# Introducing BIBCAT - a BIBFRAME Catalog

Jeremy Nelson

The Library of Congress efforts to replace MARC21 with a new linked-data vocabulary called BIBFRAME is generating much discussion in the library community. A new open-source library catalog and digital repository called BIBCAT - a project funded by the Library of Congress - demonstrates what a new linked-data catalog could look like for libraries and with a prototype available at http://bibcat.org/.

## Introduction

The library's catalog, usually the core of what is called the library's integrated library systems and now being marketed as "library services platform" (Breeding 2012), touches most areas of the library's operations, collections, and public services. The functionality needed to support acquisition and circulation work-flows - not mention the increasingly complex electronic resource management for journals and e-books - traditionally requires these large enterprise software systems to include all this functionality into large monolithic "black-boxes".

In the Spring 2014 issue of Reference and User Services Quarterly, Diane Cmor and Rory Litwin argue if libraries should keep the traditional library OPAC in favor of discovery systems or what Athena Hoeppner describes as "web-scale discovery services" (Hoeppner 2012). Cmor's position, that because library patrons prefer to using discovery systems to the traditional OPAC for finding the library's materials, the added costs of maintaining two separate systems argue in favor of dropping or eliminating the traditional library ILS/OPAC or catalog. Litwin's counter-argument for retaining the traditional catalog revolve around the catalog's utility for advanced searching by scholars and experts is better for their specific information needs than the current crop of discovery systems that focus on the general undergraduate at a academic library.(Cmor and Litwin 2014)

Although that might not have been the intentions of either Cmor or Litwin in their debate on retiring the library catalog, there is another alternative to what either of them suggest for library catalog. Instead of purchasing a commercial discovery system or deploying one of the open-source discovery layer alternatives, libraries should consider if their rate-of-return on their technology investment for resource discovery would be better spent through online adverting of their collections on the major commercial general search like Google or Microsoft Bing. This radical approach has already been investigated by libraries as Utrecht University Library, where their student patrons start their search 83% of the time from a search engine with under 1% using a online databases and the library website. (Kortekaas 2012)

With such library luminaries as Roy Tennent calling for the demotion of the library catalog by encouraging libraries to "Take that anachronistic library catalog and turn it back into what it really only ever was - an inventory control system." (Tennent 2014) or as Diane Cmor says "Obviously, we still need back-end catalogs (or the equivalent) to feed our holdings into our discovery systems, but the user interface is no longer necessary." (Cmor and Litwin 2014) Lacking from these analysis is that poor user interface for backend enterprise systems, even stripped down to bare functionality as envisioned by Tennent and Cmor, have real and significant costs that are borne by library staff and administration when trying to accomplish their workflow in the library.

## BIBFRAME - Library of Congress's MARC21 Replacement

BIBFRAME, short for Bibliographic Framework Initiative, refers to both the vocabulary as well as the communities and organizations that understand the importance of bibliographic data to libraries and other cultural heritage institutions. BIBFRAME as described from the Library of Congress website, is

Initiated by the Library of Congress, BIBFRAME provides a foundation   
for the future of bibliographic description, both on the web, and in   
the broader networked world…In addition to being a replacement for MARC,   
BIBFRAME serves as a general model for expressing and connecting   
bibliographic data. A major focus of the initiative will be to determine   
a transition path for the MARC 21 formats while preserving a robust   
data exchange that has supported resource sharing and cataloging cost   
savings in recent decades. (Library of Congress BIBFRAME website, 2015)

As one of the first adopters of BIBFRAME, Jeremy Nelson's early experiments in creating bibliographic discovery layers and catalogs started by creating a BIBFRAME catalog using Django, an open-source web front-end and the NoSQL Redis datastore to store information on BIBFRAME entities. Following the Modeling BIBFRAME entities in Redis that gained traction with the release of the formal BIBFRAME vocabulary at the first iteration of the http://bibframe.org website. In early 2013, Nelson published another Code4Lib article(Nelson, 2013), further describes a Redis-based bibliographic web applications with designs for Linked Data peer-to-peer and consortia datastores. At the 2013 Code4Lib conference in Chicago, Nelson co-presented on the current status of his collaborative work with the University of Denver in a presentation titled, Evolving Towards a Consortium BIBFRAME Redis Datastore, that further detailed a multi-institutional BIBFRAME application design. Invited to present at the BIBFRAME forum at annual ALA conference in the summer of 2012, Nelson demonstrated a consortium BIBFRAME discovery layer and catalog that included MARC records from the member institutions of the Colorado Alliance of Research Libraries, all 40,000+ RDF/XML records from Project Gutenberg, and over 5,000 MODS records harvested from Colorado College's Fedora 2.7 digital repository. An example of this minimum viable product of catalog is still accessible at http://tuttdemo.coloradocollege.edu/.

## Catalog Pull Platform Simplifying and De-coupling Library Technology

With ideas originating in the Lean Startup movement for technology entrepreneurs and that builds upon the success Toyota and other manufacturers have had in implementing Lean Manufacturing ideas; the Catalog Pull Platform is a fundamental shift in thinking about library systems. Moving away from trying to anticipate the needs and then "pushing" services and technology to patrons and staff, the Catalog Pull Platform instead identifies and responds to needs and demands for library technology by "pulling" directly from various constituencies served by libraries and cultural heritage organizations.

