



## File Submission Specifications

Updated: November 19, 2024.

### 1. XML Coding

NLM LitArch requires full text content in XML format. The XML must be valid against a mutually agreed upon Document Type Definition (DTD). Submissions are preferred in the [Book Interchange Tag Suite \(BITS\) DTD](#). For detailed information on using the BITS DTD for submissions to NLM LitArch, please read the [Bookshelf Tagging Guidelines](#).

Alternative DTDs must provide explicit and detailed mark-up of all structural and textual elements of a source, including, but not limited to:

- Bibliographic metadata, such as titles, contributor names, or publication dates
- Organizational units, such as chapters, appendices, sections, or subsections
- Typical textual elements, such as paragraphs, figures, tables, footnotes, or reference citations
- Links, for example from the text body to figures, tables, or cited references

Regardless of the XML DTD used, the following metadata information must be present and tagged with correct values in every XML file:

- Author/Editor (may only be omitted if same as publisher)
- Book Title
- Edition (may only be omitted if it is a first edition)
- Publisher Name
- Place of Publication
- Date of Publication
- Language (if other than English)

NOTE: If you are using the Book Interchange Tag Suite (BITS), LitArch strongly suggests that you tag your documents so that they conform to the PMC Style Checker.

See also additional XML tagging guidance.

### 2. Images

#### 2.1. Content Images

Generally, authors submit raw image data files to a publishing house in various formats (ppt, pdf, tif, jpg, xml, etc.). The files are then normalized to produce print or electronic output. NLM LitArch requires the normalized output, which is high-resolution, and of sufficient width and quality to be considered archival. Images generated at low resolution for display purposes are not acceptable.

- Uncompressed high-resolution TIFF or EPS files are required for all images, or, where these are unavailable, PNG, JPEG, or GIF format, at a resolution equivalent to the original source images published in the document
- All files must be cross-platform compatible
- Do not include thumbnail versions of any full-size images that are deposited in NLM LitArch
- Graphics must be legible throughout all submissions
- Graphics must be submitted in the expected orientation for display

NLM LitArch image specifications correspond to the PMC specification for journal submissions. More specific details on figure, equation, and table image quality can be found on the [PMC Image Quality Specifications page](#).

## 2.2. Cover Image

NLM LitArch displays a cover thumbnail for each book. A high-resolution source image of the cover or an alternative image representing the book should be submitted at time of delivery.

The aspect ratio for the image should be 4:5.

## 3. PDFs

One or more separate **non-password protected** PDFs which directly correspond to either (a) the complete book or (b) to each of the individual chapters in the book should be provided.

If print-quality PDFs are available, please submit them. If the book is not printed, the resolution of the images in the PDF should be no less than: Line Art 800 dpi, Halftones 300 dpi, Color 600 dpi. All fonts used in the file need to be fully embedded. Compression for images should be lossless (zip) or highest-quality JPEG. Illustrations should be encoded as vector data with no erroneous conversion to bitmaps.

## 4. Supplementary Data

NLM LitArch requires all available supplementary material to be submitted in a portable format, such as PDF, .doc, .csv, etc. Supplementary material should not be externally linked from the document text to an online location as a substitute for submission. Supplementary material has been defined to include all of the following:

- Voluminous material that was used to support the conclusions of the narrative, such as a genomic database or any data set which can never accompany a document based on sheer mass
- Material added to the work for enhancement purposes, such as a quiz, a PowerPoint presentation, or videos
- Textual or semi-textual content for which elements cannot be adequately described by the XML schema employed, for example fillable forms, questionnaires, or flowcharts

### 4.1. Video

NLM LitArch expects good-quality video and will downsample for web streaming if necessary. If the meaning of the video is not clear due to low quality, it must be improved prior to submission.

- Preferred Settings:
  - Audio codec: AAC
  - Sample audio bit rate: 128 kbit/s
  - Video codec: H.264
  - Video resolution: 480 vertical lines or better
  - Format: MPEG-4 (mp4) container
- Accepted file formats: mov, avi, mpg, mpeg, mp4, mkv, flv, wmv

Video files larger than 1GB should be split to several episodes, each less than 1GB.

## 5. File Naming and Delivery

NLM LitArch requires that data is named and packaged in a compressed archive, such as a zip file, and transferred to the NLM LitArch FTP site. Please write to [bookshelf@ncbi.nlm.nih.gov](mailto:bookshelf@ncbi.nlm.nih.gov) for an FTP account.

