

Open Data Requirements and Needs for Shared Print Goals Partnership for Shared Book Collections

202208

Overview

The Partnership is a federation of [these shared print programs](#) with the goal of coordinating collaboration to support cost-effective retention of and access to print book collections to ensure the long-term preservation, accessibility and integrity of these scholarly print resources. One of the Partnership's primary goals is to foster aggregate and openly share retention commitment data from shared print monograph programs and libraries to support cross-program analysis, resource sharing, collection management and other functions. There are new and emerging opportunities in the landscape of library technology and shared print that can support open retention data. Leveraging current and emerging projects and technologies may serve the business modeling needs: open, flexible, low-cost, accessible to a wide range of collection holders, and sustainable.

The Partnership provides these use cases to advocate for platforms which support this data aggregation and to be used as a guideline while evaluating platforms.

Use Cases for Shared Print Open Retention Data Repository

The ultimate goal of the “shared open retention data” effort that was identified at the EAST Shared Print Monograph Summit is to aggregate and openly share retention commitment data from shared print monograph programs and libraries to support cross-program analysis, resource sharing and other functions.

For the purposes of this document, “retention commitment data” or “retention commitments” means information describing that certain library holdings have been committed for long-term retention, usually 10-25 years or more, under a formal shared print program or by a library individually. This data identifies the library, the specific holdings (at the title or item level), the retention period and/or end date of the commitment, the shared print program, and optionally whether and what kind of validation was performed (existence, condition) and level of access and holdings location data (e.g. open or closed stacks and environmental controls).

The sections below provide a very high-level outline of the use cases that are required or envisioned for shared open retention data. See Attachment A for more specific descriptions of the metadata needed for each of the primary use cases.

The primary use cases are similar to the concept of “minimum viable product” used in the software development industry. These are the minimum required functions that must be in place in order to provide a usable service to participating libraries and shared print programs.

Each of these functions should be designed and implemented to support the total number of retentions across all shared print monograph programs in the United States and Canada, currently (2022) estimated at 40 to 50 million retention records.

BOLD = top priority for initial implementation

Summary of top priorities for use of aggregated data

This assumes that libraries or programs are able to submit data to the aggregation.

- **Identify titles at risk (low number of retained copies) and those not at risk (high number of retained copies) and their storage conditions.**
- **Allow libraries to look up retention status of individual titles or a file of titles.**
- **Qualify search for retention commitments based on program or hierarchy of programs.**
- **Support sharing data with other systems.**
- **Expose data that indicates the level of access available for retained holdings (e.g. digital surrogates, standard/in-house/digital circulation, restricted allowable borrowers, etc.).**

Disclosure and Discovery of Retention Commitment Data

Support publicly-available web-based online search to look up the retention status of individual titles:

- **Allow libraries to look up retention status of individual titles or a file of titles.**
- **Qualify search for retention commitments based on program or hierarchy of programs.**
- **Support sharing data with other systems.**

A non-exhaustive list of features speaking to these requirements includes:

- Search by identifier, e.g. ISBN, ISSN, local bib ID
- Search by text data e.g. title keywords, other fields as shown in Attachment A

Holdings Analysis and Decision Support

- **Identify titles at risk (low number of retained copies) and those not at risk (high number of retained copies) and their storage conditions .**
- **Analyze current retention commitments based on bibliographic data (e.g. edition, work families, subject, publication year, language, place of publication).**
- **Identify subject areas of both deep and shallow holdings and retention overlap.**
- **Analyze distribution of commitments by shared print programs, by consortia, by resource-sharing networks.**
- **Quantify retention levels across libraries.**
- **Analyze geographic distribution of commitments.**

- Analyze shelving environments where retained volumes are maintained (e.g., which commitments are held in climate-controlled, access-controlled storage facilities).
- Compare retained holdings to all holdings to show level of duplication/redundancy throughout a given region, the U.S., and the world.
- Compare aggregate retention commitments to all unique print book manifestations in WorldCat to benchmark the community's progress in securing all of those manifestations
- Identify candidates for digitization in support of preservation (low number of retention commitments) or resource sharing (e.g. controlled digital lending).
- Rebalance commitments in the future to reduce unnecessary commitments.
- Allow libraries and network-level analysis to understand the level of validation and known condition issues for retained copies.

A non-exhaustive list of features speaking to these requirements includes:

Predefined reports that analyze retention data to analyze risk and support resource sharing/access:

- Report pre-calculated statistics on retentions by attributes including:
 - overlap count (*i.e.*, retained by 1, retained by 2, etc)
 - program
 - library (OCLC symbol(s) or Marc Organization Code)
 - geographic area
 - level of validation performed
 - reported condition issues

Access

- **Expose data that indicates the level of access available for retained holdings (e.g. digital surrogates, standard/in-house/digital circulation, restricted allowable borrowers, etc.).**
- Enable analysis of resource-sharing activity for items committed for retention/shared print.
- Support interoperability with resource-sharing systems to allow retained holdings to be preferred or deprecated for lending.
- Enable identification of highly accessed copies as candidates for “digitization FOR access.”

Related Functionality

- Facilitate reallocation of retention commitments to other libraries in the same or different programs
- Allow access to aggregated data via an API

Attachment A

Metadata Requirements for Primary Use Cases

This section identifies the metadata for monograph retention commitments that are deemed minimally necessary to support the use cases identified in this document. The categories below are descriptive only; this attachment does not intend to specify the record format(s) or data structure for any of this metadata.

Ingest and retention of full MARC records, while desirable, is not required.

Category	Field	Required	Desirable	Note
Bibliographic and item metadata	OCLC number		Y	OCN is often used as a match point.
	ISBN		Y	
	Local bib ID		Y	For library matching to the local system
	Local item ID		Y	Can be item barcode, e.g. for WMS libraries
	Title (<i>i.e.</i> , 245a)	Y	Y	
	Author (1xx a)	Y	Y	
	Publisher	Y	Y	
	Publication Year	Y	Y	
Institution metadata	OCLC symbol		Y	
	OCLC holding library code		Y	
	State or province		Y	Could be system derived from OCLC Symbol or Marc Organization Code
Retention metadata	Action (for 583 \$a)	Y	Y	“committed to retain”. Support multiple entries if the title was committed to more than one program

	Program name (for 583 \$f)	Y	Y	Name of shared print program (from a controlled vocabulary)
	Action date (for 583 \$c)	Y	Y	Date commitment was made (YYYYMMDD)
	Retention commitment expiration date (for 583 \$d)	Y	Y	Date the retention commitment expires (YYYYMMDD)
	Retaining Institution (for 583 \$5)	Y	Y	
	Holdings retained (for 583 \$3)		Y	Free-text summary holdings field to indicate volumes retained for multipart monographs
Validation metadata	Action (for 583 \$a)		Y	(Optional) i.e. no data necessary if no condition review was performed. If provided, use "condition reviewed"
	Action date (for 583 \$c)		Y	Date review was performed (YYYYMMDD)
	Status (for 583 \$l) (nature of problem, if any)		Y	Type of condition problem found (from a controlled vocabulary)
	Public note (for 583 \$z)		Y	Specific information about condition issues (e.g. "missing pages 10-11" or "significant text marking").