



Building a Data Warehouse: Boston College's work with Ex Libris to move library data into BC's Enterprise Data Warehouse

Margaret Briand Wolfe - Boston College
Josh Weisman - Ex Libris
IGeLU 2019 - Singapore

About

BOSTON COLLEGE LIBRARIES



- Private, Jesuit Catholic University located in Chestnut Hill Massachusetts, USA
- 9,400 Undergraduates
- 3,500 Graduate / Professional Students
- 860 Faculty
- 8 Libraries: 3.6 Million Volumes 850,000 e-books, 45,000 serials (print & electronic), 800 electronic databases
- Member Association Research Libraries (ARL)
- Development partners with Ex Libris for Alma
- First institution to go live with Alma in July 2012

About Boston College Libraries



3.6 Million
Volumes



183,447 Book
Checkouts



First institution to
go Live with Alma



80,000 Books borrowed
from other Libraries



\$13.6 Million
Annual
Budget



4.5 Million
Uses of
Electroni



\$598 Est. Average Cost of e-
Resource per person / year

What's the task?

Over the past year BC Library Systems has been working with BC IT to move library data into the Enterprise Data Warehouse

- Alma data successfully exported using a variety of Alma tools: Publishing, APIs, Analytics, Webhooks
- ILLiad Borrowing & Lending Data exported using SQL queries





Boston College - Enterprise Data Warehouse





Boston College Enterprise Data Warehouse

Data is sourced from:

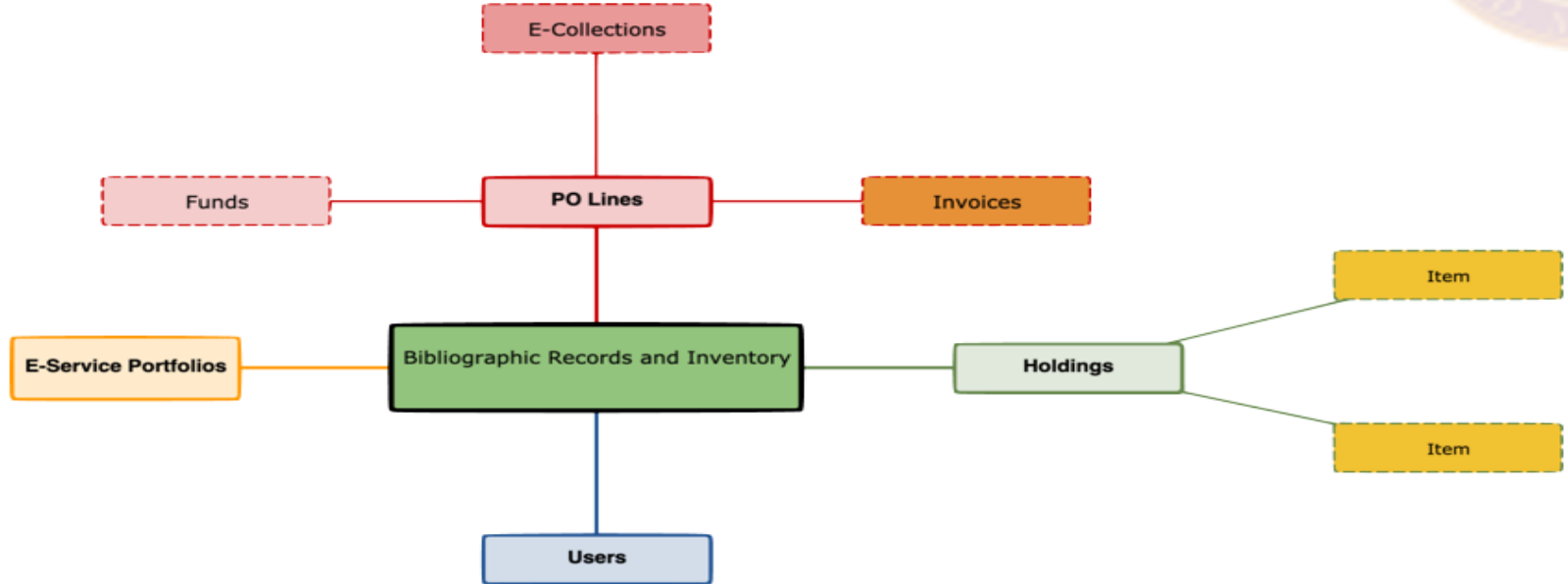
- Student System (University Information System)
- Undergraduate Admission System
- PeopleSoft Finance & HR, University Mission & Ministry
- Financial Aid, Student Accounts, Survey's, IPEDS
- 24 Subject Areas comprising of 300 Dimension Tables, 28 Fact Tables, couple of Bridge Tables
- Multiple star's joined with Conformed Dimensions
- Reporting Tool - Cognos, Information Server is our Data Integration and Governance, Data Quality Tool