Data files:

- Names of image files and supplemental data files must match the names referenced in the XML file
- The name of the cover image should correspond to a book identifier used in the XML, for example an ISBN or a publisher ID
- Files should be grouped by type in appropriately named subfolders:
  - XML files should be delivered in subfolder named xml
  - Image files should be delivered in subfolder named images
  - PDF files should be delivered in subfolder named pdf
  - Supplementary data files should be in subfolder named suppl
- In case of multiple PDFs and XML files, the PDF and corresponding XML base file names must match
- All file names must be unique within a package
- All files names must contain a valid file type extension
- File names should not contain spaces, and should only contain letters, numbers, dashes, periods, and underscores
- XML file names should not exceed 20 characters in length

ZIP file package:

- NLM LitArch accepts .zip, .tar, .gz, and .tgz
- bzip2 compressed .zip files are not accepted
- The name of the package should, if possible, correspond to a book identifier used in the XML, for example an ISBN or a publisher ID
- Send an email notification to [bookshelf@ncbi.nlm.nih.gov](mailto:bookshelf@ncbi.nlm.nih.gov) for each submission including the compressed file name and the source title
- Do not send files as attachments to an email message
- For corrections please explicitly include what has been changed in the email notification. The name of replacement files (.xml, .pdf, .tif, etc.) must be identical to the original (replaced) file names

## 6. Additional XML Tagging Guidance

In addition to the minimum requirements above, below is guidance on XML tagging issues that arise commonly in content in LitArch.

### 6.1. Copyright Statement

If the content submitted to LitArch is under copyright, then a copyright notice, or statement must be included in the XML. See the [copyright tagging guidelines](#) for detailed tagging information and examples.

### 6.2. Open Access and License Statements

#### Terms.

A **license** is a statement which specifies the copyright permission granted for a document.

*continues on next page...*

*continued from previous page.*

An **open access** (OA) document, in the context of LitArch, is a document that is published with a [Creative Commons](#) license or a similar license that allows any user to redistribute the document without requesting permission from the copyright holder. Additionally, this license may allow or prohibit the creation and distribution of derivative works or modified versions of the document, and/or may limit reuse and redistribution to non-commercial purposes only.

An **open access** document, in the context of LitArch, is NOT simply a document that is freely available at the time of publication.

The following conditions apply to all OA documents in LitArch:

- The applicable license will be clearly indicated in both the XML and PDF versions of each OA document. The license data in the XML must conform to LitArch's guidelines for recording the terms of a license. See License Tagging Information below
- LitArch will include the content (full-text XML and other files) for all OA documents in PMC's Open Access subset ("OA subset")
- Content for documents in the OA subset will be freely available to users for downloading via the LitArch FTP service, and similar services that allow automated downloading of documents
- Reuse or redistribution by LitArch users of content that is in the OA subset will be subject only to proper attribution of the original source and authorship, and to the license included in a particular document

### 6.2.1. License Tagging Information

Any document that is to be open access must indicate that explicitly in the source XML file. For content that is supplied in the Book Interchange Tag Suite (BITS), see the [license tagging guidelines](#) for detailed tagging information and examples.

For content supplied to LitArch in a form other than BITS, the guidelines are:

1. Include a precise summary license statement in the source XML file.
2. If you are referring to any standard license – such as a Creative Commons license – there should be a URI included for an unambiguous reference.

## 6.3. Book and Chapter Excerpts

Book or chapter records may be submitted to PubMed and the citation in PubMed may include an excerpt. See the [excerpt tagging guidelines](#) for detailed tagging information and examples.

## 6.4. Accessibility

NCBI Bookshelf supports XML tagging practices as defined by the [JATS4Reuse Recommendation on Accessibility](#), ensuring that content is accessible to all users, including those who rely on assistive technologies.

Key accessibility practices, as outlined in the [Bookshelf Tagging Guidelines](#), include:

- **Headings:** Use structured heading elements to create a clear document hierarchy that enhances navigation for screen readers.
- **Image Descriptions:** Provide meaningful alternative text for images, figures, and tables, ensuring essential information is accessible to users of assistive technologies.
- **Audio and Video:** Include descriptions or transcripts for multimedia content, making audio and video accessible to all users.

- **Table Structure:** Implement clear headers and logical structuring in tables to ensure that assistive devices can accurately interpret data.
- **Lists:** Use list elements solely for true list content, ensuring proper nesting for clarity, and avoid using list tags for non-list formatting purposes.
- **External Links:** Clearly tag external links to distinguish them for assistive technology users, aiding navigation.
- **Mathematical Content:** Represent mathematical expressions using MathML, following best practices for compatibility with screen readers.

## 7. Tools

For content submitted in the BITS Book DTD and marked up as per the Bookshelf Tagging Guidelines, a number of tools are available:

### 7.1. Online XML Validator

Use the [PMC XML Validator](#) to validate XML files against a DTD. The DTD must be identified with a properly formed DOCTYPE declaration. The book XML document must be uploaded as a single file of less than 1Mb.

### 7.2. PMC Style Checker

The [Online PMC Style Checker](#) is an interactive tool which provides a detailed report of all items in a document tagged using the NLM Book DTD that do not comply with the [Bookshelf Tagging Guidelines](#). The report will list items as either warnings or errors. Errors are required fixes, and warnings are suggested fixes.

### 7.3. Math Previewer

Use the [PMC Math Preview Tool](#) to create a GIF or PNG rendering of MathML or LaTeX code.

## Submission sample

Download a [sample submission file](#) corresponding to [this](#) report. See [Bookshelf Tagging Guidelines](#) for additional examples of fully-tagged samples.