ScholarsArchive@OSU
2023 Preservation Assessment Report

The goal of the preservation assessment is to proactively evaluate the contents of OSU's institutional repository, ScholarsArchive@OSU (SA@OSU) within local and broader contexts, in order to enable OSULP to ensure the availability of our content over time. More information can be found in the ScholarsArchive@OSU Preservation Policy.

The first annual ScholarsArchive@OSU Preservation Assessment was conducted in May and June 2023. Since this was the inaugural preservation assessment, extra time was spent establishing procedures and capturing those as part of local documentation.

◆ ◆ ◆

Part 1: Inventory Repository File Formats

◆ Compile an annual inventory of file formats stored in the repository, and plan to address format migrations if indicated.

As of May 2023, SA@OSU held a total of **78,281** files.

PDFs constitute **94.03%** of SA@OSU files, with 761 PDF/A and 72,846 standard PDF files in the repository.

In total, the inventory shows:

➜ **60** unique MIME types
➜ **108** unique file extensions

Compared against the current confidence table in the SA@OSU Preferred File Formats Guide, excluding PDFs, the repository holds:

➜ 1,548 files rated Highest Confidence
➜ 2,300 files rated Medium Confidence
➜ 583 files rated Lowest Confidence
➜ 203 files with unrated file formats

97 erroneous filesets were identified and scheduled for remediation or deletion.

No files were indicated for format migration during this year's assessment.

◆ ◆ ◆

Part 2: Update Preferred File Formats Guidance

◆ Review and make appropriate revisions to the ScholarsArchive@OSU Preferred File Formats Guide to ensure that the confidence table remains up to date with current national and international standards and best practices.

The ScholarsArchive@OSU Preferred File Formats Guide informs users about file formats that are more likely to be accessible and usable over time. During the preservation assessment, updates were made to this guidance to reflect current standards as well as the reality of the content held in the repository.

Notable updates are listed below, and full details are provided in change log files attached to the Guide. During this year's assessment, **no listed formats were removed or had their rating revised.**

➔ Language was added to the introduction about long-term versus near-term use, and about OSULP's commitment to format migration.
➔ The formatting of the confidence table was updated for a more readable presentation.
➔ Based on guidance from national and international organizations and peer institutions, two Highest Confidence, one Medium Confidence, and one Lowest Confidence formats were added to the confidence table.
➔ Based on the SA@OSU file format inventory:
  ◆ A section was added to the confidence table for ebooks;
  ◆ Five Highest Confidence, three Medium Confidence, and two Lowest Confidence formats were added to the confidence table;
  ◆ Several alternate file extensions were added to listed formats.

Resources consulted during the update:
● Library of Congress, Recommended Formats Statement
● Open Preservation Foundation, International Comparison of Recommended File Formats
● Digital Preservation Coalition, Bit List
● Preferred or recommended file formats guidance from peer institutions:
  ○ University of Washington
  ○ Duke University Libraries

◆ ◆ ◆

Part 3: Review Fixity Reports

◆ Review repository fixity reports and restore relevant files from backups as needed.

Fixity reports give an indication that the files stored in the repository are unchanged since their submission. Fixity checks for ScholarsArchive@OSU run automatically on the 15th of the month. No fixity check failures have been reported, and no restorations from backups were needed during this year's assessment.

The most recent fixity report was produced on June 16, 2023.
➔ 153,936 files were checked;
➔ 153,936 files were reported as PASSED;
➔ The process took 9 hours, 12 minutes, and 23 seconds.

During the preservation assessment, it was discovered that expected email reports from fixity checks are not being received. This issue was corrected by the OSULP LIT team in June 2023. Further analysis is needed to understand slight discrepancies in the number of files checked compared with the number of files identified in the file format inventory.

◆ ◆ ◆

Part 4: Update Preservation Policy

◆ Review and make appropriate revisions to the ScholarsArchive@OSU Preservation Policy.

Recommended updates to the ScholarsArchive@OSU Preservation Policy were submitted to the SA@OSU Users Group and subsequently incorporated into the public ScholarsArchive@OSU Policies page in June 2023. The majority of the 2023 changes consist of minor formatting revisions, and of removing specific implementation details that are covered in procedures and other documents. Meaningful changes are limited to:

➔ Clarifying that, as a required format, standard PDF files are included in the commitment of future usability and thus would be migrated if the format were deprecated;
➔ Including the production of a report as part of the Annual Preservation Assessment.

◆ ◆ ◆

Resources

ScholarsArchive@OSU Home Page: https://ir.library.oregonstate.edu
ScholarsArchive@OSU Policies: https://wiki.library.oregonstate.edu/confluence/x/bivWAw
ScholarsArchive@OSU Preservation Policy:
      https://wiki.library.oregonstate.edu/confluence/display/RP/ScholarsArchive@OSU+Po
      licies#ScholarsArchive@OSUPolicies-PreservationPolicy
ScholarsArchive@OSU Preferred File Formats Guide:
      https://guides.library.oregonstate.edu/Scholars-Archive/PreferredFileFormats
OSU Libraries & Press Home Page: https://library.oregonstate.edu

Submitted June 2023 by Cara Key
Digital Repository Librarian, OSULP