

Transitioning from LDP to Metadb

Metadb is a new streaming data integration platform that will eventually replace the Library Data Platform (LDP 1.x) software used by FOLIO project participants for reporting on FOLIO data. Metadb is still under development. Some institutions are running early versions of Metadb for testing purposes. The Cornell Library is currently running LDP 1.x on its production reporting database to provide users with access to FOLIO data for reporting.

Data updates

The LDP software provides reporting users with a copy of yesterday's data from the FOLIO application server. Metadb will provide continuously streaming data updates from the FOLIO application server. This will provide reporting users with access to FOLIO data in near real time instead of yesterday's data. While LDP 1.x's architecture is limited to running on a single node, Metadb's architecture is horizontal, allowing for linear scaling of data updates. This allows updates to run in parallel within a single Metadb cluster.

Multiple Data Sources

Metadb is designed to be a general purpose data integration platform that supports multiple data sources. This feature will allow the Cornell Library to expand library reporting to include non-FOLIO data from multiple sources in a single reporting database. Additional data sources will also be continuously streaming to provide near real time access to that data for reporting users.

Types of Data sets

In addition to providing data from the FOLIO application server to use for reporting purposes, the LDP and Metadb architecture supports the creation and use of several different types of data sets, extending reporting functionality.

- historical data - history tables are created as data are updated so the LDP or Metadb database retains data that was updated in the past, which can be used for trend reporting
- transformed data - FOLIO stores data in 3 different formats (JSON, ERM, MARC JSON mapped), all of which get transformed into relational data tables as part of the LDP/Metadb ETL (extract, transform, load) process, making it possible make joins and queries across those heterogeneous native data sources
- derived tables - the FOLIO reporting community writes derived table queries that transform the denormalized FOLIO data into joined tables with data arrays broken into columns to both improve query performance and simplify query creation
- external data sets - institutions can import their own data into the LDP or Metadb database, which Cornell has done in bringing select data tables from its legacy Voyager ILS into its LDP reporting database
- user-defined data sets - data analysts can use LDP and Metadb to create their own data tables to be used in queries, which the Cornell Library reporting team has done for queries in areas with large data sets (e.g., annual statistics)

Report Development

Building report queries using the LDP software requires extracting some of the data from data arrays. By design, Metadb tables have data arrays broken out into columns, making it easier for query developers to use.

Special Features

Both LDP 1.x and Metadb include special features of interest to library reporting users: `ldpmarc` and the LDP app.

- Having access to MARC data is essential for Cornell Library reporting users. The `ldpmarc` tool currently converts SRS/MARC records from JSON to a tabular format for LDP 1.x and will be ported to support Metadb.
- The LDP app provides a FOLIO interface for interacting with an LDP database. The Query builder in the LDP app allows you to build a simple query without using SQL with the option to save it for later use. The LDP App is currently available for LDP 1.x and will be included in Metadb in the future.

Borrow Direct move from Relais to Reshare

Cornell's Borrow Direct service is in the process of moving from Relais to Reshare. Institutions that participate in Reshare are running Metadb.

Architectural differences between LDP 1.x and Metadb

	LDP 1.x	Metadb
FOLIO/ReShare support	FOLIO (no ERM)	FOLIO and ReShare
Data update frequency	Once per day	Continuous
Historical data snapshots	Once per day	Continuous
Data sources	Single source	Multiple sources
Scalability	Single node	Horizontal

(Image provided by Nassib Nassar, Index Data)

Resources

- Metadb documentation is available here: <https://metadb.dev/doc/>
- Metadb-based query development for FOLIO: https://github.com/folio-org/folio-analytics/tree/main/sql_metadb
- <https://github.com/metadb-project/metadb>
- [Wolfcon 2022 LDP Presentation](#) - (skip ahead to the 4-minute mark in the recording)
- Metadb software development Roadmap and Issues: <https://github.com/metadb-project/metadb/issues>
- LDP to Metadb table equivalents - https://metadb.dev/doc/#_table_names