

XML File Submission Specifications



DRAFT ONLY UNDER REVIEW

- [Naming Data Files](#)
- [Compressed File Package Requirements](#)
- [Resubmitted Package Requirements](#)
- [FTP Deposit Requirements](#)
- [Manifest file \(manifest.txt\)](#)
 - [Example Manifest](#)
- [XML Meta Information \(meta.xml\)](#)
 - [Supported Metadata Tags](#)
 - [Bookshelf Meta XML Submission DTD \(book-submit-xml.dtd\)](#)
- [Sample Packages](#)
 - [Books Submission](#)
 - [Chapter Submission](#)

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 for an FTP account. Only publishers who have applied for their content to be included in the Bookshelf archive and meet the Bookshelf's scope and scientific quality standards are eligible for an FTP account. For more information see [How to Include Content in Bookshelf](#).

Naming Data Files

- 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 and PDF file names should not exceed 20 characters in length

Compressed File Package Requirements

- NLM LitArch accepts .zip, .tar, .gz, and .tgz
- bzip2 compressed files are not accepted
- Must be [deposited to FTP](#)

Each submission package must contain the following:

- Source XML file(s) for the content, depending on workflow type. Each submission must contain either a single XML file for the entire book OR a single XML file for a chapter in the book. Only under mutual agreement can multiple XML files be packaged as an entire book submission, in which case a "main_xml" file type should be designated for compiling the book during processing.
- All associated image files referenced in XML submissions
- All supplementary files referenced in the XML submission according to the [Bookshelf Supplementary Data Policy](#)
- A [manifest file](#)
- A [meta information XML file](#)

All these files must be compressed into a single compressed archive file, for example:

"<book-id_name>.tar.gz" for book submissions

OR

"<chapter-id_name>.tar.gz" for chapter submissions

Resubmitted Package Requirements

- The compressed file package name should be identical to the originally submitted package name
- All replaced files must be named identically to the originally submitted file and identified in a [manifest file](#). *No version number should be added. Version numbers should be controlled in the [meta information XML file](#).*
- All IDs in the [meta information XML file](#) must be identical to the IDs in the originally submitted meta information XML file.
- If possible, an addendum or notes file should be included to explain the changes made.

FTP Deposit Requirements

Submission packages should be placed directly in the root directory of each publisher's FTP directory.

Example list of submission packages:

submission_id_name1.tar.gz
 submission_id_name2.tar.gz
 submission_id_name3.zip

Each compressed submission package must contain the following files:

1. [Manifest file](#) (manifest.txt)
2. [Meta information XML file](#) (meta.xml)
3. Files to upload (all content files required for this submission)

Note: Files in the compressed archive package should be placed directly in the archive root; if the archive contains a directory with files, it is the wrong format.

Manifest file (manifest.txt)

The manifest file is a tab-delimited, plain text file that lists all of the other files in the package, The format of the contents of the manifest file is as follows:

<file_type> <file_name>

1. <file_type> is a string literal. The field is case-sensitive and must be one of the types in the table below.
2. <file_name> is the name of the file in the submission packages.

file_type	Description
meta	Required meta information file, described in the next section. A meta file is required for each submission, whether a book or chapter submission.
book	XML file for the book
hybrid	XML file for an alternative book structure (a document that combines book and chapter features); for more information, see: https://www.ncbi.nlm.nih.gov/pmc/pmcdoc/tagging-guidelines/book/genprac.html#bk_struct
fm	XML file(s) for each front matter part in the book.
chapter	XML file(s) for each chapter in the book.
part	XML files(s) for each part in the book (such as volume, section, part).
appendix	XML file(s) for each appendix or annex in the book.
reflist	XML file for a reference list for an entire book.
addendum	XML file(s) for addenda, corrigenda, or errata to a book or chapter.
toc	XML file for a table of contents for chapter-processed submissions only.
supplement	Supplementary portable data files, including videos. Names and file extensions of supplemental data files must match the names and file extensions referenced in XML file submissions.
image	Image files called out in XML submissions. Names and file extensions of image files must match the names and file extensions referenced in the XML.
cover	A cover image file for book submissions.
main_xml	If, under mutual agreement, more than one XML file is packaged for a book submission, designate a master or main_xml file for how to compile the book files during processing.
book_pdf	PDF for download and / or quality assurance. Each XML submission should have a corresponding PDF file for each book and/or chapter(s) files, unless a PDF is not available, because the source is an online HTML page only, in which case a URL should be provided in the book_metadata_xml for the online HTML version of record.
chapter_pdf	PDF for download and / or quality assurance. Each XML submission should have a corresponding PDF file for each book and/or chapter(s) files, unless a PDF is not available, because the source is an online HTML page only, in which case a URL should be provided in the book_metadata_xml for the online HTML version of record.

notes	Special processing instructions, queries, or notes, including, if relevant, a narrative of changes in resubmitted files.
test	Text XML file for QA of a temporary preview

Example Manifest

```
manifest.txt
book 9780429470035.xml
book_pdf 9780429470035.pdf
chapter_pdf front-matter.pdf
chapter_pdf chapter0.pdf
chapter_pdf chapter1.pdf
chapter_pdf chapter2.pdf
chapter_pdf chapter3.pdf
chapter_pdf chapter4.pdf
chapter_pdf chapter5.pdf
chapter_pdf chapter6.pdf
chapter_pdf chapter7.pdf
chapter_pdf back-matter.pdf
meta meta.xml
```

XML Meta Information (meta.xml)

A single XML meta information file of file_type="book_meta_xml" containing book AND/OR chapter meta information for submissions is required to be included in each submission package. It will be validated against the [books-bulk.dtd](#).