Boston College Enterprise Data Warehouse

Missing Library Data:

- Alma: bibs, items, holdings, courses, reading lists, po lines, invoices, funds, vendors, licenses, electronic collections & portfolios, users, loans, requests
- Interlibrary Loan - lending, borrowing
- SUSHI / COUNTER Usage
- Ezproxy
- Gate Counts

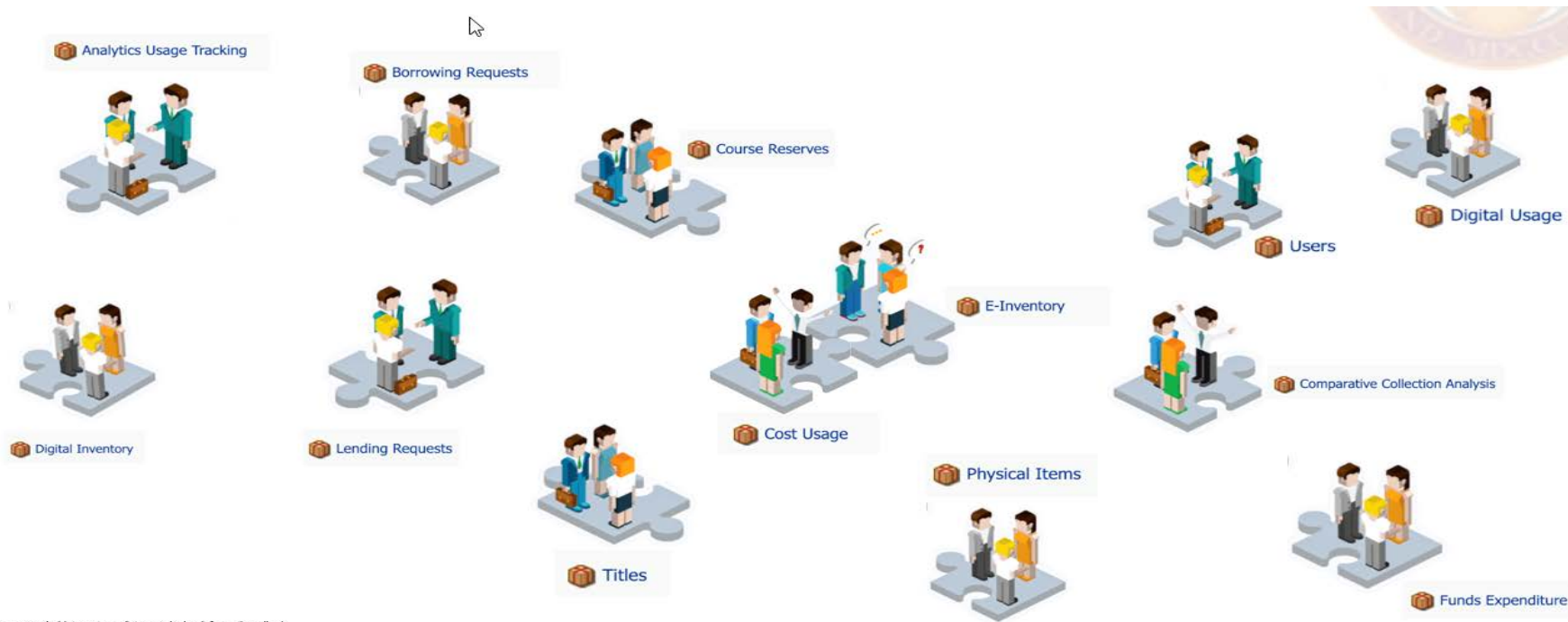
Alma Data added to EDW:



