



Full Content Indexing Data Submission Guide

Benefits of Submitting Full Content to Bowker

Bowker offers a free program that will allow for improved discoverability of all new and older titles. Simply upload an electronic version of the full book and we will scan the document for keywords and related information. We accept digital or print editions, wherever a PDF or EPUB file is available. This data will be indexed for search. Without making your content viewable, Bowker will show the most relevant items searched, and rank the importance of the word to the work. This indexing will allow Bowker users, over 100 million library patrons, librarians, retailers, and school systems, to find your titles more readily.

Program Governance

- Bowker's program is governed by our Statement of Use, which can be found at http://www.bowker.com/assets/downloads/services/datafeeds_statementofuse.pdf
- Metadata will be extracted from our Books In Print® database. Therefore, please ensure you are supplying ebook and print metadata regularly and in a timely fashion. Metadata and cover image files should be supplied under our [metadata program guidelines](#).
- Please note that metadata such as tables of contents, marketing copy, and author biographies supplied in your regular metadata feed to Bowker are exempt from the terms of agreement for this program. We encourage you to continue to supply the metadata you always have to Bowker.
- Bowker will not resell or make your full content viewable. Our indexing program does not have a view inside option. We strictly take the indexed content (words and phrases that are scored for relevance and frequency) and embed it in our products to allow our users to find the most appropriate items quickly.

Submission Options

Web Portal (for publishers with less than 20 files to submit)

Publishers can upload content directly to our web portal, www.MyIdentifiers.com, after entering their metadata for a book.

- You should obtain a copy of your entire book in a PDF format from your printer or self-publishing service provider after your book has been printed or published as an ebook. Please do not submit draft content or manuscripts.
- PDFs are the only format for full works that can be uploaded through this website. Bowker cannot accept Word documents or design files.
- Each PDF should be named by ISBN (example 9780123456789.PDF).

Here are detailed instructions for this:

1. The username and password that you previously created on the site is fine. Alternatively, you can use your www.BowkerLink.com credentials on this site.
2. Once logged in, under My Account on the menu bar, click on Manage ISBNs.
3. Click on the ISBN you want to update.
4. Upload your PDF to the box under the PDF image.

Bowker welcomes publishers of all countries to participate in the full content indexing program. Bowker customers are worldwide and are often searching for books to buy in their local currency. Publishers located outside of the U.S., however, need to purchase ISBNs from their [local agency](#). While we have opened the MyIdentifiers site to international publishers for this full content indexing program, please do not attempt to use the ISBN functionality of MyIdentifiers.com for this purpose.

FTP (for publishers with less than 20 files to submit)

Using our secure FTP site enables us to set up an automated process to retrieve and process the files. Bowker will set up a separate FTP account for full content indexing. Please do not supply PDF files to the FTP site used for metadata and cover image delivery.

We ask that you please supply a few sample files for the Bowker IT department to review, prior to sending regular submissions. The submissions can be sent to Patricia.Payton@Bowker.com, and please indicate that you wish to participate in the full content indexing program. Once you supply test files via email, your FTP account will be set up and instructions will be supplied to you via email.

While our preferred method of upload is FTP, we can support the transfer of content in many ways, including FTP, DVD, and USB drive. Please email if you prefer to send the files via another means.

Format Options and Security

File Format

PDF or EPUB is the preferred format for full text. We can also work with XML files, or most other formats. Each file should be named by ISBN.FileExtension.

Files are first run through Optical Character Recognition (OCR) software. Bookmarks or other formatting is appreciated, but not mandatory. Because of the OCR step, Bowker will also accept predominantly illustrated works. In this case, we will use the text only, including captions, to embed in our products.

File Content

For titles where you have produced several formats, Bowker only needs to receive a single file. Since we link formats within our system to title and author, our product will be programmed to tie in all related product ISBNs. In this case, you can choose whether you name the file by the hardback, paperback, or ebook ISBN.

Bowker is interested in indexing both frontlist and backlist titles. Because many libraries already own copies of your works, this is an opportunity to enhance the metadata patrons see, which can potentially increase circulation. Also, the more metadata you have available for your titles, the more discoverable they will be to potential buyers, regardless of the age of the product.

Limitations of indexing: This program serves nonfiction titles well, as our lexicon is based on factual content. While we also accept fiction titles into this program, the breadth of keywords indexed is typically smaller than for nonfiction titles. Foreign language titles are also accepted, but, as with fiction, the lexicon is limited and produces fewer keywords.

Security

Files do not leave our internal server. Files will be dropped to a separate FTP where we control the number of staff that has access. Further, all files are automatically removed from the FTP within 10 days of original posting.

Files are not accessed directly by customers searching our products. Only the XML string extracted from the content will be placed in the programming code for our products. This means that Bowker will not apply digital rights management to your titles, as it is not necessary for indexed-only content. Nor will we show your actual book content within our products.

Bowker does not destroy old files, they do remain on the server for indexing purposes. Our system's lexicon will always continue to grow with the amount of content received. In turn, re-indexing the content against an enhanced vocabulary allows new keywords and word strings to be identified from existing PDFs. This ensures our customers are getting the best results when searching, so we continue to reprocess files on a regular basis. Should you need to replace a PDF due to errors in the original file, just upload the new content with the original file name, and it will be overwritten.

Timing

Content is updated based on your publishing cycle. We are ingesting content and indexing continuously, so we can accept new files whenever you release them to us.

In Summary

The full content indexing program offered by Bowker allows for enhanced discoverability for your titles through keywords. We do not resell your content, nor do we show it on a preview basis to our customers. It is as simple as sending the PDF files you have created for other view inside programs, or the EPUB files created for ebook sales.

If you would like further information on indexing, or any of our products and services, please contact us at Bowker Publisher Relations:

Patricia Payton (Patricia.Payton@bowker.com)

