

Data Specifications

I. General Information for databases, books, journals, and holdings information

A) Accepted File Formats

- 1) For Licensors Metadata, OCLC accepts MARC, Onix 3.0 ([reference names](#) or long form tags), XML, and Dialog B. Other formats are permissible as agreed by both parties in writing. Documentation should accompany format. Bibliographic records made available through the Service must meet or exceed minimal level MARC standards (level K) as outlined in OCLC's Bibliographic Formats and Standards document: <https://www.oclc.org/bibformats/en.html>.
- 2) For Licensors Full-Text Content, OCLC accepts PDF, epub, and embedded XML. Other formats are permissible as agreed by both parties in writing. There must be a way to link Full-Text to the records. For a PDF or epub file, the naming convention should be parallel (i.e., 1234.xml for the metadata, and 1234.pdf for the Full-Text file). OCLC may not load the Full-Text concurrently with the metadata and will display Full-Text (in accordance with Section 3.5 of the Agreement) at its discretion.

B) Required Links

Each record must contain a link (DOI or URL) to the full document on Licensors website. The link must either lead to (a) metadata representing the item or (b) the item in full form (full-text, streaming video, image, etc.). If a resource-specific link is not available, other options for document fulfillment will be discussed by both parties.

C) Required Information Authentication

OCLC requires authenticated access to Licensors content for testing: Authentication is needed to test the federated search tool in the Service and credentials for testing full record links provided with metadata.

D) Specifications

Licensors will supply physical media and machine-readable form specifications for each Licensors Metadata and full-text file; data and record formats; time periods covered by each Licensors Metadata and full-text file; and a complete written description of each Licensors Metadata and full-text file, including size and the characteristics, contents, and limitation of each field that can be a part of a record in each Licensors Metadata and full-text file; and will incorporate specific Licensors technical and product documentation.

E) Sample Record Set

For review of Licensor's data, OCLC requires an agreed-upon number of records as a representative sample in the file format of the data records to be provided to OCLC.

F) Record Delivery Mechanism

OCLC prefers record upload via sFTP. Other options are available as agreed by both parties in writing. OCLC and Licensor will coordinate initial and succeeding file deliveries.

G) Frequency of Updates

OCLC and Licensor will agree on an appropriate update schedule for each data feed from Licensor.

H) Required Indexing Information

Metadata for every record in each database provided

Unique, stable record identifier	Title	Publication	Publication Year
Publication Volume for Serials/Series	Publication Issue for Serials/Series	ISSN and/or eISSN	ISBN and/or eISBN
Link to full record (DOI or URL)	Pagination	Publisher	

Holdings Information will at a minimum additionally include: embargo period, OCLC holding symbol, coverage dates, mutual customer information (including name and location), date and number of first & last volume and issue online, and coverage depth for Licensor Subscribers.

I) Optional Indexing Information

Metadata for every record in each database provided

Author	Editor/Narrator/Artist	Artist Biography	Language
Abstract/Summary	Keywords	Genre	Awards
Publication	Language	Subject Area(s)	Pagination
Controlled vocabulary	Chapter-level record	Chapter-level abstract/summary	Chapter-level DOI or URL
Document Type	Other standard numbers or identifiers (Please specify ->)	Other (Please specify if provided ->)	OA Indicator (identifying OA content in each record)

II. Licensors' eBooks

A) As to Licensors' eBooks, the following additional terms and conditions shall apply:

- 1) "Licensor's eBook" means Licensor's compilation of eBook records, each of which characterizes the complete, machine-readable representation of a single tome, described in these specifications.
- 2) "OCLC Control Number" means the numerical identifier assigned by OCLC to OCLC metadata records.
- 3) "OCLC Data File" means a data file in a mutually agreed upon format containing pairings between Licensor ID Numbers supplied by Licensor and corresponding OCLC Control Numbers that OCLC is able to determine through the use of its reasonable efforts.
- 4) Within 30 days after making Licensor's eBook Records available via the Service, OCLC will provide Licensor with an OCLC Data File for such WorldCat® eBook Records via a mutually agreed upon method.
- 5) OCLC hereby grants Licensor a nonexclusive, nontransferable right to use OCLC Control Numbers supplied by OCLC in OCLC Data Files in accordance with the "Guidelines for Use of OCLC Control Numbers" below.
- 6) Subject to the rights granted in the Agreement, the OCLC Control Numbers contained in the OCLC Data Files, and all copies thereof that are in Licensor's possession shall be and remain OCLC's property.
- 7) Upon termination of the Agreement for any reason, (i) OCLC will cease making OCLC Data Files available to Licensor, and (ii) the OCLC Control Numbers contained in the OCLC Data Files received by Licensor prior to the effective date of termination may continue to be used in accordance with the Agreement.
- 8) For the avoidance of doubt, the parties agree that OCLC has no obligation under the Agreement to provide metadata other than OCLC Control Numbers to Licensor or any Licensor Subscriber.