Information we can't get from Alma Analytics alone

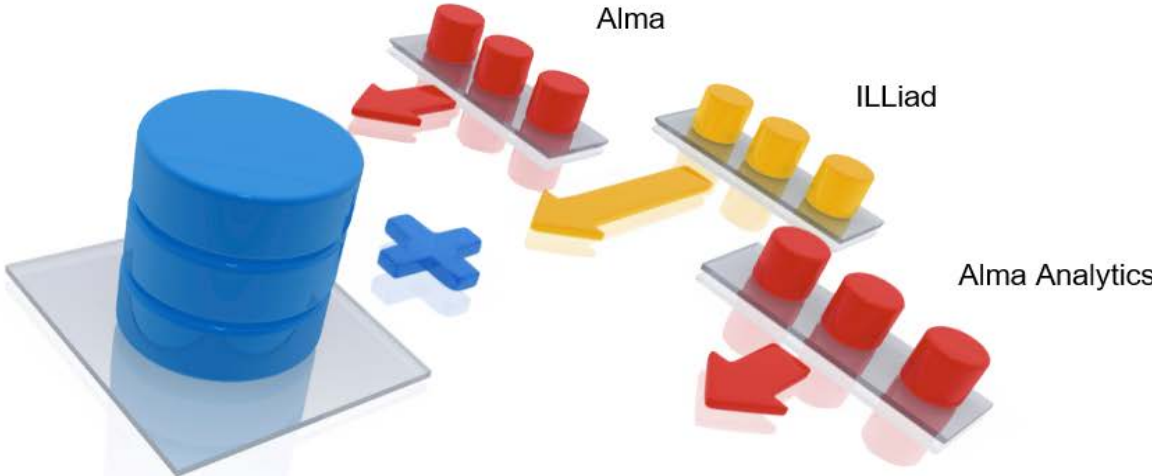
- Whether e-resource is allowed to be cancelled combined with e-resources ILL lending/borrowing
- Is faculty's use of Libraries and engagement correlated with their scholarly output?
- Is library usage an indicator of Student Retention?
- Who is using E-Resources, both on and off campus?
- What % of students / faculty / staff are using EZ Proxy?
- ARL Reporting - need a centralized place where staff can input their monthly or annual numbers for ARL reporting

Alma Analytics Subject Areas - only select data available between subject areas



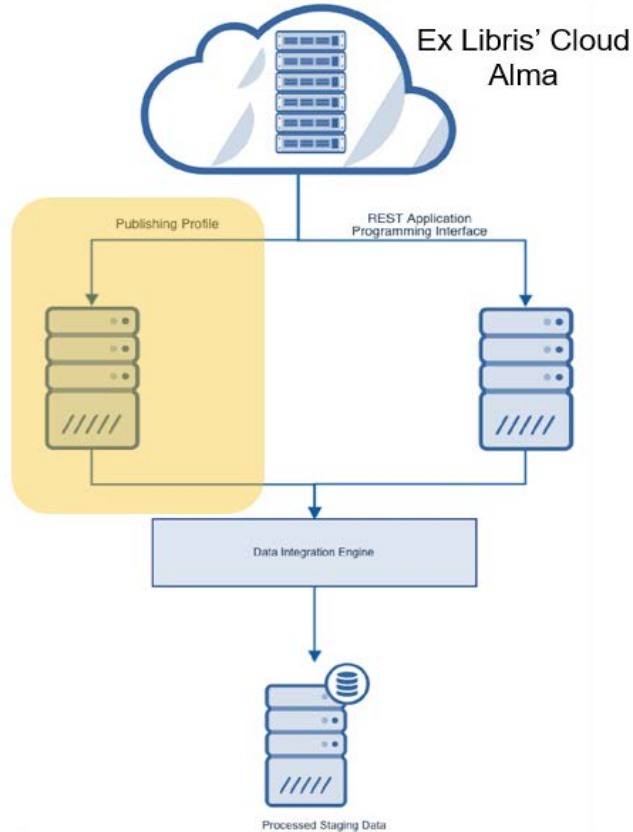


Enterprise Data Warehouse Solution



Co-Exist with other subject areas
(Academic & Administrative)

Data pulled from Alma via Publishing Profiles



Alma Publishing Profiles

- Full Bibliographic Records enriched with electronic portfolio information

Full Publishing: 3.6 Million Records, 3 Days runtime

Daily Incremental publishing: ~3000 new, updated, deleted records - 30 Mins runtime

