

To: Rebecca Cremona
Re: Library Analytics Toolkit Final Report (# 514/703)
Date: November 15, 2013

Overview

We have received three generous grants from the Harvard Library Lab to build the Library Analytics Toolkit. The Toolkit serves as a dashboard that pulls library data together in a way that allows both librarians and library users to identify and respond to trends and changes in the Harvard Library's collections, usage and other data.

The first Library Lab grant funded research to determine the types of analytic information that Harvard librarians would like to see in the tool. Analytics projects at other schools were evaluated as a way of gauging what other libraries had found to be useful. This stage of the project culminated in [a report](#), which made recommendations about the types of data that should be included in the initial version of the Toolkit and ways that it could be expanded in the future.

Library Lab funding for the second stage of the project allowed us to begin building the tool. The focus was on visualizing collection and circulation data in a way that would be helpful for collection development professionals and interesting for library users. This stage culminated in the completion of a working data visualization tool that allowed users to explore data about Harvard Library's collection. That version of the Toolkit is available online at <http://librarylab.law.harvard.edu/toolkit/>.

In October 2012, we received a grant for the third and final stage of the project. This funding allowed us to add new functionality and content to the Toolkit and to revamp its user interface. New features included additional data points, such as electronic resources usage and library location; a more granular subject breakdown to provide more focused search opportunities; and a more user friendly interface. We also modified the tool to feature a more interesting data visualization design that pulled together the different functions of the existing Toolkit.

Accomplishments

Third stage work began in March 2013. We purchased a dataset from the Library of Congress to allow for greater precision in subject searching so that the users of the Toolkit could get a more granular look at how the Harvard Library collections are being used. We mapped LC call numbers to items in the current collections that were not LC-classed, including to COUNTER statistics on e-resource usage. The historical range of the application was expanded in areas for which we had corresponding metrics. We modified the overall architecture of the application to allow for these expanded metrics and to streamline search and the overall user experience. The backend data wrangling work was done in house at the Harvard Library Innovation Lab. The data visualization and front end architecture was handled by the same outside contractors we used for the development of the original Toolkit application.

Use cases were developed and handed off to the data visualization developers for wireframe production. A wholesale redesign of the previous application was suggested, whereby items in the Harvard Library collection were represented by bubbles aligned on an x/y graph that expanded, contracted, and otherwise moved as search terms were modified. Additional functionality included:

- Adding dynamic links from the individual visualizations to item level information produced from HOLLIS and StackLife data;
- Improving ability to break down search functions by user type, publication date range, owning library and classification schema;
- Refining the bubble like visualizations to provide for easier retrieval;
- Adding an export function for results;
- Adding a scroll bar for better search precision;
- Improving visuals for Library of Congress Classification schema.

Once the new features were implemented, testing began and work on the final user interface was completed. Further work on refinement of search retrieval and the breakdown of classification visuals took up the remainder of the developers' time.

Code

All Library Analytics code is freely and publicly available at Github: <http://github.com/harvard-lil/analytics-dash>

Budget Spent:*

\$3,975.00 – For the purchase of MARC Classification records for all Library of Congress Classification printed schedules from A-Z.

(<http://www.loc.gov/cds/products/product.php?productID=14>)

\$13,715.00 – Payments made to developer, Rosten Woo, for data visualization development work.

\$18,000.00 – Payments made to developer, Rachel Witte, for data visualization development work.

Total: \$35,690.00

* Attached CREW report does not reflect the latest invoicing. We will submit a supplemental CREW report with updated information next week.