Supported Metadata Tags

Element	Attribute	Notes
book-submit	book-id	Populate this with the applicable BOOKID required when submitting a book submission or a chapter submission part of a book
book-submit	chapter-id	Populate this with the applicable CHAPTERID required when submitting a chapter (or book-part) as an independently-processed unit
book-submit	collection-id	Populate this with the applicable COLLECTIONID required when submitting book or chapter content as part of a collection
book-submit	version	If published versions are supported for the chapter submission provide the published version number. If no published version number is provided, it will be by default added as "1".
book-submit	workflow	Must be populated with "xml"
book-submit	submission-type	Either "book" or "chapter" depending on the unit being submitted per agreement with Bookshelf.
book-submit	pmid	Populate this with the PMID if one exists for the submission
book-submit	doi	Populate this with the DOI if one exists for the submission
URL	url-type="citation"	URL for the citation data source
URL	url-type="full-text"	URL for the full-text content
contacts		Wrapper for authors and contact people, including name, email, affiliation, and contact type
person	person-type="author"	Use this option if the person is an author of the content and is not the Reviewer for the submission.

person	person-type="editor"	Use this option if the person is an editor of the content and is not the Reviewer for the submission.
person	person-type="publisher"	Use this option if the person is the publisher of the content and is not the Reviewer for the submission. Publisher contact and email is required for publishers to receive email notifications of submission errors / issues.
person	person-type="supplier"	Use this option if the person is the data supplier of the XML content and is responsible for fixing errors. Data supplier contact and email is required for data suppliers to receive email notifications of submission errors / issues.
person	person-type="reviewer"	Use this option if the person is to be the Reviewer of the submission to address any queries about it during processing. Reviewer contact and email is required if reviewers must receive and approve a preview of submitted content.
grant-information		Wrapper for grant information
grant	id	The full award number
grant	funder	Specifies the supported funder who issued the grant / contract
grant	PI	The main PI or awardee of a supplied grant / contract
keywords		Wrapper information for adding data accession information or genes and substances, as well as publisher supplied keyword terms
previous-version		Populate this with the applicable BOOKID of a previous edition or version of the submission to build a version link between the two units in Bookshelf.

Bookshelf Meta XML Submission DTD (book-submit.xml.dtd)

[books-submit.xml.dtd](#)

```
<!ELEMENT book-submit (URL?, contacts?,
grant-information?, keywords?, previous-version?) >
```

```
<!ATTLIST book-submit
book-id CDATA #REQUIRED
chapter-id CDATA #IMPLIED
collection-id CDATA #IMPLIED
pmid CDATA #IMPLIED
doi CDATA #IMPLIED
version CDATA #IMPLIED
workflow (xml) #REQUIRED
submission-type (book | chapter) #REQUIRED
xmlns:xlink CDATA #FIXED 'http://www.w3.org/1999/xlink'
>
```

```
<!-- ===== -->
<!-- URLs information links to
citations and full text of the submission -->
<!-- ===== -->
```

```
<!ELEMENT URL (#PCDATA) >
<!ATTLIST URL
url-type (citation | full-text) #REQUIRED>
```

```
<!-- ===== -->
<!-- CONTACT information about the people (at least one is required) -->
<!-- ===== -->
<!ELEMENT contacts (person+)>
```

```
<!ELEMENT person EMPTY>
<!ATTLIST person
fname CDATA #REQUIRED
mname CDATA #IMPLIED
lname CDATA #REQUIRED
email CDATA #REQUIRED
affiliation CDATA #REQUIRED
person-type (author | editor |
```

```

publisher | supplier | reviewer)
#REQUIRED >

<!-- ===== -->
<!-- GRANTS information about contract and grant funding -->
<!-- ===== -->

<!ELEMENT grant-information (grant+)>

<!ELEMENT grant (PI?) >
<!ATTLIST grant
id CDATA #IMPLIED
funder ( nih | ahrq | cdc | epa | va | wt ) #REQUIRED
>

<!ELEMENT PI EMPTY >
<!ATTLIST PI
fname CDATA #REQUIRED
lname CDATA #REQUIRED
email CDATA #IMPLIED
>

<!-- ===== -->
<!-- KEYWORDS information about related data -->
<!-- ===== -->

<!ELEMENT keywords (keyword+)>

<!ELEMENT keyword (#PCDATA | sub| sup | b | i)* >
<!ATTLIST keyword
source (nct| bioassay | compound | substance |
publisher) #REQUIRED
accession CDATA #IMPLIED
gene CDATA #IMPLIED
substance CDATA #IMPLIED
casrn CDATA #IMPLIED
term CDATA #IMPLIED >

<!-- ===== -->
<!-- OTHER information about related data -->
<!-- ===== -->

<!ELEMENT previous-version (#PCDATA | sub| sup | b | i)* >

<!-- ===== -->
<!-- FORMATTING -->
<!-- ===== -->
<!ELEMENT sub (#PCDATA | sub| sup | b | i)* > <!-- subscript -->

<!ELEMENT sup (#PCDATA | sub| sup | b | i)* > <!-- superscript -->

<!ELEMENT b (#PCDATA | sub| sup | b | i)* > <!-- boldface -->

<!ELEMENT i (#PCDATA | sub| sup | b | i)* > <!-- italic -->

```

Sample Packages

Books Submission

[9780429470035.zip](#)

Chapter Submission

[cadth-1-175.zip](#)