

Portico Content Type Action Plan: E-Journals

1. Content Type Introduction

- 1.1. E-journal content is comprised of scholarly journal articles in digital format. This content can include full or header-only information in XML, SGML, or other format; page images in PDF or some other format; figure, formula, table, or other graphic images in various still image formats; and media and other supplementary information files in various digital formats.
- 1.2. E-journal content may also include RSS XML and referenced images and other files.

2. Policy Statement

- 2.1. The fundamental archival unit for e-journal content is an article.
- 2.2. E-journal content as received from the publisher is batched and processed into Submission Information Packages (SIP) by the Portico ConPrep System.
 - o As there is no standard package format for e-journal content, each content provider's content stream package structure is analyzed by Portico staff to determine the directory structure and file naming conventions used by the publisher to associate the components which comprise a single e-journal article.
 - o The results of this analysis are described in the Turnover Document associated with the content stream. The results are also expressed more formally in the XML profile which informs the ConPrep system as it processes publisher content and assembles the components of each archival unit (article).
 - o Analysis sometimes reveals extraneous files that are included in the publisher package, but which do not comprise components of the articles (e.g., thumbs.db files). These files are not ingested into the Portico archive.
 - o The original package will be preserved separately in the Supplied Files for E-Journal Content content-type.
- 2.3. With the exception of these "excluded files," all assets provided by the publisher are validated and characterized with the JHOVE tool, associated with the appropriate article, and are associated with a Portico metadata file containing technical and descriptive metadata, in the archival unit.
- 2.4. Should Portico make the determination that any formats of the assets comprising the e-journal content are at risk, these components will be migrated in conformance with the Format Action Plan for that format.
- 2.5. All e-journal content will be associated with a business artifact comprising the contract authorizing Portico to store and manage those assets.
- 2.6. Portico will maintain descriptive and technical metadata associated with the technical artifact in the Archival Information Package (AIP) containing the technical artifact.

3. Implementation Examples

- 3.1. Portico regularly performs two migrations on e-journal content – transforming the content provider's mark-up to a Portico version of the JATS and transforming the content provider's



packaging into the Portico content model. The Turnover Document associated with each stream of content from a specific content provider is considered the Format Action Plan for these two initial migrations. The migration of the packaging and the migration of the publisher-provided XML and SGML files in the publisher's "native" XML or SGML format to the Portico profile of the JATS XML format are performed for ease of collection management and the delivery needs demanded by the trigger event model associated with this content-type. The publisher-provided XML/SGML file is preserved in the archive along with the other assets comprising the article.

- 3.2. Additionally, for publishers who, in lieu of a single PDF file constituting the page image of an article, provide individual pages as TIFF or other files, Portico processes that collection of individual pages into a single PDF file. Again, the publisher-provided files are preserved along with the PDF file in the archive along with the other assets comprising the article.

4. Document History

- 4.1. Approved by: Amy Kirchhoff
 4.2. Last Review Date: 3/27/2016
 4.3. Reviewed by: Sheila Morrissey, Stephanie Orphan, Kate Wittenberg
 4.4. Change history:

Version	Date	Change	Author
0.1	06/04/2009	Drafted document	Sheila Morrissey
0.2	07/07/2009	Edited and formatted to new template	Sheila Morrissey
0.3	07/27/2009	Incorporate Stephanie's copy-edits	Sheila Morrissey
1.0*	7/29/2009	Synched with Format Monitoring and Migration Policy.	Amy Kirchhoff
1.1*	8/5/2009	Added reviewed by line	Amy Kirchhoff
1.1.1	4/4/2011	Updated the logo	Amy Kirchhoff
1.2*	3/27/2016	Updated to mention JATS and RSS	Amy Kirchhoff, Sheila Morrissey

* An approved version of this document.