

Data management plan

The Internet Archive serves as a trusted digital repository for all data in the Pop Up Archive system. Pop Up Archive bundles each audio file with an XML file containing standardized metadata that is then stored as a record at the Internet Archive. The mission of the Internet Archive is to offer “permanent access for researchers, historians, scholars, people with disabilities, and the general public to historical collections that exist in digital format.” Material uploaded to the Internet Archive is transcoded into lower fidelity formats (.wav, .mp3, .aif, .ogg) to achieve compatibility with more systems and, thus, wider distribution. When files are updated in Omeka, they will simultaneously be updated at the Internet Archive. Users choose whether or not they would like to index the material they store with Internet Archive; however, in order to prevent accidental deletion, users must contact the Internet Archive to remove individual files or collections.

Digital files are stored using the Internet Archive’s hardware, which consists of PCs with clusters of IDE hard drives. Data is stored on DLT tape and hard drives in various appropriate formats, depending on the collection. Web data is received and stored in archive format of 100-megabyte ARC files made up of many individual files. Alexa Internet (currently the source of all crawls in Internet Archive collections) is proposing ARC as a standard for archiving Internet objects. The Internet Archive is also mindful of the obsolescence of digital formats. They note, “As advances are made in software applications, many data formats become obsolete. We will be collecting software and emulators that will aid future researchers, historians, and scholars in their research.” (<http://archive.org/about/about.php>)

Pop Up Archive has also evaluated the LOCKKS model for preservation, and is currently acting under their advice that the Internet Archive provides adequate backup storage. However, our team is open to reconsidering LOCKKS as a storage mechanism for the software artifacts created over the course of this project. We are also working with the Knight Foundation and PRX to safeguard and ensure the continued existence of the software we engineer. All software will be made accessible on GitHub for further development using an open-source license.

The Pop Up Archive system outputs PBCore-compliant XML records that can be distributed to other institutional repositories such as the Library of Congress (LOC) and the California Digital Library (CDL). The Pop Up Archive team will also explore the feasibility of a one-click plug-in for data exchange with the LOC and CDL. Supplementing distribution through individual producers’ and oral history archives’ websites with distribution through the repositories of these significant cultural institutions will expand the number of probable users of archival audio artifacts to the tens of thousands.