



OAI-PMH Service

Updated: July 26, 2018.

The NCBI Bookshelf OAI-PMH service (Books-OAI) provides access to metadata of all items in the Bookshelf archive, as well as to the full text of a subset of these items.

Books-OAI is an implementation of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), a standard for retrieving metadata from digital document repositories. Visit the [Open Archives Initiative site](#) for more information about the protocol and other activities of the OAI group.

Books-OAI supports OAI-PMH version 2.0. It does not support earlier versions of the protocol.

If you have questions or comments about this, or any of the other services provided by the NCBI Bookshelf, please write to bookshelf@ncbi.nlm.nih.gov.

Copyright

Most of the items in this archive are copyright protected, with copyright held by the author(s) or the depositing publisher. In general, the OAI service cannot be used to retrieve the full text of documents in the NCBI Bookshelf. The only exceptions to this policy are for documents that are in the public domain and those that are made available under an Open Access provision (as defined in the [Open Access Subset](#)). See the Bookshelf Copyright Notice for more information.

Using the Books-OAI Service

The base URL for the service is <https://api.ncbi.nlm.nih.gov/lit/oai/books/>.

See the examples, below, for details about how to use this service to retrieve metadata for books and documents.

High-Volume Retrievals

If you are using a script that makes more than 100 requests of any kind, please run it outside of the Bookshelf system's peak hours. Also, please make sure that your system does not make concurrent requests, even at off-peak times. Peak hours are Monday to Friday, 5:00 AM to 9:00 PM, U.S. Eastern time.

The Books-OAI Service supports HTTP compression, and the Accept-Encoding HTTP header must be set to 'gzip, deflate' for the most efficient data transfer. See also [HTTP compression](#) for further information about this capability.

Access to Full Text

Some Bookshelf participants allow harvesting of the full text of all items, others allow it for only some items, and many do not allow it at all. See the [Open Access Subset](#) for specifics.

In addition, the parameter `set=book-open` identifies the complete collection of items in Bookshelf for which the full text may be harvested.

Supported Data Formats

Records may be retrieved from the Bookshelf archive in either the “NCBI Book Tag Set” or the “Book Interchange Tag Set” XML format for metadata or full-text book records. A DTD, W3C Schema, and complete documentation for these formats are available from <https://dtd.nlm.nih.gov> and from <https://jats.nlm.nih.gov/extensions/bits/>. The prefixes are:

- Metadata: `metadataPrefix=nbk_meta`
- Full Text: `metadataPrefix=nbk_ftext`

OAI Books identifiers

The identifiers used by this system are of the form:

`oai:books.ncbi.nlm.nih.gov:nbkid`

where *nbkid* is the numerical portion of the document’s accession ID. For example, the OAI identifier for the document NBK315783 is `oai:books.ncbi.nlm.nih.gov:315783`.

Automatic Segmentation of Large Result Sets

If a `ListIdentifiers` request results in more than 500 hits, Books-OAI will return the first 500 with a `<resumptionToken>` that can be used to get the remaining items.

If a `ListRecords` request results is more than 25 hits for full text or 50 hits for metadata, Books-OAI will return the first 25, or 50 records, respectively, with a `<resumptionToken>`.

Examples

Identify

- Identify the Books-OAI interface.
<https://api.ncbi.nlm.nih.gov/lit/oai/books/?verb=Identify>

ListMetadataFormats

- List all metadata formats.
<https://api.ncbi.nlm.nih.gov/lit/oai/books/?verb=ListMetadataFormats>
- List metadata formats available for a document with the identifier `oai:books.ncbi.nlm.nih.gov:315783`.
<https://api.ncbi.nlm.nih.gov/lit/oai/books/?verb=ListMetadataFormats&identifier=oai:books.ncbi.nlm.nih.gov:315783>

ListIdentifiers – return a fixed number of identifiers per request.

Typically, the response to this request will contain 500 identifiers, and will include a resumption token at `OAI-PMH/ListSets/resumptionToken`.

- List (the first set of) identifiers for all documents released in Bookshelf after 1/1/2014 in metadata format.
https://api.ncbi.nlm.nih.gov/lit/oai/books/?verb=ListIdentifiers&from=2014-01-01&metadataPrefix=nbk_meta
- If the above request returned the resumption token "oai%3Abooks.ncbi.nlm.nih.gov%3A179164!2014-01-01!!nbk_meta!", then the following request returns the next set of identifiers:
https://api.ncbi.nlm.nih.gov/lit/oai/books/?verb=ListIdentifiers&resumptionToken=oai%3Abooks.ncbi.nlm.nih.gov%3A179164!2014-01-01!!nbk_meta!
- List identifier for documents in the setSpec "micad", released between March to July 2012, and available in full-text XML.
https://api.ncbi.nlm.nih.gov/lit/oai/books/?verb=ListIdentifiers&from=2012-03-01&until=2012-07-31&metadataPrefix=nbk_ftext&set=micad

ListRecords – return a fixed number of records per request.

Typically, the response will contain 25 full-text XML records.

- List (the first set of) full text XML for the setSpec "assayguide" released between 3/22/2014 to 6/12/2014
https://api.ncbi.nlm.nih.gov/lit/oai/books/?verb=ListRecords&from=2014-03-22&until=2014-06-12&set=assayguide&metadataPrefix=nbk_ftext

ListSets – return a fixed number of sets per request

- This response will typically contain the first 500 sets. Use the "resumption token" to get the next set.
<https://api.ncbi.nlm.nih.gov/lit/oai/books/?verb=ListSets>

GetRecord – retrieve a record by identifier

- Get record in Bookshelf metadata XML format with an identifier.
https://api.ncbi.nlm.nih.gov/lit/oai/books/?verb=GetRecord&identifier=oai:books.ncbi.nlm.nih.gov:315783&metadataPrefix=nbk_meta
- Get record in Bookshelf full-text XML format with an identifier.
https://api.ncbi.nlm.nih.gov/lit/oai/books/?verb=GetRecord&identifier=oai:books.ncbi.nlm.nih.gov:315783&metadataPrefix=nbk_ftext