

# Book Analytics Service - Data Onboarding Background

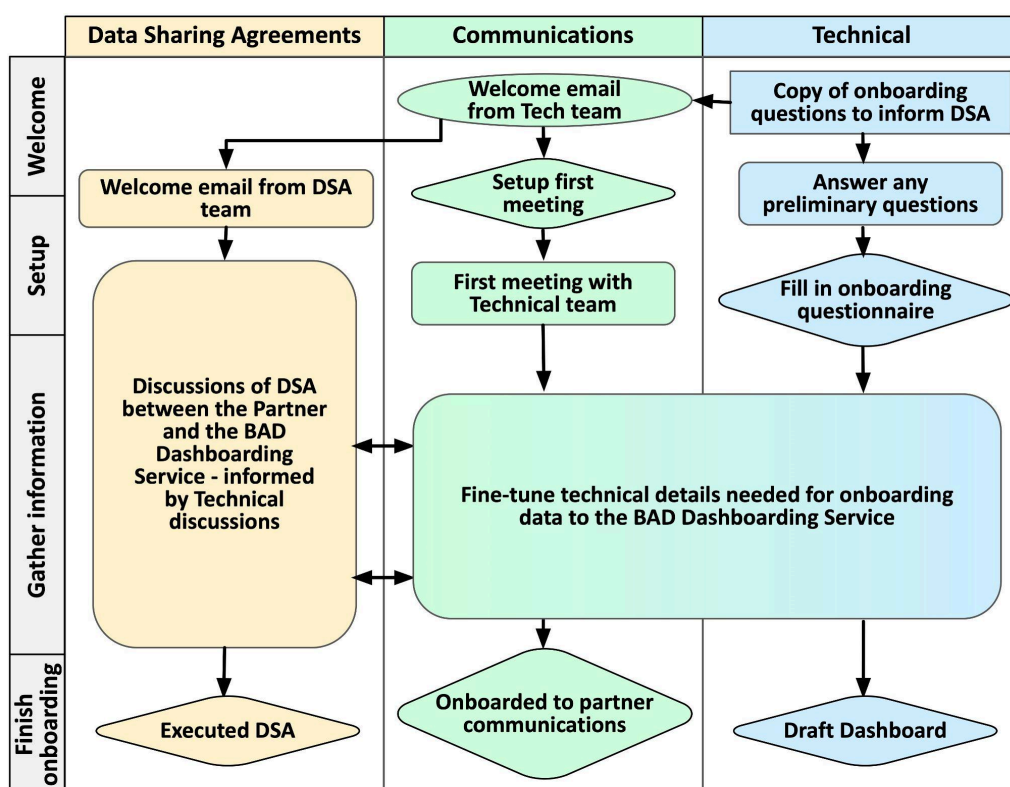
Welcome on board for the Book Analytics Service (BAS)! The purpose of the data onboarding process is to gather technical and access information about the sources of your organisation’s book usage data so that we can import this data into the BAS system. The process includes:

1. Telling us about your book usage data sources
2. Giving us access to your book usage data sources
3. Telling us of any changes you need to the default data fields that map book usage data to Views, Downloads, etc.
4. Branding and other key information about your organisation
5. The parallel process of the Data Sharing Agreements (DSAs)

At any stage, if you have any questions please contact the BAS technical team at: [info@book-analytics.org](mailto:info@book-analytics.org)

## Overview of the onboarding process

During the onboarding process you will need to action the technical steps outlined in the purple-shaded boxes further ahead in this document, as well as fill in the onboarding questionnaire that we will send you. Alongside this data onboarding process we will be in contact with you to discuss data sharing agreements (DSA) and other non-technical parts of the onboarding process.



## BAS background

BAS is the main output of the [Book Analytics Dashboard \(BAD\) project \(2022-2025\)](#), which is focused on creating a sustainable OA book-focused analytics service. During the current phase, we are scaling workflows, infrastructure, and customer support; and developing a long-term plan for housing, maintenance, and funding of the analytics service as a sustainable community infrastructure. The BAD project is led by the [Curtin Open Knowledge Initiative \(COKI\)](#) at [Curtin University](#) in partnership with [OAPEN](#) and [Educopia](#), and is generously supported by the [Mellon Foundation](#) and [COARD](#). This project builds on the momentum we established in our [OAeBU project \(2020-2022\)](#).

## 1. Data sources

For each data source, the questionnaire that we send you will guide you through the information that we need to onboard your organisation’s data. In the sections below there is some background information to each data source, and some steps that you will need to take at your end.

