
Evidence-Based Decision Making Using New Library Data

C. Heather Scalf
University of Texas Arlington, USA

Abstract

Evidence-based decision making is becoming more and more important across the academy and in libraries. University of Texas Arlington (UTA) Libraries implemented card swipe access in the Central Library as an opportunity to combine new data with other university data in order to make data-driven decisions about services and partnerships. The project has gathered data from library access points and university sources and combined them into a single secure database, linking student information such as major and classification and de-identifying the student ID number; this was then shared for analysis with the research team as a structured dataset. This analysis provided data pivotal to discussions with campus partners about new or continuing services.

Background

In the summer of 2012, UTA Libraries was joined by a new dean of libraries. Shortly thereafter, the libraries initiated wide-scale assessment projects to benchmark with peers across the country in several key areas and to learn more about users' research habits and the use of the spaces in various facilities across campus, including the existing five library locations. Findings from the space evaluation activities (affectionately called "Where's Waldo?") revealed that, while the traffic pattern was consistent with most academic libraries in that it resembled a bell curve throughout the day, there was a noticeable contingent who came into the Central Library late in the evening and stayed into the wee hours of the morning. Observations conducted during this ethnographic work showed that there were more than 300 people in the building at 2:00 a.m. over the three-week observation period. After sharing the results, university administration became very interested in partnering with the libraries to learn more about those using the facility as well as what activities they were engaging in while they were there. With the arrival of a new university president in 2013, funds were made available to add entry and exit gates using university identification card swipe

access when the first floor to the Central Library was renovated.

The Project

Discussions about creating a database to gather and evaluate usage data from the access points began in earnest when planning for the renovation began. The libraries began to identify what would be needed to accomplish a secure merge of entry and exit data and demographic data about its students. The libraries are using the same entry and exit gate technology that is also used across campus for access to various labs and classrooms. There were a number of issues to address in the initial design phase of the project. First, most online resource datasets do not contain much if any demographic data about users, which is a positive thing when considering the need for patron privacy, but can be problematic when attempting to design a project that will incorporate data sources with different attributes. Library use data is stored in a number of disparate systems, many of which are vendor hosted or supported, and this allows for a variety of user identities, which complicates the connection and merging of data. The libraries have existing access to the university's Central Enterprise Directory and Authentication Realm (CEDAR), which populates the patron records in the Integrated Library System (ILS), Voyager. This access made it possible to leverage the data in order to create the project and build the database. The demographic information contained in CEDAR provides what is needed to begin to tell a more comprehensive story about library use.

Security of the data is a key consideration for the project, and no data could be gathered or merged until a secure server was procured and provisioned by our Office of Information Technology (OIT). Once this was made available, the systems librarian for programming and analysis began to create the database structures and scripts necessary to import and manipulate the data into LIBLAND (Library Learning Analytics Database). Greater detail about the technical aspects of the development of the

database can be found in Michael Doran's recent Library and Information Technology Association (LITA) Forum presentation, and is available at <http://rocky.uta.edu/presentations/>.

Data and Security

Extensive measures have been taken to not only secure the database, but also provide cryptographic security of the unique 10-digit student identifiers that are used to link usage data from different data sources. As frequently happens when providing online library resources, there are several potential identifiers that are used by patrons to authenticate access. Each of these must be ascertained, linked to the primary 10-digit identifier in order to allow for analysis, and then removed from the exported Access database used for analysis. Before export, the unique 10-digit identifier is subjected to two separate and non-reversible cryptographic processes to create a 256 character string as the new unique identifier. The unique identifier is first put through a cryptographic random "salt" process, blending numbers and letters with the original string. This new character set is then subjected to a one way hash process to further de-identify the resulting 256 character string. This process is completed once for each semester's data, just after the first week of the following semester. It is then exported from the database and shared with the director of assessment for analysis in the form of an Access database.

As the project was beginning, the two principal investigators completed a protocol for submission to the university's Institutional Review Board governing human subject research. After a review of the protocol and many subsequent discussions about the encryption processes, it was determined that the LIBLAND project was exempt, based upon the category that identified it as research involving the collection or study of existing data, where the subjects cannot be identified, either directly or through unique identifiers. Throughout the project, great care was taken in the selection of the attributes that would be retrieved from the data sources. While the unique ID number can be retrieved, names of subjects cannot. Similarly, age is calculated and retrieved, but not date of birth and the zip code is made available for analysis, but not street address. These are but a few of the decisions that have been made in order to protect the individuals whose data is represented in the dataset.

