



Portico Content Modification and Deletion Policy

1. Policy Statement

- 1.1. Portico will preserve the intellectual content of items submitted to it for preservation in a manner which will allow the item to be recreated or rendered on current or future delivery platforms.
- 1.2. Portico will not modify content delivered by the content provider without thorough review, approval, and documentation.
 - The exception to this is that Portico will correct, as a matter of course, minor errors in the publisher's markup that do not affect the intellectual content of the item being changed and which cause an item to fail to validate against the supplied DTD.
- 1.3. Portico will not delete content ingested into the Portico archive without thorough review, approval, and documentation.
- 1.4. Content provider supplied files intended for ingest will not be deleted from the Portico file system until content is confirmed to be ingested. Deletions from the file system may occur only for content that will not be ingested and when specific criteria are met:
 - Content is delivered for titles not on the licensor annex and the licensor has stated it will not add the titles.
 - Original deliveries are found to be problematic and the content provider redelivers the entire set of content with corrections.
 - When it is determined that delivery of certain problematic content was not intentional and Portico will not receive corrections.
 - Other situations, which are reviewed on a case-by-case basis, with content deletions requiring approval and documentation.

2. Implementation Examples

2.1. Content Modification:

- There are limited circumstances, in which Portico staff may make, as a matter of course, manual changes to publisher content, typically to correct minor markup errors that cause XML or SGML files to be invalid against the supplied DTD. These changes must not affect the intellectual content of the file being changed and may include such alterations as changes to case, adding the closing bracket to an end tag, or moving a tag so that it is properly nested. This class of manual changes should only be made if problems are truly "one off"; otherwise content corrections will be requested from the content provider or one of Portico's tools will be modified to accommodate the anomaly.
- Any other instances of content modification must be thoroughly reviewed, approved, and documented prior to any action being taken.
 - Where it has been determined that a publisher has systematically included extraneous and unnecessary files with its submission packages that may be excluded from ingest into the archive by the profile, all exclusion rules must be approved by the Archive Service Product Manager and outlined in the Turn Over Document.



- Where manipulation of a supplied package is required so that content can be deleted or moved to create a correct package that can be properly ingested, a standard "Modified Original SIPs or Portico AUs" document must be created and approved by the archive service content manager prior to any changes being made.
- Where "preprocessing" of a publisher's supplied packages is required in order to create packages that contain the necessary sets of content to allow the content to be ingested into the archive. For example, some publishers deliver content for a single issue across multiple packages, which also include partial contents for other issues. In these instances, preprocessing is necessary to create discrete units that can be loaded into the Portico content processing system.

2.1. Content Deletion, Portico Archive:

- On occasion, after content has been ingested into the Portico archive it is discovered that a) an incorrect transformation of the content has occurred, b) incorrect information has been associated with the content, or c) some other aspect of the content is found to be incorrect. In such cases, in order to correct the content, it must be pulled from the Portico archive (and delivery site if applicable), so that the necessary corrections to the tools can be made and the content re-ingested to reflect these changes.
- No such deletions of content may be made without thorough investigation and documentation of the problem(s) found and necessary corrections. This information is documented in a standard "Modifying Original SIPs or AUs" document and must be approved by the Archive Service Product Manager prior to any action being taken (see also "Portico Specified Escalation Path for Archive System Problem Resolution").
- The "Modifying Original SIPs or AUs" document requires that the following information be clearly presented:
 - Affected publisher, stream, and, where applicable, affected content set IDs.
 - All locations affected by proposed change (e.g., directory system, archive, delivery site, etc.)
 - All the steps that must be taken to correct the problem
 - A recovery strategy
 - Detailed explanation of the problem and reasoning behind proposed steps to correct it.
 - Complete list of affected content to be pulled from the archive (includes ARK IDs)
 - Supporting documentation (e-mails exchanged, etc. leading to the resolution put forth in the document)

2.2. Content Deletion, Portico File System:

- When content is delivered for titles not on a publisher's agreement and the publisher has confirmed that it does not intend to add the title(s) to the agreement, the publisher content coordinator will request permission from the



archive service product manager to delete the content. The deletion is typically noted in the publisher’s Fact Sheet.

- When a publisher redelivers content to correct problems for content not yet ingested into the archive, typically the new deliveries will replace the original deliveries.
- When it is determined that Portico received unintentional deliveries of content that it cannot ingest, the content may be deleted. Alternatively, such content may be maintained in an isolated directory along with explanatory text, at the discretion of the archive service project manager. In either instance, the action is typically documented in the publisher’s Fact Sheet. (One example of “unintentional deliveries” that Portico cannot ingest is when metadata only is provided for digitized print content for which full-text [PDF] is not available through the publisher.)
- Depending on the nature and breadth of proposed content deletions and determined on a case-by-case basis, content deletions may be documented in a standard “Modifying Original SIPs or AUs” document, which must be submitted for approval to the Archive Services Product Manager.

2.3. The Portico file system used for publisher content is read-only to all but a very limited group of staff:

- Publisher Content Coordinator
- Operations Ingest Team Lead
- Data Team Lead
- Data Team/Operations Team Members with consent of Data Team Lead (for content that requires preprocessing prior to ingest)

3. Definitions

3.1. Preprocessing: When supplied content is not packaged in such a way as to allow Portico’s tools to identify and ingest defined sets of content (e.g. articles, issues, etc.), Portico will reorganize the content and create new packages that will allow the content to be ingested into the Portico archive.

4. Document History

4.1. Approved by: Amy Kirchhoff

4.2. Last Review Date: 8/5/2009

4.3. Reviewed by: Stephanie Orphan, Sheila Morrissey

4.4. Change history:

Version	Date	Change	Author
0.1	6/1/2009	Drafted.	Stephanie Orphan
0.2	6/22/2009	Extensive revision	Stephanie Orphan
1.0*	7/28/2009	Slight mods to make more generic – e.g., “publisher” to “content provider”.	Amy Kirchhoff



1.1*	8/5/2009	Added reviewed by line.	Amy Kirchhoff
------	----------	-------------------------	---------------

* An approved version of this document.