surveying the Partners

The project asked project partners to describe their existing collections and systems according to:
• Collection & Repository Information
• Collection Data Management
• Preservation Assessment
• Ingest & Recovery

The variety of responses indicated that tools must be lightweight and easy-to-use.

Collecting Sample Data

The project also examined sample data from project partners, discovering a variety of:
• file naming schemes
• object identifiers
• folder structures
• amounts of metadata

Tools must be adaptable to each organization's policies.

Interviewing Stakeholders

The project talked to a number of newspaper stakeholders including a social media reporter, a newsroom librarian, a state archivist, and a university librarian.

The interviewees requested better resources to make sense of standards, published methods to incorporate standards into existing workflows, and easy-to-use tools.
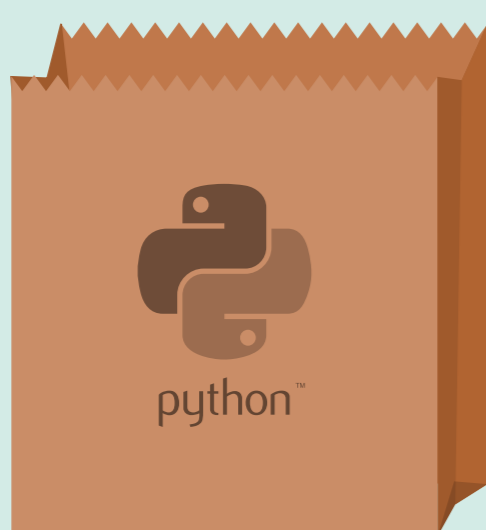
Tools

BagIt Data Model

Bags are a simple data model. They consist of:
• a data directory
• a manifest listing files and checksums
• a file with metadata about the bag

The BagIt specification does not impose a file organization structure, so newspaper holders can package collections as-is. Bag creation tools also accomodate a range of skill levels with both command line tools such as bagit.py and graphical user interface tools such as Bagger.

To promote the usage of BagIt with better documentation, the project will publish a set of simplified BagIt Usage Instructions and a suite of scripts for enhancing the use of BagIt for the preservation of digital newspapers and other digital content in early 2014.

PREMIS Creation and Usage

The project extended the use of two PREMIS creation utilities.

Format Descriptions Service - a web service developed by the FCLA that performs batch file format idenitifcation using JHOVE and DROID.

The project developed workflows to integrate technical metadata into collection bags.

UNT PREMIS Event Service - a web service developed by UNT to record server actions such as the copying of files and fixity checks in PREMIS objects.

The project is helping to prepare this tool for release with an open source license in early 2014.

Preservation Readiness Plans

The project tested how to prepare a collection for ingest into DDP systems. Four actions were considered essential for preparation:
• recording administrative metadata
• creating a file inventory
• recording file checksums
• identifying file formats

The BagIt and PREMIS tools accomplish these goals, but preparing a collection is more complex than applying tools to data. The project formalized a workflow for each partner with a standards-based Preservation Readiness Plan.

A generalized version of these plans will be published in early 2014 for other institutions to reference.

Digital Preservation Business Case Toolkit
http://wiki.dpconline.org/

NDSA Levels of Preservation

The project evaluated the metadata tools with the Levels of Preservation, a lightweight digital preservation metric published by the NDSA (National Digital Stewardship Association) in 2013.

The project tools address the metadata requirements in Levels 1 and 2, including:
• Creating file fixity
  (Data Security - Level 1)
• Creating an inventory of content
  (Metadata - Level 1)
• Storing administrative and transformative metadata
  (Metadata - Level 2)
• Inventorying formats in use
  (File Formats - Level 2)

The common philosophies of Chronicles and the Levels demonstrate that the community is building practical tools.

http://www.digitalpreservation.gov/ndsa/activities/levels.html