2016-10-03

Implementing Library Analytics at the University of Michigan

Varnum, Kenneth J.

https://hdl.handle.net/2027.42/133959

http://creativecommons.org/licenses/by-nc-nd/4.0/

Implementing Library Analytics at the University of Michigan

Ken Varnum
Senior Program Manager

Digital Analytics Tools for Libraries Webinar

A bit about me & my role at U-M. Discovery, Delivery, Library Analytics. Largely intertwined: helping our users find resources, access resources, and helping the library understand how and where that happens -- and how well (or poorly). Will inform multiple levels of the library; back- and front- ends of website. Services we offer (in-person or technologically mediated). And, ultimately, be an input into a much larger campus effort to understand what it is the University of Michigan does that enables success for students.

Background & Context

Campus Context

Library Context

Updating the Library's Privacy Statement

Campus learning analytics program.
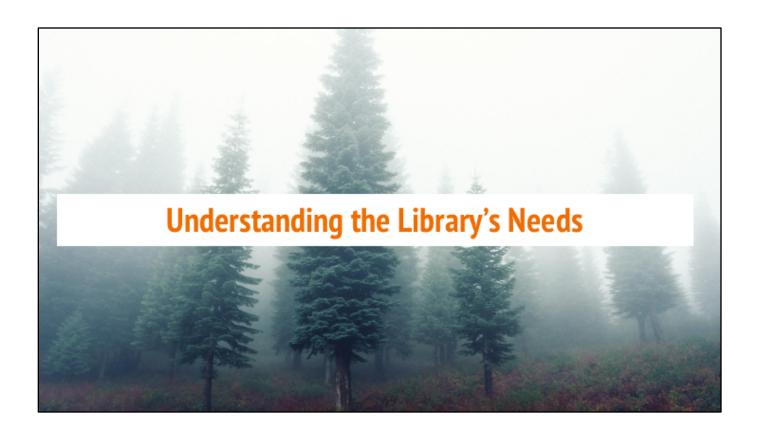
Library's Dean is also Vice Provost for Academic Initiatives → focused on studying how we teach, and the impact various methods and tools have on outcomes.

Library discussed these changes for many months, from first proposal to ultimate decision.

Privacy statement -- updated in September to come into line with campus model; allows the library to use transaction log and other data to understand how users utilize our services, and -- this is the big campus-wide goal -- how various inputs to education result in outcomes. What is that we (that is, we the university or we the library) do that has a positive causative effect on educational outcomes? We don't know. And worse, we don't as a broad institution know what the challenges, specifically, or therefore how to address them. The campus -- and the library -- hopes to learn this.

In the near term, however, we are taking advantage of this opportunity to understand how our users interact, at a large scale, with our services, and what works for them and what doesn't. Even if we can't relate our data to campus-level outputs (e.g., graduation rates or post-graduation success), we can figure out how to more broadly improve our services. This is a significant user experience improvement tool.

Once we had updated our privacy statement to enable a range of data collection, next step was to figure out how to start.

Understanding the Library's Needs

You can see the forest, or you can see the trees. We spent a lot of time working out our approach, which boiled down to identifying specific research questions to answer. We did a library-wide survey. That turned out, to nobody's surprise, to be varied in terms of services (virtual and real-world), data needed, time spans desired, etc. And when we through the data points that Design & Discovery (my unit) has for user research and web tool optimization, it became crazy.

So, we rethought our approach. We reduced our short-term scope to library staff members' questions that could be answered, theoretically, from data gathered from online interactions through the website. We merged that with services that D&D is looking at in the near term for updates, changes, or general understanding. And then we tried to come up with specific data points from specific sources that might be useful to answer those questions.

So, to answer questions about what users are accessing online journals and how they are doing so (an important question in terms of our journal finder, main site search, and acquisitions and promotion efforts), we wanted data from a range of sources:
What is searched on the library site or catalog that leads to an online journal?
Where do people enter our link resolver from?
Where do people go from the link resolver?
Who are these people, anyway (e.g., faculty, undergrads, grads; school/college affiliations; perhaps even courses enrolled in)
And someday… Perhaps what do users do with the information they access (e.g., cite it in a paper, etc.)