If you are providing an ONIX data feed to the BAS project, the first step is to provide an email address associated with a Google account so we can set up your access to the BAS password manager:

### **STEP 1: Provide contact details for the BAS password manager (Google's Secret Manager)**

The BAS technical team requires the email address of the person who will be configuring your organisation's ONIX feed - this email needs to be associated with a Google account. Please send contact details for this person to:

`info@book-analytics.org`

### **STEP 2: Confirm your access to the Secret**

We will add the user to the BAS password manager for the contact that you provide, and will email you the link to access the Secret containing the credentials to configure your ONIX feed. Please let us know if you do not receive an email from us with these details. Once you have access to the credentials, you will be able to proceed to step 3.

Note, Steps 1 and 2 are not required if your ONIX data feed is obtained from the OAPEN Library (for example, in the case of book collections from multiple publishers) or from Thoth Open Metadata.

## **1.1 ONIX feed**

The core data in BAS is the [ONIX](#) data feed containing metadata for your organisation's books, which we use to link usage data to individual book ISBN-13s. ONIX is a standard that book publishers use to share information about the books that they have published. Having a valid [ONIX 3.0](#) feed is a requirement for including your organisation's book usage data in BAS. As part of the onboarding process we will verify that your organisation's ONIX feed is suitable for use in BAS, including using [ONIX checking software tools](#).

There are currently three methods to provide the ONIX data feed:

1. **Via SFTP feed** - for most organisations who have the metadata of their books in the ONIX 3.0 standard, configuring your ONIX feed to be sent to the BAS SFTP server is the preferred method (see 1.1.1).
2. **Via the [OAPEN Library \(for book collections\)](#)** - If your book collection is hosted on the OAPEN Library, we can automatically extract the collections metadata in ONIX format from the OAPEN website (see 1.1.2)
3. **Via [Thoth Open Metadata](#)** - If your books metadata is hosted on Thoth Open Metadata, we can automatically extract the books metadata for a given publisher in ONIX format using the Thoth API (see 1.1.3).

### **1.1.1 If your organisation has your books metadata in ONIX 3.0 standard:**

The ONIX 3.0 standard requires an ISBN-13 to be transmitted with the record. If you wish to aggregate your organisation's book data by "Work" or "Family" then your organisation's ONIX feed must also contain valid data in the optional "RelatedWorks" field:

- **ISBN-13** - one per book format (note that titles may not be unique)
- **Work** - groups all ISBN-13s from the same edition
- **Family** - groups all editions/works

BAS uses an SFTP server to gather your ONIX data. SFTP (SSH File Transfer Protocol) is a secure file transfer protocol. To send your organisation's data to a dedicated folder on our SFTP server you will need to configure your ONIX using the steps below.

### **STEP 3a: Configuring the supply of your ONIX feed**

The credentials you need to configure your organisation's ONIX feed are stored securely using Google's Secret Manager. Please configure your ONIX feed from your ONIX supplier (e.g. ONIXsuite, BooksoniX etc.) as follows:

- **Server:** `sftp.book-analytics.org`
- **Username:** <your SFTP username as captured in your organisation's Secret>
- **Password:** <your SFTP password as captured in your organisation's Secret>
- **SFTP Port:** 22

**Please send your organisation's ONIX feed with the following settings:**

- **ONIX version:** ONIX 3.0. Please contact us if you do not use ONIX 3.0
- **Encoding:** UTF-8
- **File naming:** Name the files sequentially or by date
- **Schedule:** Please schedule sending the feed fortnightly
- **Completeness:** Even when there is an update/correction, please send the entire record every time (NOT incremental updates), preferably once a week
- **Images:** You do not need to send any cover images with the feed, although it is OK for you to send them

**ONIX data - Minimum data field requirements**

Onboarding ONIX data into BAS requires the following minimum data fields:

- **ISBN-13** as the product identifier
- Use the **"RelatedWork"** field to indicate related works. **"Manifestation of"** is a good relationship to use. ISBN-13 is a good identifier to use for related works. Pick one ISBN-13 to represent the work.
- Use the **"RelatedProduct"** field to indicate related products. You could use the **"Alternative format"** relationship to indicate softcover, hardcover, pdf, etc. ISBN-13 is also suitable to use as an identifier for related products. Related products are used to aggregate and link things into 'works' and 'work families' in the BAS workflows.
- To indicate different editions, either use **"Derived from"** RelatedWorks, or **"Replaces"** or **"Replaced by"** or **"Is later edition of first edition"** RelatedProduct codes. For simplicity you may wish to pick one and stick with it.