B) Guidelines for use of OCLC Control Numbers

Licensor may do the following:

- Print the OCLC Control Number of a record associated with an item in print catalogs of library materials sold by the Partner. If printed, it must be cited in one of the following forms:
 - OCLC#xxxxx
 - OCLCxxxxx
 - OCoLC#xxxxx
 - OCoLCxxxxx
 - (OCoLC)ocmxxxxx
 - (OCoLC)xxxxxwhere xxxxx is the numeric component of the OCLC Control Number.
- Provide the OCLC Control Number associated with an item, cited in the same manner as noted above for print catalogs, in online catalogs.
- Provide a list of OCLC Control Numbers in a spreadsheet or in a similar consistent file to library customers for titles they purchase from that Partner.

- Return the OCLC Control Number to OCLC in WorldCat Cataloging Partner manifests for exact matching purposes.
- Store the OCLC Control Number in internal Partner database(s).
- Add the OCLC Control Number to a MARC record in a field defined to contain local data or a System Control number field (never the 001 field). It must be cited as: “(OCoLC)xxxxx” where xxxxx is the numeric component of the OCLC Control Number. Examples:
 - 035 (OCoLC)4567892349
 - 590 (OCoLC)56789403
 - 945 \$c (OCoLC)61756948

Licensors may not:

- Enter or add the OCLC Control Number to the 001 field in any MARC records.
- Sell the OCLC Control Numbers in list form or any other form to any party. The OCLC Control Number may be included in MARC records in a local field, as noted above, and those (non-OCLC) MARC records may be sold, but the Control Number in and of itself or in any compilation may not be sold.

If Licensors wish to use the OCLC Control Number in a manner not specifically permitted above, Licensors must make a prior written request to OCLC. OCLC will respond in a timely manner to all such requests.

III. Licensors Holdings Information and Licensing and Pricing Data

C) Global Data Format for Licensors Data (Bibliographic Level Data)

This format is based on the KBART Phase 2 [NISO RP-9-201x recommendations](#).

Column number	Column heading	Description
01	publication_title	Publication title for serial or monograph (Conference proceedings series titles are treated as serials while volume titles are treated as monographs.)
02	print_identifier	Print-format identifier (i.e., ISSN, ISBN, etc.)
03	online_identifier	Online-format identifier (i.e., eISSN, eISBN, etc.)
04	date_first_issue_online	Date of first issue available online in ISO format: YYYY-MM-DD or YYYY-MM or YYYY; <i>applies only to serials</i>
05	num_first_vol_online	Number of first volume available online; applies only to serials
06	num_first_issue_online	Number of first issue available online; applies only to serials

07	date_last_issue_online	Date of last issue available online (or blank, if coverage is to present) in ISO format: YYYY-MM-DD or YYYY-MM or YYYY; applies only to serials
08	num_last_vol_online	Number of last volume available online (or blank, if coverage is to present), applies only to serials
09	num_last_issue_online	Number of last issue available online (or blank, if coverage is to present), applies only to serials
10	title_url	Title-level URL; applicable to both serials and monographs For conference proceedings, the title url for the proceedings series and the title_url for each volume should be different.
11	first_author	First author (for monographs only)
12	title_id	Title ID – A unique identifier for this title in the data provider's system (this item must be persistent to facilitate update matching), applicable to both serials and monographs. For conference proceedings, the title_id for the conference proceedings series and the title_id for each proceeding volume should be different.
13	embargo_info	Embargo information; describes any limitations on when resources become available online
14	coverage_depth	Coverage depth (e.g., abstracts or full text)
15	notes	Free-text field to describe the specifics of the coverage policy
16	publisher_name	Publisher name (not to be confused with platform hosting name)
17	publication_type	Serial or monograph. Use serial for journals and conference proceeding series. Use monograph for books, e-books, and conference proceeding volumes.
18	date_monograph_published_print	Date the monograph is first published in print (in ISO format: YYYY-MM-DD or YYYY-MM or YYYY)
19	date_monograph_published_online	Date the monograph is first published online (in ISO format: YYYY-MM-DD or YYYY-MM or YYYY)

20	monograph_volume	Number of volume for monograph, applicable to e-books and conference proceedings. For proceedings, the volume within the conference proceedings series.
21	monograph_edition	Edition of the monograph
22	first_editor	First editor, applicable to monographs, i.e., e-books or conference proceedings volumes
23	parent_publication_title_id	Title identifier of the parent publication. For a conference proceeding volume, the parent_publication_title_id is the title_id of the conference proceedings series.
24	preceding_publication_title_id	Title identifier of any preceding publication title, applicable to serials and conference proceedings series.
25	access_type	Access type may be fee-based (P) or Open Access (F)
26	oclc_number	The OCLC number (if available)

D) Global Data Format for Collection Description File (Bibliographic Level Data)

A collection is a grouping of titles that may be sold to libraries as subject-, year-, or country-based groups of titles that are covered by one subscription. If custom purchases are offered, the WorldCat knowledge base supports a single, large collection that allows a subset of titles to be selected by the library. One file should be sent for each collection or package offered by the provider.