Next Steps (A Work in Progress)

If we tried to capture all the data to answer all the questions we had at present, we'd be faced with a massive task.

So we took a staged approach, and are now beginning to work on a pilot of data collection. There were a number of data elements from a number of sources that were common needs across a number of questions. Few are sufficient to answer any one question thoroughly, but they represent a common core of data. So we are starting the process of building this data collection process.

On the back end, we are implementing Elasticsearch to store and access selected data -- extracted either from server logs or from custom-built data pipelines -- and the Kibana plugin for visualizing the data we're storing.
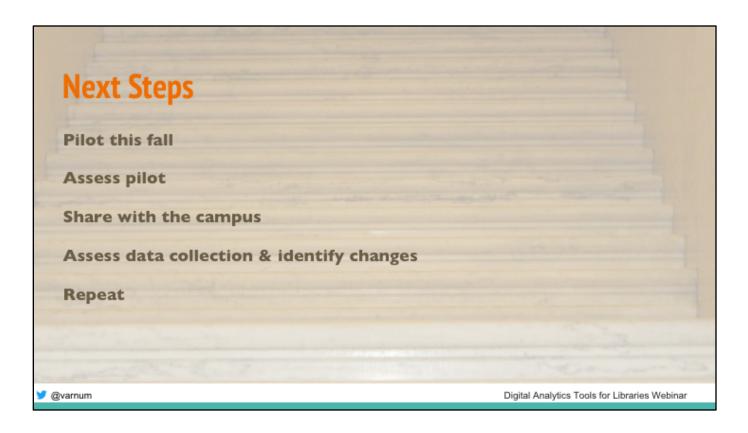
Realized the pure log analysis and Google Analytics (or other traditional analytics engines) are not likely to give us the data we need. Much of what we want to know requires closer integration with the applications themselves.

Pipelines. Sometimes, we can get good data through the web server logs. Where we can obtain useful data, we can extract useful bits from it and store it in ElasticSearch. In other cases (e.g., our new Umlaut link resolver, going into production later this month), we are building in logging of the details we want while not capturing absolutely everything that is possibly capturable.

We want to build a baseline, and not slurp in everything that anyone might do.

Proposed Initial Data Gathering

Data Sources	Web Site Server Logs	Mirlyn Server Logs	Proxy Server Logs	Aleph Circulation History and related data	Campus status & affiliation data
Data Attributes	timestamp	timestamp	timestamp	timestamp	
	uniqname	uniqname	uniqname	uniqname	uniqname
	event/action type (search, display details, external link, etc.)	event/action type (search, display details, get this, external link, etc.)	event/action type	event/action type (charge, discharge, renew). Other related events of interest (e.g., hold, recall, overdue sent, library-to-library delivery, etc.)	user category (undergrad, grad, faculty, staff, sponsored, etc.)
	Database/Package/ Resource identifier(s)	Aleph resource IDs (barcode, bibliographic record number, etc.)	Resource identifier	Aleph resource IDs (barcode, bibliographic record number, etc.)	user department
	Search string (when event is a search)	Search string (when event is a search)	Result (user passed through or dead end?)		additional user demographics
	Search results	Search results			
	Referrer string (where user came here from)	Referrer string (where user came here from)	Target		
	Target (when event is external link)	Target (when event is an external link)			

Colors represent potential linking points to connect data from different sources. Note that some connections (e.g., uniqname) are relatively strong and unambiguous while other (e.g., timestamp, referrer, and target strings) may be circumstantially related, but we rely on inference when postulating a relationship as a user session or path between multiple systems.

Pilot the core data sets described in previous slide

Check in with library researchers and work with them to access and assess the data we're collecting

Update data collection until we can answer these questions

Share our process and progress with campus

Identify new data needs; assess ongoing collection of data we've been gathering

Thank You!

Happy to take questions during the Q&A

Ken Varnum
Senior Program Manager
Design & Discovery / Library Information Technology
University of Michigan Library
varnum@umich.edu
@varnum