Developing the Catalog Pull Platform at Colorado College centers around creating simple, standalone web applications with minimal external dependencies. Some of these applications replaces lengthly manual MARC-based workflows for normalizing MARC records for Tutt Library's legacy ILS along with the indexing into the Aristotle Discovery Layer. Other Fedora-based institutional repository utilities including staff productivity tools that allowed repository managers to easy move collections around in Fedora as well as batch templates that added one to any number of stub Fedora objects that shared similar metadata and were all co-located in the same collection. These linked-data vocabularies were also examined to see they could be used for operational improvement of library services and expansion of access to the library's patrons. This philosophy offers loosely coupled components for rapid, iterative, Lean Startup inspired (Ries 2012) Build-Measure-Learn software development cycles for Colorado College's linked data research, experiments, and digital services. The highest profile project is a new catalog with a Fedora 4 and Flask-based semantic server made up of Colorado College's MARC records that have been converted to BIBFRAME and Schema.org linked data.

Inspired by John Hegel III's and John Seely Brown's early conception a pull platforms, the Tutt Library has embarked on approach for library catalog software development that emphasis emergent design, loosely coupled, people focused with incremental cycles that provide immediate feedback from the end users of these services (Hegal II and Brown 2008). The current technical infrastructure includes open-source projects like Fedora Commons and Elastic search for RDF and datastream management and search functionality while supporting existing MARC21 works and newer command-line and web-based tools for manipulating RDF graphs of BIBFRAME and schema.org vocabularies.

## BIBCAT

In the fall of 2014, the Library of Congress solicited bids for a new BIBFRAME Search and Display system. Aaron Schmidt of Influx Library User Interface Design and Jeremy Nelson successful proposal is BIBCAT - BIBFRAME Catalog the basis for open-source the BIBFRAME Catalog (shortened to BIBCAT) and a backend BIBFRAME Datastore that extends the Semantic Server REST API.

The Semantic Server is an open-source REST API wrapper that for managing RDF entities stored as subject graphs. The current iteration uses Fedora 4 as a subject linked-data store and binary preservation store while providing expanded and enriched search of these RDF entities through Elastic Search supported by an HTTP SPARQL endpoint using Blazegraph.

BIBCAT's current demonstration release is live at http://bibcat.org/ and currently has loaded sample MARC21 records from the Library of Congress with additional contribution of MARC records from the Colorado Alliance of Research Libraries and Colorado College. The second development iteration of BIBCAT continued through 2015 with an additional contract from the Library of Congress to add a reporting and analytics module to BIBCAT.

BIBCAT's current user interface is deliberately stripped down and simplfied to a single search input box that dynamically provides typeahead search completion as a patron enters a search query. Search results from either the typeahead search or running a search query from the search box generates a listing of results that as the patron scrolls down the page are dynamically loaded in a technique similar to how Netflix and Facebook implement their search results and social streaming. Patrons can also resort their search results as well as narrow their search focus through dynamic cluster of their results by BIBFRAME Resource categories like Work, Instance, Topic, and Agent.

## Reimagining the Colorado College Library's TIGER Catalog and Website

An ongoing experiment for replacing Colorado College's legacy ILS, the first iteration of the TIGER Catalog Minimum Viable Product was released in 2014 using Aaron Schmidt's design he outlined in a blog posting (Schmidt, 2014). The first release of this TIGER catalog illustrated the front-end viability of using Flask and is currently available at http://catalog.coloradocollege.edu/.

Requirements gathering and project planning is continuing for the next release of the TIGER catalog that will extend the user interface being developed for BIBCAT as a base platform for creating and eventually replacing current library operational workflows that focus on the library's legacy ILS.  
With the current topology of the Catalog Pull Platform that separates the library's linked data from any specific front-end application, Tutt Library is just beginning to explore the research and operational possibilities of providing comprehensive library linked-data services to the library and to the larger Colorado College campus. Part of these linked data services is moving knowledge, in the form of RDF statements of fact about the library's physical and digital resources coupled with organizational structures and history feed into an emerging common institutional knowledge graph for Colorado College. These services are also being integrated into Tutt Library's website as an improvement to the current method of information architecture and management of these statements of fact that are at the core of a RDF triples made up subjects, predicates, and objects.

By becoming institutional authority and linked data publisher of record for the college, the library's continues to evolve its role to maintain it's relevancy on the modern liberal art college campus.

## References

Breeding, M. 2012. Automation Marketplace 2012: Agents of Change. Accessed at <http://www.thedigitalshift.com/2012/03/ils/automation-marketplace-2012-agents-of-change/>

Cmor, D. Litwin, R. 2014. Should we retire the catalog? Reference & User Services Quarterly; Spring2014, 53(3):213-216.

Hagel, J. Brown, J. Davison, L. 2010. The power of pull: How small moves, smartly made, can set big things in motion. New York (NY): Basic Books.

Hoeppner, A. 2012. The Ins and Outs of Evaluating Web-Scale Discovery Services. Computers in Libraries; April 2014, 32(3):6-10.

Kortekaas, S. 2012. Thinking the unthinkable: a library without a catalogue - Reconsidering the future of discovery tools for Utrecht University library. LIBER General Annual Conference. Accessed at <http://libereurope.eu/news/thinking-the-unthinkable-a-library-without-a-catalogue-reconsidering-the-future-of-discovery-tools-for-utrecht-university-library/>

Library of Congress BIBFRAME website. Accessed at <http://www.loc.gov/bibframe/>

Nelson, J. 2013. Building a Library App Portfolio with Redis and Django. Code4Lib Journal; 19 Accessed at <http://journal.code4lib.org/articles/7349>

Schmidt, A. 2013. Catalog by Design. Accessed at <http://www.walkingpaper.org/5979>

Tennant, R. 2014. The OPAC is Dead. The Digital Shift. Accessed at <http://www.thedigitalshift.com/2014/02/roy-tennant-digital-libraries/the-opac-is-dead/>