Digital Preservation for Small Repositories

Documentary Heritage Services for New York

Bonnie Weddle
March 14, 2018

Additional Resources

Data Accessioner

To install version 1.1 of Data Accessioner, download the zipped package of Data Accessioner files from **http://dataaccessioner.org**, unzip the files, and save them to a folder on your processing computer's internal hard drive (e.g., "C:\DataAccessioner"). To start Data Accessioner, go into the bundle of unzipped Data Accessioner files you've saved on your computer, double-click on the folder named "Data Accessioner.jar", and double click on the "start.sh" or "start.bat" file. Data Accessioner may take a few seconds to open.

Data Accessioner cannot process audio CD's that contain CDA files (most commonly found on commercially produced CD's) or DVD discs that are configured to play video in DVD players. It can process CD's and DVD's that house data files.

Data Accessioner requires the use of Java, which may already be installed on your computer. If your computer doesn't have Java, you'll find it at **https://java.com/en/**

Seth Shaw, **http://dataaccessioner.org**

> Shaw created Data Accessioner while working for the Duke Univerisity Archives, and older versions of the software were named Duke Data Accessioner. This site furnishes background and a link to download Data Accessioner version 1.1.
>
> Instructions for using a slightly older version of Data Accessioner:
> **http://dataaccessioner.org/downloads/da-1-0-0/DataAccessionerGuide_v1_0.pdf**

Moving Beyond Data Accessioner

Other Tools

Ricky Erway, *You've Got to Walk Before You Can Run: First Steps for Managing Born-Digital Content Received on Physical Media*,
https://www.oclc.org/content/dam/research/publications/library/2012/2012-06.pdf

Very short and very practically oriented OCLC Research publication that details how to locate, survey, and transfer to stable storage the unprocessed backlog of digital content on the portable electronic media found in paper collections.

Lara Friedman-Shedlov and Carol Kussman, "Demystifying Digital Records Processing: Step-by-Step, Byte-by-Byte," **http://digitalcommons.macalester.edu/libtech_conf/2015/sessions/60/**

Information about a variety of open-source (i.e., free to download, use, and modify) tools that can help you identify duplicate files, generate checksum/fixity information, rename files en masse, delete empty folders, and perform other processing tasks. Includes detailed information about using an older version of Data Accessioner and Hash My Files, a checksum generator.

Gretchen Gueguen, "Management of Born Digital Special Collections," **http://www.aserl.org/wp-content/uploads/2013/04/Intro_DP_2013-3_Born_Dig_Docs.pdf**

2013 Association of Southeastern Research Libraries webinar that discusses inventorying portable media, technological dependences, and setting up a relatively sophisticated processing workstation.

Digital Preservation Software, Consortia, and Services

Archivematica, **https://www.archivematica.org/en/**

Open source suite of software tools that supports the processing and preservation of digital content. It was developed to conform to the Open Archival Information System Reference Model. It runs only on Linux computers; however, if you have some technical savvy and willingness to learn, you can install Oracle VirtualBox (**https://www.virtualbox.org**) to create a "virtual" Linux computer on a Windows or Apple desktop or server and run Archivematica within the virtual computer.

ArchivesDirect, **http://archivesdirect.org**

ArchivesDirect is a cloud-based version of Archivematica available to subscribers.

MetaArchive Cooperative, **https://metaarchive.org**

International digital preservation network network composed of libraries, archives, and other cultural heritage institutions. Each member is responsible for maintaining its own hardware and storage resources and using shared software to host its own data and preserve data created by other, physically distant members of the cooperative.

Preservica, **https://preservica.com**

Preservica is a digital preservation repository that was developed to conform to the Open Archival Information System Reference Model. It addresses all facets of digital preservation. Customers can purchase licenses to install Preservica software on servers they own and maintain or to use Preservica's cloud-based subscription service.

Self-Assessment and Planning

Megan Phillips, Jefferson Bailey, Andrea Goethals, and Trevor Owen, "The NDSA Levels of Preservation: An Explanation of Uses," http://ndsa.org/documents/NDSA_Levels_Archiving_2013.pdf

Overview of the National Digital Stewardship Alliance Levels of Preservation, " a tiered set of recommendations on how organizations should begin to build or enhance their digital preservation activities," and discussion of its development. Includes chart summarizing the levels.

Northeast Document Conservation Center, Planning for Digital Preservation: A Self-Assessment Tool, https://www.nedcc.org/assets/media/documents/DigitalPreservationSelfAssessmentfinal.pdf

Four-page list of questions. Keep in mind that even the largest and best-funded institutions

Preservation-Friendly File Formats

National Archives [United Kingdom], *Selecting File Formats for Long-Term Preservation*, **http://www.nationalarchives.gov.uk/documents/selecting-file-formats.pdf**

Brief explanation of the characteristics of preservation-friendly formats. Does not include detailed format recommendations.

Core Theoretical Readings

Brian Lavoie, *The Open Archival Information System (OAIS) Introductory Guide (Second Edition)*. **http://dx.doi.org/10.7207/TWR14-02**

The Open Archival Information System (OAIS) Reference Model is a profoundly influential theoretical standard that outlines all of the roles, responsibilities, and functions associated with preserving digital content over the long term. The standard itself is extremely long and incredibly detailed, and Lavoie succinctly covers its essential points. If you're new to OAIS, start here.

Research Libraries Group and OCLC, *Trusted Digital Repositories: Attributes and Responsibilities,* **https://www.oclc.org/research/activities/past/rlg/trustedrep/repositories.pdf**

> Product of a groundbreaking effort to define the core attributes of and responsibilities assumed by repositories that reliably preserve data over the long term. It can be a bit overwhelming if you're new to digital preservation or work for a repository that has extremely limited resources, but it can also be very useful when you're ready to start thinking about purchasing digital preservation services or making the case for increased resources to senior management.

Continuing Education

Online Tutorials

Cornell University Library and Inter-University Consortium for Social and Political Research, *Digital Preservation: Implementing Short-Term Strategies for Long-Term Problems*, **http://www.dpworkshop.org/index.html**

> Self-paced, outlines the tenets of digital preservation and the basic components of a digital preservation program.

North Carolina State Archives and North Carolina State Library, **http://digitalpreservation.ncdcr.gov/tutorials.html**

> Short video tutorials and recorded webinars focusing on file naming conventions, cloud computing, digital preservation, preserving Facebook data, and other topics.

Organizations providing no-cost online training (live or recorded)

Council of State Archivists, **https://www.statearchivists.org**

New York State Archives, **http://www.archives.nysed.gov/workshops**

San Jose State University School of Information, **http://ischool.sjsu.edu/about/webcasts**