- Full Physical Item Records

Full Publishing: 2.8 million Records, 2 hours runtime

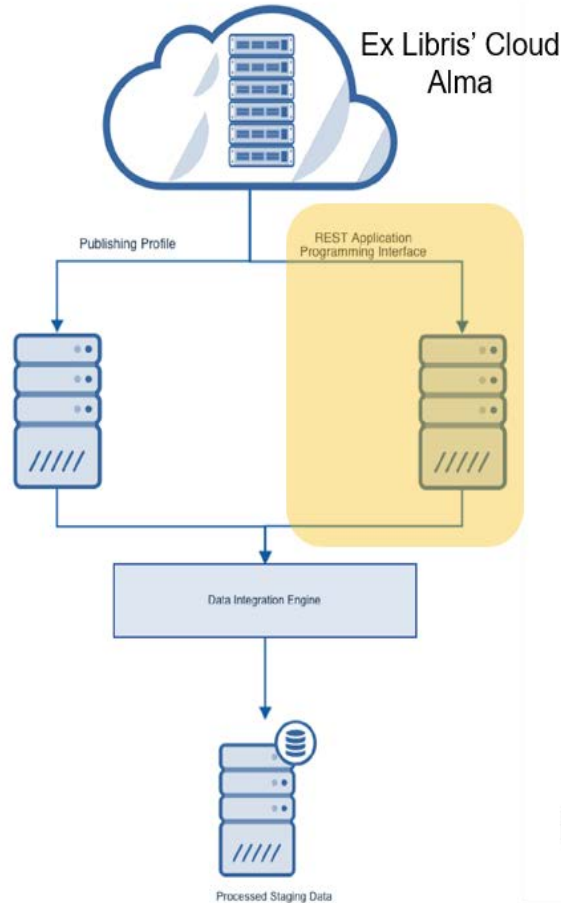
Daily Incremental publishing: ~5300 new, updated and deleted records - 13 Mins runtime

- Full Physical Holding Records

Full Publishing: 2.1 million records, 1 hour runtime

Daily Incremental publishing: ~2600 new, updated and deleted records - 8 Mins runtime

Data pulled from Alma via REST APIs



Alma REST API's

- Active Users - Includes active loans, hold requests, fines / fees - 30,000 records in 300 API calls per day (100 records returned per call)
- Active Courses - 4500 records in 45 API calls per day
- Active Purchase Orders - 980,000 records 9800 API calls per day
- Active Invoices - 128,000 records in 1280 API calls per day
- Active Vendors - 6,600 records 67 API calls per day
- Active Licenses - 1000 records 10 API calls per day
- Active Electronic Collections - 2200 records 22 API calls per day
- Electronic Usage Research Statistics
BR1, BR2, BR5, JR1, JR5, DB1, PB1 - 7 Alma Analytics API calls per month

Alma Webhooks

- Alma Webhook defined in Integration Profiles
- Webhook looks for publishing profiles to finish processing
- When each publishing profile finishes a file is written to one of BC's servers
- IT jobs waiting for the publishing profiles to complete keep checking BC server for the files to appear, one each for bibs, items, holdings
- When files are found IT jobs start running to load published files into staging tables
- More about webhooks:

<https://developers.exlibrisgroup.com/alma/integrations/webhooks/>

IT DWH Architecture

If Publishing Profiles:

- XML files published from Alma are exported to a BC server
- Webhook is written to another BC server to indicate Alma publishing is complete
- IWS Scheduling System monitors server watching for webhook
- Once webhook is found IWS Scheduling System launches Unix script to retrieve files and move them to IT Server
- IWS Scheduling System then launches IBM Datastage job

If API calls:

- Datastage jobs calls Alma API
- API returns data in XML

Both jobs:

- Datastage job parses XML files and loads data into Oracle staging tables
- Staging tables are source for Library ETL into EDW (Oracle Database)

Data Pulled from Alma Using Alma Analytics

- Loan History
- Patron Physical Item Request History
- Electronic Inventory Usage

Analytics Reports are run as scheduled jobs and CSV results are delivered via FTP to a BC server

IT jobs pick-up report results from BC server and load the data into staging tables for loading into EDW

