

## **Portico Content Type Action Plan: D-Collections**

### **1. Content Type Introduction**

- 1.1. D-collection content consists of digital primary source materials and a variety of born-digital collections that do not fit the standard e-book or e-journal model.
- 1.2. D-collection content can include full or header-only information in XML, SGML, or other format; page images in TIFF, PNG or some other format; figure, formula, table, or other graphic images in various still image formats; web archive files; and media and other supplementary information files in various digital formats.

### **2. Policy Statement**

- 2.1. The fundamental Archival Information Package (AIP) or *archival unit* for d-collection content varies by collection. For example, it may be a single document, a newspaper issue, an image, or even a dictionary entry.
- 2.2. D-collection content as received from the content provider is batched and processed into Submission Information Packages (SIP) by the Portico ConPrep System.
  - As there is no standard package format for d-collection 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 content provider to associate the components which comprise a single unit.
  - 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 provider content and assembles the components of each archival unit.
  - Analysis sometimes reveals extraneous files that are included in the content provider's package, but which do not comprise components of the content to be preserved (e.g., thumbs.db files). These files are not ingested into the Portico archive.
- 2.3. With the exception of these "excluded files," all assets provided by the content provider are validated and characterized with a file format identification tool (such as JHOVE), associated with the appropriate content unit, and associated with a Portico metadata file containing technical and descriptive metadata, in the archival unit.
- 2.4. Portico will evaluate any file formats of the assets comprising a collection to determine whether they are at risk and/or require migration in conformance with the *Portico Format Monitoring and Migration Policy*. Should any file formats require migration, the strategy will be documented in the collection's Turnover Document, which will be used as that content's Format Action Plan.
- 2.5. All d-collection 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 unit containing the technical artifact.

### **3. Implementation Examples**

- 3.1. D-collection content is sometimes treated differently than e-journal or e-book content. Due to how large the collections are and business reasons, Portico sometimes treats the XML formats provided as de-facto standards. During initial analysis in such instances, Portico writes a tool to extract all the descriptive metadata for the archival units from the provided XML files in order to place it into Dublin Core in the Portico preservation metadata file for the item. As with e-journal and e-book content, the content provider's packaging is transformed into Portico's archival unit packaging. The Turnover Document associated with each stream of content from a specific content provider is considered the Format Action Plan for this content. By extracting the descriptive metadata and transforming the packaging into an archival unit, Portico can manage the content in the archive and quickly deliver the content to end users when needed.

#### 4. Document History

4.1. Approved by: Kate Wittenberg

4.2. Last Review Date: 8/7/2023

4.3. Reviewed by: Amy Kirchhoff, Karen Hanson, Kate Wittenberg, Sandra Parr

4.4. Change history:

Version	Date	Change	Author
0.1	3/16/2016	Drafted document	Amy Kirchhoff
1.0*	3/27/2016	Changes as suggested by SO and SM. Approved	Amy Kirchhoff
2.0*	8/7/2023	Updated wording to allow for a wider variety of content providers. Clarification of file format handling. More emphasis on born digital collections. Addition of web archives format.	Karen Hanson

\* An approved version of this document.