



Harvest Ingest Requirements

Thank you for preserving your materials with the community governed CLOCKSS Archive. Your content will be ingested into the archive via “harvest”. The requirements we articulate below are adapted from Google best practices documents. Compliance will ensure your content is effectively preserved AND optimally available to search engines for discovery.

CONTENT REQUIREMENTS

Site browsability

- Build your site with a consistent link structure. Every object should be reachable from at least one static text link. Objects with single-use URIs are not easily preserved.
- Content that comes from CDNs should have a stable URI.
- Use a text browser, such as Lynx, to examine your site. If features such as AJAX, JavaScript, cookies, session IDs, frames, or other dynamic layout keep you from seeing your entire site in a text browser, then preservation is likely to be impaired.
- Every article and all accompanying material must be reachable by following a text link. List all articles in a volume or year with links to all aspects of the article, as well as supplemental materials, on a single HTML page. This can be the site’s tables-of-contents or some other time-based index that eventually closes.

Site behavior

- Use HTTP 5xx codes to indicate temporary errors that should be retried soon, such as temporary shortage of backend capacity.
- Use HTTP 4xx codes to indicate permanent errors that should not be retried for some time, such as file not found.
- If you need to move your articles to new URLs, set up HTTP 301 redirects from the old location of each article to its new location.
- Use If-Modified-Since and HTTP 304 to better identify content updates

Required metadata

- Metadata needs to be supplied for each article or book: DOI, publisher, publication date, ISSN or ISBN, and as appropriate: publication title, article/book title, series title, volume, issue, and page number.
- Metadata needs to be machine-readable using metadata tags in html or standard metadata formats, like RIS.



Harvest Ingest Requirements

ACCESS REQUIREMENTS

Permissions statements

- As appropriate, content will need one or the other, or both of these permission statements. The Stanford LOCKSS team will tell you where on your site to post these statements.
 - CLOCKSS system has permission to ingest, preserve, and serve this Archival Unit
 - LOCKSS system has permission to collect, preserve, and serve this Archival Unit
 - Note: The LOCKSS permission statement should be displayed when the requesting IP address has permission to view the volume. This statement can be placed on the volume TOC.

IP Addresses

Enable IP address access to all content you wished preserved:

171.66.236.0/24 (i.e. 171.66.236.0 through 171.66.236.255) Stanford University

128.42.174.11 Rice University

128.42.174.12 Rice University

156.56.241.164 Indiana University

156.56.241.166 Indiana University

- Note: You may need to separate these IP addresses from the Rice, Indiana, and Stanford subscription records and set up a new CLOCKSS Archive subscription.