Development Challenges

- Combining 2 different XSD Files for MARC & Alma specific XSD into one
- The XML generated by Alma Analytics is OBIEE specific. Schema information embedded into the XML File
- Generating the Publishing profile for Bib's, Items and Holdings into MARC 21 format
- Adding e-inventory enrichment to BIB's publishing profile created duplicate records for BIB's which generated thousands of additional records. Held up our publishing jobs for 3 months until Ex Libris found cause and resolved
- Electronic portfolio API takes too long to run. Workaround by adding e-portfolio enrichment to bib publishing profile

Identified Shortcomings

- No SQL Access to Data
- Create and modification dates are at the lowest level of the API calls, need to make several API calls to get this information
- Alma APIs return active records only. Deleted records cannot be determined
- Funds API does not return Parent Fund Code

Data Warehouse Toolset

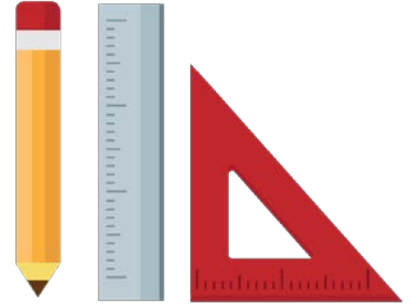


The Data Warehouse Toolset

BC used a number of Alma tools to export data into its warehouse. The tools include:

- Publishing
- Analytics
- APIs
- Webhooks

The following slides will discuss the considerations for when to use each tool in the context of a data warehouse.



Publishing

Coverage

- Repository of bibliographic data
- Enrichment with inventory information (print, electronic, digital)

Benefits

- Available in files or via HTTP (OAI)
- Incremental - change/delete detection

When to Use

- Preferred for all supported entities due to ability to publish only changed records

More information:

- [Online help](#)
- [Developer Network](#)



Analytics

Coverage

- Entire scope of Alma Analytics

Benefits

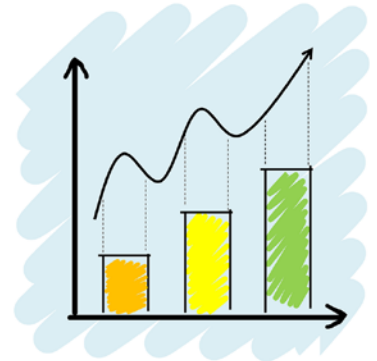
- Large coverage of data
- Ability to manipulate data and criteria
- API available to automate retrieval of data

When to Use

- When the desired data is available in analytics but not in publishing

More information:

- [Online help](#)
- [Developer Network](#)



APIs

Coverage

- 200+ REST APIs
- Wide coverage of entities in BIBs, Users, Acquisitions, Electronic

Benefits

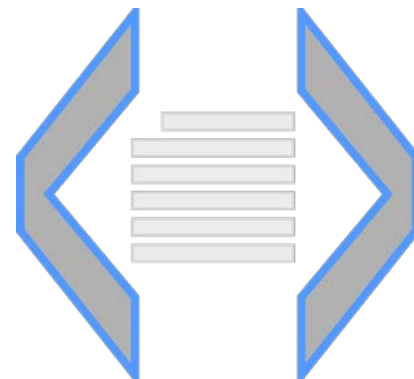
- Large coverage of data
- Ability to automate retrieval
- Retrieve paginated lists of objects

When to Use

- When access to the raw data entities is required

More information:

- [Developer Network](#)



Webhooks

Coverage

- Users
- Bibliographic records and inventory
- Loans
- Requests

Benefits

- Receive only updated entities

When to Use

- Alternative to API to retrieve only records which have been updated

More information:

- [Developer Network](#)



Other Options

There are other options as well, including:

- [Metadata Export](#)
- [Alma Local Backup service](#)
- And more

Toolset Summary

There are many options for extracting data from Alma. Boston College has successfully met their goals with the tools mentioned here. Other solutions are available depending on the specific requirements of the institution and the warehouse.

Summary and Next Steps

Next Steps

- Historical Staging for all the objects completed
- Daily incremental loads for BIB's, Holding, Items and other related objects
- Data Quality checks are in progress
- Dimensional Modeling in Progress
- Better understanding of required analytics to build functional star schema
- Analyze reporting tools : Cognos, Tableau

Questions?



Contact Information

Margaret Briand Wolfe - Boston College -

briandwo@bc.edu

Josh Weisman - Ex Libris -

josh.weisman@exlibrisgroup.com