Archivematica
http://www.archivematica.org, Accessed December 14, 2014
Reviewed by Brad Houston, CA, University Records Archivist, University of Wisconsin–Milwaukee

As the importance of long-term preservation of born-digital and digitized records grows, so too does the need for systems to manage those records in a standardized fashion. Archivematica, an OAIS-compliant, open source digital preservation system created by Canadian company Artefactual Systems Inc., is one attempt to put such a system within reach of a broader segment of the archival community. Artefactual Systems issued Archivematica's 1.0 (production) release in January 2014, with several point releases throughout the year to add to the system's functionality.

Archivematica is not itself a digital preservation repository, but is designed to help prepare Archival Information Packages (AIPs) for long-term storage in a number of different repository settings, including a specially hosted DuraCloud service. The system, which runs on a local host server through a web browser, creates AIPs using what Artefactual describes as a "micro-services" model. Archivematica combines custom Python scripts with a number of open-source tools, such as the File Information Tool Set (FITS), ClamAV anti-virus, and the BagIt packaging utility, to move digital objects through these steps with minimal user intervention. At pre-defined intervals, information about the files and the changes being made during processing is extracted and packaged with the objects in a PREMIS-in-METS metadata file, to maintain authenticity and facilitate management within the eventual repository.

The key component of Archivematica for digital preservation is the Format Policy Registry, a server maintained by Artefactual which syncs format information and best preservation practices with individual Archivematica instances. During processing, users can specify whether to normalize files for preservation and/or access. Based on this choice, Archivematica runs a number of scripts through packaged open-source software to convert identified obsolete or unstable formats to established preservation or stable access formats, such as TIFF or PDF/A. The process significantly streamlines the decision-making process for file-level records preservation. Institutions may customize their local version of Archivematica to normalize to alternate formats, or to convert file formats not identified for normalization, though extensive customization requires knowledge of Python or Linux Bash scripting. Artefactual does offer fee-based technical support, training, and consultation to assist institutions with customization as needed.

In addition to packaging digital objects for preservation, Archivematica can also create Dissemination Information Packages (DIPs) for upload into institutions' access systems. Archivematica comes packaged with Artefactual's own ICA-AtoM access system, and includes protocols to upload DIPs directly to ContentDM instances. There are many opportunities during processing to add Dublin Core metadata to entire collections, as well as to individual digital objects or folders, thus improving object searchability on the access system of choice. Other metadata schema are not directly supported by manual entry, but Archivematica includes metadata import functionality to allow users to map large datasets to the Dublin Core defaults.   Even if not using an external access system, objects within AIPs are indexed by ElasticSearch, making it possible for archivists to find objects within Archivematica and serve them to users as requested.

Although Archivematica greatly simplifies the digital preservation process, it is not without its frustrations. Most of these stem from the requirement that Archivematica run in a LAMP (Linux, Apache, MySQL, PHP) environment, under the Ubuntu 12.04 distribution (as of this writing the latest point version of Archivematica was being tested in Ubuntu 14.04). Archivists-cum-administrators without experience in UNIX-based systems or in running Apache web servers, or without an allied IT professional with experience in same, will run into difficulties configuring the system. To its credit, Artefactual maintains an open Google Group community of active and potential Archivematica users, which can be extremely helpful in answering basic questions about the system. Documentation for the system is also available on the Archivematica site, although much of it assumes a certain level of knowledge in scripting and server maintenance as well.

For institution-specific installation and technical support, Artefactual will provide services on a fee basis. "Regular" users will have fewer problems once the system is up and running, though here the "micro-services" model shows a key weakness: if certain services late in the process fail or output incorrect metadata, the user's only recourse is to remove the AIP and start over, often at the cost of hours of processing time.

Despite these problems, Archivematica provides a valuable and timesaving service to any institution that can afford the support package or has the technical expertise to manage a production instance. By bundling together numerous open-source tools in an easy-to-use interface and providing a selection of ready-made processing scripts, Archivematica allows archivists with even modest technical knowledge to process electronic records according to best practices, thereby greatly increasing the efficiency and effectiveness of institutional digital preservation.