

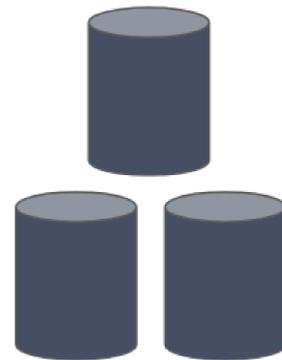
FOLIO's Library Data Platform (LDP)

Kevin W. Walker, PhD

With support from: Nassib Nassar (Index Data), Sharon Beltaine (Cornell University), and Angela Zoss (Duke University)



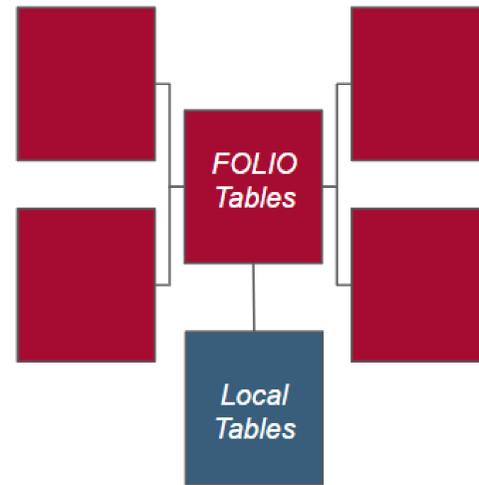
FOLIO apps that feed data to LDP



Real-time data streaming



LDP (data warehouse)

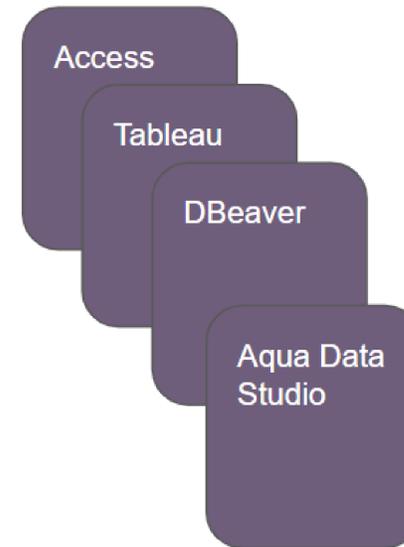


Send queries



Receive results

Connect to LDP via reporting software



Other Data Sources (e.g., University-level student data, assessment system data, etc.)



Background

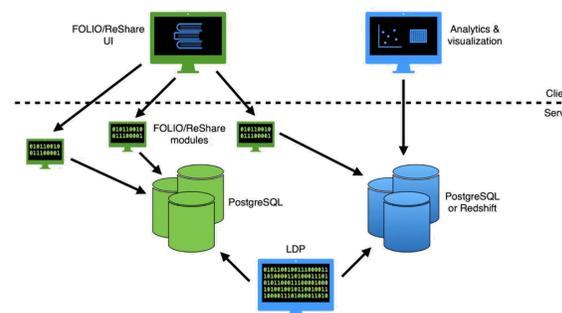
- FOLIO is an open-source library services platform (LSP) utilizing a micro-services architecture
- The Open Library Foundation (OLF) steers FOLIO development
- The Library Data Platform (LDP) is a community-developed, open-source analytics platform
- LDP supports reporting workflows for metadata management, assessment, as well as more advanced workflows and data storage solutions

Development Process

- Development is led by the FOLIO Reporting Special Interest Group (SIG)
 - Librarians and information technology specialists from over 16 research universities
- Development resources provided by Index Data (ID) and OLF
- Community need was defined (data and functionality), then began process of working with developers to design, test, and implement a software solution
- Working groups lead focused efforts to develop support for specialized reporting workflows
 - Metadata Management
 - User Management
 - Resource Access
 - Resource Management / E-Resource Management
- Report prototyping and creation of community-based GitHub repository for report templates
- Beta deployment and testing at the University of Chicago, Duke, Cornell, and the University of Alabama

Design

- Designed specifically with analytics functionality in mind
- Hybrid database warehouse architecture utilizing relational and JSON schemas
- Built atop stable database technologies – PostgreSQL and Redshift
 - Supports both row store and column store data storage models
- Automated extract, transform, and load (ETL) functions currently in development
 - Support for historical data
 - De-normalizing data to reduce the number of joins needed for queries
 - Transforming JSON to relational data
- Extend ETL pipeline to support connecting with databases other than FOLIO

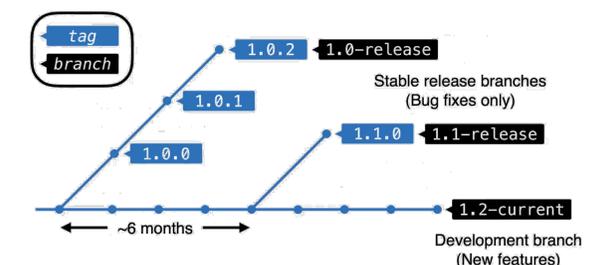


Functionality

- Ad-hoc, cross-domain SQL analytics and reporting
- Stores all historical FOLIO data in support of longitudinal analyses
- Provides scalable platform for integrating multiple subdomain datasets
- Supports advanced data science workflows (e.g., machine learning, predictive analytics, etc.)
- Automatically captures and processes changes to data schemas from diverse, inbound data sources
- Visual query builder (currently in development)
- Support for real-time analytics dashboard creation

Timeline and Versioning

- June 2020: LDP 1.0
- September 2020: LDP 1.1
- December 2020: pre-LDP 1.2
- March 2021: LDP 1.2



More Information

- FOLIO: <https://www.folio.org/>
- Reporting SIG: <https://wiki.folio.org/display/RPT/Reporting+SIG+Home>
- LDP Software: <https://github.com/library-data-platform/ldp>
- LDP Query Repository: <https://github.com/folio-org/folio-analytics>