



FTP Service

The FTP service for the [NLM Literature Archive](#) (NLM LitArch) may be used to download the source files for any book in the [NLM LitArch Open Access \(OA\) Subset](#) and can be used as a source for data mining. The NLM LitArch FTP site is normally updated once a day.

If you have questions or comments about the NLM LitArch FTP site, please write to the Support Center, using the “Support Center” link at the bottom right corner of any Bookshelf web page.

Source Files from the NLM LitArch Open Access Subset

This FTP service may be used to download the source files for any book in the NLM LitArch Open Access subset. The source files for a book may include:

- Files with the .xml suffix, which are XML files encoded in the NLM/BITS DTD
- PDF file(s) if available
- Image files or graphics for display versions of mathematical equations

The URL to access the FTP site is <ftp://ftp.ncbi.nlm.nih.gov/pub/litarch/>

All the source files for a book are packaged in a single .tar.gz file. The FTP site has a 2-level-deep folder (directory) structure. Folder names are randomly generated and the .tar.gz file for a book is randomly assigned to a second-level folder.

Locating OA Source Files

To locate a specific title or set of titles in the OA subset, refer to this file list which is available in 2 formats, text (.txt) and comma-separated values (.csv).

Text: ftp://ftp.ncbi.nlm.nih.gov/pub/litarch/file_list.txt

CSV: ftp://ftp.ncbi.nlm.nih.gov/pub/litarch/file_list.csv

Each list entry includes the following information:

- The fully qualified name of the .tar.gz file for a book, with directory structure
- Citation details:
 - Book title
 - Publisher
 - Publication date
- The accession number for the title which is a unique, persistent ID for the title that can be used to display it in Bookshelf
- Last updated (YY-MM-DD HH:MM:SS)

Example:

```
cc/8d/healthus08_NBK19617.tar.gz    Health, United States, 2008: With Special Feature on
the Health of Young Adults    National Center for Health Statistics (US)    2009
NBK19617    2012-03-12 15:42:52
where
File:    cc/8d/healthus08_NBK19617.tar.gz
Title:    Health, United States, 2008: With Special Feature on the Health of Young Adults
Publisher:    National Center for Health Statistics (US)
Publication date:    2009
Bookshelf Accession ID:    NBK19617
Last updated:    2012-03-12 15:42:52
```

Display a Title in Bookshelf Using the Accession Number

Insert the accession number (accid) to the Bookshelf URL as follows:

```
http://www.ncbi.nlm.nih.gov/books/<accid>/
```

For example, to display

Health, United States, 2008: With Special Feature on the Health of Young Adults

National Center for Health Statistics (US) 2009

Accession number: NBK19617 (from the example above)

Use: <https://www.ncbi.nlm.nih.gov/books/NBK19617/>

Suggested FTP Client Configuration

After a series of experiments using ftp clients with NCBI's FTP server, we've found that the configuration of ftp clients can seriously affect performance. NCBI recommends setting the TCP buffer size to 32Mb. For more information, please see <ftp://ftp.ncbi.nlm.nih.gov/README.ftp>.