**1.1.2 If your organisation does not have your books metadata in ONIX 3.0 standard, but does have a collection hosted on the OAPEN Library:****STEP 3b: Confirm the URL of your book collection hosted on the OAPEN Library**

We will ask you to confirm the URL of your book collection hosted on the OAPEN Library website. This will be used to automatically extract the book collection's metadata in ONIX format.

**1.1.3 If your organisation has your books metadata hosted on Thoth Open Metadata:****STEP 3c: Confirm your publisher ID on Thoth**

Thoth is a free, open metadata service that publishers can use as a solution for metadata storage. We will ask you to confirm your Thoth publisher ID (which can be found using Thoth's [GraphiQL API](#)).

**1.2 JSTOR - partner usage data**

[JSTOR](#)<sup>1</sup> is a digital library, offering over 7,000 open access eBooks. Publisher usage reports offer details about the use (views and downloads) of eBooks by title, institution, and country.

In the case of a book collection (with titles from multiple publishers), we will speak with you directly to facilitate access to JSTOR usage reports for book collections.

**STEP 4: Give BAS access to your JSTOR data**

BAS is able to gather monthly JSTOR usage data reports. Details of the data can be found on their [provider website](#).<sup>2</sup> To give BAS access to your organisation's JSTOR data, please contact JSTOR to give permission to provide 'admin' access to the following service account email: [data@book-analytics.org](mailto:data@book-analytics.org)

<sup>1</sup> <https://about.jstor.org/librarians/books/open-access-books-jstor/>

<sup>2</sup> <https://support.publishers.jstor.org/hc/en-us/articles/360043921594-Books-at-JSTOR-Publisher-Reports>

## 1.3 Google Books - partner usage data

[Google Books](#) is one of the sources of usage data available in BAS. The Google Books Partner program enables selling eBooks through the Google Play store and offering a preview on Google Books. The program makes books discoverable to Google users around the world on Google Books. When readers find a book on Google Books, they can preview a limited number of pages to decide if they're interested in it. Readers can also follow links to buy, borrow, or download the book when applicable.

### **STEP 5: Give BAS access to your Google Books data**

As an eBook publisher you can download usage reports on Google Books data from the *Google Books Partner Center*. Of the three report types available, BAS uses data from the:

- Google Play sales transaction report
- Google Books Traffic Report

BAS uses an SFTP server to gather your organisation's monthly Google Books data. To give BAS access to your organisation's Google Books Partner Center reports you need to follow the procedure explained in the '[Manage additional Partner Center users - Google Play Books Partner Center Help](#)'. You will need to add BAS's `data@book-analytics.org` service account email with the access type 'Analytics & Reports'.

## 1.4 IRUS OAPEN - partner usage data

[IRUS OAPEN](#) is one of the sources of usage data available in BAS. IRUS provides COUNTER standard access reports. Almost all eBooks hosted on [OAPEN Library](#) are provided as a PDF file for the whole book. The original reports show access figures for each month, and the location (IP address) of the access. Within the OAPEN Google Cloud project (which is located in Europe to comply with European GDPR regulations), IP addresses are replaced with de-identified geographical information (city and country). This means that only de-identified geographical information is transferred and stored to and by BAS.

### **Technical steps to onboard your IRUS OAPEN data to BAS**

BAS is able to gather your organisation's monthly IRUS OAPEN usage data via their API without any other steps required by you. Learn more about IRUS OAPEN data on their [website](#).

## 1.5 IRUS Fulcrum - partner usage data

[IRUS Fulcrum](#) is one of the sources of usage data available in BAS. IRUS provides COUNTER standard access reports for books hosted on the [Fulcrum platform](#).

Technical steps to onboard your IRUS Fulcrum data to BAS

BAS is able to gather your organisation's monthly IRUS Fulcrum usage data via their API without any other steps required by you. Learn more about IRUS data on their [website](#).

## **2. Other information**

### **2.1 Branding**

We will need to have a square copy of your organisation's logo to add to your institution's BAS dashboard. The preferred format of your logo image file (e.g. JPG, PNG) is with a resolution of 250 x 250 pixels.

### **STEP 7: Send your organisation's logo to BAS**

Please send your organisation's logo to: `info@book-analytics.org`