Findings

Once card swipe data began to be merged in LIBLAND, the first goal was to determine not only how many entries we had to the building during a certain interval, a typical library question, but also the duration of visits to the library, as this would build upon the results of the earlier ethnographic work. While entry and exit data continued to support the bell curve distribution of visit counts, new patterns began to emerge from using the date/time element now available to us. Students who entered the library during typical business hours of 8:00 a.m. to 5:00 p.m. in April 2015 stayed an average of 80 to 100 minutes, but students who entered between 8:00 p.m. and 1:00 a.m. stayed more than 144 minutes on average. More details about this can be found on the UT Arlington Libraries Tableau Public page.

This information was of critical importance in our discussions with the food service vendor who operated the coffee and snack shop in the Central Library. Libraries' leadership was able to bring evidence to the discussions to show that students were indeed present in the library for extended hours during the overnight hours, and previous observations showed that food and/or drink were visible on the desks and tables for more than 60% of the observed students. The food service vendor then agreed to stay open and provide coffee and snacks until 1:00 a.m. five nights a week, resulting in a quadrupling of its sales.

A close corollary of this work was the first combination of demographic data with entry data to determine the percentage of visits from each school's enrollees. The result that 28% of visits to the Central Library were from the College of Engineering was particularly surprising, given that there is a library in the basement of the Science and Engineering building, and that the Engineering Research Building also includes a lot of study space. Analysis of the trends by hour of entry showed that while students enrolled in one of the other seven colleges or schools displayed the same general bell curve distribution, engineering students had a disproportionate increase in late night and very early morning entries. Subsequent observation of user behavior in the Science and Engineering Library led to our understanding that many students were migrating from there to the Central Library, as it was open all night. Based upon this, library administration agreed to extend the hours at SEL to 2:00 a.m. in order to better meet the students' need for space and

resources. Analysis of data from the current semester may reveal new or different trends in this area.

After this initial success, the project began to expand and other data sources were added to LIBLAND. Entry and exit information was joined by other usage data from sources such as Voyager, InterLibrary Loan (ILL), EZproxy authentication logs and our OpenRoom room reservation software. Each dataset was analyzed prior to the development of a load script and all of its particular attributes were screened before inclusion in the dataset. For instance, EZproxy logs include not only the student identifier but also the uniform resource locator (URL) of the resource that was accessed. Knowledge of this prompted the development of a script that truncates the URL while the file is loading so that the only remaining URL data available for analysis is the name of the source database and not the article title.

Analysis of EZproxy logs provided quantitative data to support the anecdotal evidence that our online-only students enrolled in the College of Nursing and Health Innovations (CONHI) accounted for a substantial percentage of the usage of our online resources from off campus. While the university had a global enrollment total of over 45,000, there are about 39,800 Texas-based students as of this semester, which reflects a growth of 7.3% since fall 2015. Just over 4,000 of those students actually live on campus. Meanwhile, CONHI enrollment increased by 23% between 2015 and 2016 and accounts for more than 23% of the courses offered during the fall 2016 semester. Evaluation of EZproxy logs showed that students enrolled in CONHI courses facilitated by our commercial partner Academic Partnerships (AP) use our resources at least as much as students in all other schools and colleges combined. This evidence was provided to the administration in order to develop a business case for the addition of a library fee for AP students to support the increasing costs of e-serials and other online resources.

In 2014, UTA was designated as a Hispanic-Serving Institution (HSI), with over 25% of enrolled students identifying themselves as being of Hispanic or Latino heritage. In addition to that, almost 67% of students enrolled for fall 2016 are transfer students. In 2015, the university received a \$2.62M grant that will assist in the creation of a new IDEAS Center—for Innovation, Diversity, Excellence, Access and Success—as a resource to increase graduation and retention rates among Hispanic and lower-income students. This center is housed in the Central Library, as part of our second floor academic plaza. In order to get an understanding of the traffic flow and to staff for their opening in the fall of 2016, analysis of baseline data was provided to them. In the fall of 2015, 31% of the student entries to the Central Library were transfer students, and 20% of the total entries were students who identified themselves as Hispanic or Latino.

Looking Ahead

The conclusions that have been reached at this point in the project have proven useful in making a number of business cases within the libraries and with outside partners. Future work will include an analysis in 2017 to determine if there is any correlation between library usage and academic success, specifically grade point average. New data sources are being developed to be included in the database, and the libraries are a very active partner in the development of a university-wide data warehouse supported by SAS. As the university continues to focus increasingly on evidence from its academic units, the libraries' goal is to be an active partner in demonstrating how effective decision making can be done by providing evidence using data available in a variety of sources.

—Copyright 2017 Heather Scalf

Correspondence concerning this paper should be directed to Heather Scalf, UT Arlington Libraries, Arlington, TX 76016. Contact scalf@uta.edu.