This format is based on the NISO-RP-9-2010, KBART recommendations.

Column number	Column heading	Description
01	collection_name	Human readable name for this collection
02	collection_id	Identifier used for this collection by the data provider's system (sometimes called a database ID). This string must remain static and free of punctuation, and it also must be contained in the collection file name described above.
03	selectable	Yes/No indication of whether individual records can be purchased/selected from this collection by a customer

04	custom_coverage	Yes/No indication of whether the coverage dates (for individual journals) can be changed on a per customer basis
05	open_access	Yes/No indication of whether this collection contains only open access titles
06	pda	Yes/No indication of whether this collection contains only titles available for Patron Driven Acquisitions (also called Demand Driven Acquisitions, PDA and DDA)
07	collection_url	URL pointing to a collection-level page (can be a search form or a listing of titles in the collection) on the provider's site
08	record_count	The number of records in this collection. This is used to help verify the contents of the data files.

E) Library-specific Holdings Data Format (optional, but recommended)

In addition to global-level data, it is also desirable to have customer-specific holdings files that describe the particular subscriptions of each customer. Columns 4 – 10 should only contain values if a particular customer's subscription is different from the data in the global data files.

Column number	Column heading	Description
01	site_id	A site/customer identifier. Supported are OCLC Symbol (https://www.oclc.org/en/contacts/libraries.html) and WorldCat Registry Identifier (https://www.worldcat.org/libraries). If it is not possible to use one of those identifiers or if the identifier is not guaranteed to be available for all customers, a provider-specific customer identifier can be used. If a provider-specific site_id is used, the customer data file (under D below) is required.
02	collection_id	A collection_id that matches to one of the collection_id values in the collections information
03	title_id	A title_id that matches to one of the title_ids in the data file for this collection; should be left blank if the entire collection is subscribed by this customer and there are no overrides in columns 4 – 10.

04	date_first_issue_online	Date of first issue available online in ISO format: YYYY-MM-DD or YYYY-MM or YYYY; for e-books the date of publication
05	num_first_vol_online	Number of first volume available online
06	num_first_issue_online	Number of first issue available online
07	date_last_issue_online	Date of last issue available online (or blank, if coverage is to present) in ISO format: YYYY-MM-DD or YYYY-MM or YYYY
08	num_last_vol_online	Number of last volume available online (or blank, if coverage is to present)
09	num_last_issue_online	Number of last issue available online (or blank, if coverage is to present)
10	title_url	Title-level URL
11	status	Values: ADD/DELETE/UPDATE. Optional field used when sending changes only (not full files) to indicate whether this is a new holding (ADD) or an old holding to be removed (DELETE) or just a changed holding (UPDATE).

Holdings data can be contributed in full replacement files each time or by delivering updates to the previous files. If delivering updates, be sure to use the “status” field to indicate the type of change from previous.

F) Identifier—Data about the Library (optional, but recommended as needed)

This file is used to convey customer specific information when the “site_id” values specified in the holdings information are not one of the two officially supported identifiers or when per customer configuration (account id or custom location id) is required by the provider’s platform in order to link to this content.

Column number	Column heading	Description
01	site_id	A site/customer identifier matching the value in the holdings file
02	site_name	The name of this site/customer
03	oclc_symbol	The OCLC symbol(s) for this library. If more than one, they should be comma separated (i.e., “TRN,SER,OZY”)

04	registry_id	The WorldCat Registry id for this library
05	local_site_id	An identifier for this site/customer used by the data provider's site
06	country	The country this library is located in. Used to help match to OCLC customer record.
07	notes	Any other pertinent information

G) Files Names

There are potentially many different files that will be transferred through this process. Specific file names are required to distinguish between each type of data. The naming scheme to use is:

- Global (title list) metadata – many files: DATA_<provider_id>_<collection_id>.txt
- Collection information file – one file: COLLECTIONS_<provider_id>.txt
- Library-specific holdings file – one file: HOLDINGS_<provider_id>.txt
- Library (about) data file – one file: CUSTOMERS_<provider_id>.txt

Notes:

- “<provider_id>” should be replaced with the agreed upon identifier for the organization providing the data.
- “<collection_id>” should be replaced with the collection_id used in the collections description file.
- The extension “.txt” can be replaced with the appropriate extension, such as “.xml” or “.csv” if providing the data in an alternate format.
- If the data is compressed as a zip file, the name should be: DATA_<provider_id>.zip.

H) OCLC Support for Holdings Information

For support issues, questions concerning data feed formatting, or to request an FTP account, send inquiries to kb-data@oclc.org and include your “provider_id” in the subject line.

I) Licensing and Pricing Data

Licensors’ information for licensing and pricing of Licensors content and Collections and all business models, such as purchase, subscription, DDA (Demand Driven or Patron Driven Access) will be supplied in a mutually agreed format. Licensors is responsible for keeping this information up to date.