Preservation Action Plan - Journals

Introduction

This document describes the preservation plan for journal content in the Scholars Portal repository. Most of the journal content is scholarly articles. The preservation plan for journal content follows from policies and practices described in the Preservation Strategic Plan and the Preservation Implementation Plan. This document explains practical steps that Scholars Portal takes to preserve the intellectual content of journal articles in digital format. It outlines the basic tools, methods, and standards used for the long-term preservation of journal content.

Content Formats

For the preservation of journal content, Scholars Portal requires PDF versions of the content, and publisher-supplied XML or SGML (in a format agreed upon between SP and the publisher) containing descriptive metadata and full-text content is strongly encouraged. PDF and XML/SGML conform with Scholars Portal's criteria for preferred formats outlined in the Preservation Implementation Plan. Some journal content includes supplementary image files, audio or video files, or data files in various digital formats. Scholars Portal accepts a wide range of well-known, commonly used formats, but cannot necessarily commit to the 'full' preservation level for formats that do not meet the respository's criteria for preferred formats. These objects will still be maintained at the 'bit-level' preservation level. Scholars Portal continuously monitors developments in file formats to determine if and when formats require migration (see Environmental Monitoring of Preservation Formats).

SIP Format

Scholars Portal works with publishers to determine and define the format of each SIP before publishers submit content (see Definition of SIP).

Analysis on Ingest

Upon ingest, every file in the repository is subject to identification of its file format using DROID and validation of that format using JHOVE. During the process of DROID identification, a file format is associated with each file, and, where possible, the file is linked to the format's entry in PRONOM, the British National Archive's format registry. The outputs of these processes are recorded in the preservation metadata for each file.

Content Excluded

Scholars Portal does not ingest files that are not referenced (either as part of a representation or as supplementary material) in the associated metadata. As the SIP is retained, these files can later be ingested if necessary.

Format Normalization

Upon ingest, the publisher's XML/SGML is converted to a Scholars Portal version of NLM XML. Where possible and when desirable, files that do not conform to Scholars Portal's preferred formats will be converted to preferred formats.

Metadata Normalization

When necessary, Scholars Portal crosswalks metadata from the publisher's XML/SGML to a Scholars Portal version of NLM XML. The repository creates preservation metadata for each file. The preservation level, explained in the Preservation Implementation Plan, is applied to each file upon ingest and recorded in the preservation metadata for each file.

Acceptable Formats

For the Full Preservation level for journal articles, currently acceptable formats include PDF and XML. XML articles may have diagrams in GIF, JPG, TIFF, or PNG format. Articles not in these formats (or those unable to be converted to these formats) may still be preserved at the Bit-level Preservation level.

Supplementary materials will be accpeted in any format, and preserved at the Bit-level preservation level.