ScholarsArchive@OSU 2024 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 annual ScholarsArchive@OSU Preservation Assessment was conducted in May 2024.

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 2024, SA@OSU held a total of **80,214** files.

PDFs constitute **94.14%** of SA@OSU files, with 936 PDF/A and 74,580 standard PDF files in the repository.

In total, the inventory shows:

➔ **60** unique MIME types
➔ **110** unique file extensions

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

➔ 1,602 files rated Highest Confidence
➔ 2,314 files rated Medium Confidence
➔ 580 files rated Lowest Confidence
➔ 196 files with unrated file formats

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

Compared with the 2023 inventory, total SA@OSU files increased by 1,933. Of this growth:
➔ 1,909 (98.76%) are PDFs
➔ 54 (2.79%) are Highest Confidence formats
➔ 14 (0.72%) are Medium Confidence formats

Numbers of Lowest Confidence and unrated file formats decreased (-3 and -7 respectively).

Thanks to work throughout the year, filesets missing format information dropped from 40 items to 6; other types of errors dropped from 57 items to 24.

The most common types of files deposited to ScholarsArchive@OSU during the past year were:

✓ PDF - 1,734
✓ PDF/A - 175
✓ CSV - 17
✓ DOCX - 17
✓ MP4 - 13

*Difference in file extension counts since the previous assessment year inventory.

◆ ◆ ◆

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 are 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 a change log file attached to the Guide.

➔ Guidance from national and international organizations and peer institutions reflected little change in the format landscape since the previous assessment year. Given this, **no formats were removed from or reclassified in the confidence table.**
➔ Based on the SA@OSU file format inventory, one Medium Confidence format was added to the confidence table.
➔ An abbreviation was added to a format entry for clarity.
➔ Language was added to the introduction about the scope of the Guide.
➔ A new Review section was added to convey the annual file format recommendation review.

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.

Overview

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, and an auto-generated report is delivered to the ScholarsArchive@OSU email inbox. In 2023 the Digital Repository Librarian established a practice of recording the monthly fixity report data in a tracking sheet to make this information available to other internal users.

Between June 2023 and May 2024:
- ➔ Fixity checks ran 9 times. The automated fixity checker was disabled between November and January, but the service was restored in February by LIT developers.
- ➔ No restorations from backups were needed during this year's assessment or at any time throughout the year.

	JUN	JUL	AUG	SEP	OCT	FEB	MAR	APR	MAY
Files checked[1]	153,936	154,513	154,869	155,489	155,983	157,268	157,390	157,607	157,943
Failures	0	0	0	0	0	14	14	12	12
Duration	09:12:22	09:40:41	07:19:06	08:56:49	08:56:33	09:14:21	06:55:06	05:52:57	07:12:21

Analysis

14 failed items were reported in the February 2024 fixity check. All 14 items were investigated with the same conclusion: The failed items should not be in the repository. In each case, a file had been deleted intentionally through the user interface during the course of deposit review; the files no longer appear in the user interface, but the fileset objects are persisting in the database.
- ➔ These items have been ticketed as evidence of a broader system issue in which some repository objects resist deletion by standard means (GH2572).
- ➔ Since the larger issue is unresolved and the fileset objects have not been deleted, the same 14 failures were reported in the March 2024 fixity check, and 12 of the same 14 were reported in the April and May 2024 fixity checks. We can expect that they will continue to be reported until the larger issue is resolved and the files are successfully removed from the database.

◆ ◆ ◆

---

[1] Fixity checks count checksum audit logs rather than fileset objects. The file format inventory counts the latter, hence the large discrepancy in the numbers.

Part 4: Update Preservation Policy

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

The Digital Repository Librarian conducted a review of the ScholarsArchive@OSU Preservation Policy and deemed that the policy was current. Input was invited from the SA@OSU Users Group and no recommendations were proposed. The Preservation Policy is unchanged as a result of this year's review.

Note that the URL for the public ScholarsArchive@OSU Policies page changed due to the OSULP wiki migration in February 2024.

◆ ◆ ◆

Resources

ScholarsArchive@OSU Home Page: https://ir.library.oregonstate.edu
ScholarsArchive@OSU Policies: https://osulp.atlassian.net/wiki/x/DABSAw
ScholarsArchive@OSU Preservation Policy:
> https://osulp.atlassian.net/wiki/spaces/RP/pages/55705612/ScholarsArchive+OSU+Policies#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 May 2024 by Cara Key
Digital Repository Librarian, OSULP