

# Library Analytics Investigation Team Recommendations

7 September 2016

Albert Bertram  
Noah Botimer  
Doreen Bradley [TASC Liaison]  
Marisa Conte  
Kathleen Folger  
Diana Perpich  
Emily Puckett Rodgers  
Jon Rothman  
Lori Tschirhart [Service Lead]  
Ken Varnum [Technical Lead]

The University of Michigan Library's Technology Alignment and Stewardship Committee charged the Library Analytics Investigation Team in the Winter 2016 cycle. The team met from April through September of 2016 to investigate processes to provide library-generated data to library staff for service improvement and/or research investigations.

## Recommendations

The investigation team recommends that an project team be formed to pilot data extraction, storage, and analysis processes for the data sources we outlined. This team should have the same (or functionally similar) membership as the current investigation team, although most of the active work of the team will be concentrated within LIT.

Through the pilot process, the library can gain an understanding of our capacity and constraints for gathering and providing library-generated data to meet specific researcher needs. The project team will consult with members of the TASC investigation team as needed through this process.

TASC should charge a new project team to accomplish the following:

- Define a pilot period
- Develop data collection and storage practices
- Identify tools to extract and analyze data from the data sources and elements specified in Table 1
- Coordinate with the library researchers who submitted research questions during Summer 2016 to provide available data and hone the data collection and reporting mechanisms.
- Develop procedures for library staff to request data collection and a workflow within LIT for adding new data points to the existing data collection and storage flows
  - A conversational, “reference interview” type approach between potential library data consumers and LIT may provide a superior way to learn about the library generated data needs of library researchers. A workflow that allows providers of library generated data a chance to ask questions will likely lead to delivery of more meaningful library generated data (see “Process” below).
- Maintain the Inventory of known/available and potential sources of data and data points within them
- Scan for gaps where we currently need data but can’t readily characterize the source

While it is beyond the the pilot Project Team’s scope, we recommend that LIT provide a page on the Intranet outlining policies and procedures for using library data, a succinct guide to other useful resources, and a contact method for initiating new analytics requests.

## Process

The team began by finalizing its charge and charter.

We initially planned to develop and distribute a survey (see Appendix 1), then apply pre-determined criteria (see Appendix 2) to survey results to select 3-5 appropriate research questions regarding user transactions that library staff would like to investigate. Our intent was

to pilot a process for providing requested data to library colleagues using the selected research questions. Upon our attempted analysis of survey responses, however, we realized that:

- Most respondents wanted more than one kind of data source, type, and attribute.
- Many respondents presented promising ideas about what kinds of data they needed, but did not fully articulate the data types and attributes desired.

Our team required more information from respondents to understand what kinds of data would fill their respective needs. We abandoned our pre-determined evaluation criteria in favor of a different approach. We collectively reviewed all survey responses to identify commonly-requested data sources, types, and/or attributes that would provide a baseline set of data as a foundation to support many of the questions. We selected five data sources that appear to have broad utility per representation across multiple survey submissions (see Table 1).

This baseline set of sources can be expanded in order to provide data for current and new questions that arise during and beyond the pilot period.

Our team followed up with survey respondents informing them of our pivot, detailing data sources that we selected to pilot, and provided sample data attributes for selected data sources.

## Table 1: Data Sources and Data Elements

This table provides some examples of what we might learn (data attributes) from various data sources.

| Data Sources    | Web Site Server Logs   | Mirlyn Server Logs   | Proxy Server Logs                         | Aleph Circulation History and related data   | Campus status & affiliation data (MCommunity and/or MPathways)   |
|-----------------|--|--|---|--|--|
| 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.) may come from elsewhere in Aleph. | user category (undergrad, grad, faculty, staff, sponsored, etc.) |
|                 | Database/Package/Resource identifier(s)                          | Aleph resource IDs (barcode, biblio. record number, etc.)                  | Resource identifier                       | Aleph resource IDs (barcode, biblio. 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.*

# Recommendations Beyond the Scope of the Investigation and Project Teams

Fall 2017

- Defining an assessment mechanism for the pilot and conducting an assessment exercise.
  - Part of this mechanism should involve follow up with library staff who have accessed library generated data through the pilot project.
- Defining an assessment mechanism for when data collection for a data point is no longer needed.
- Defining a procedure for determining when previously collected data are no longer needed and may be destroyed.
- Periodically scan for gaps where we need data but can't readily characterize the source - this work must be iterative.

## Potential Resources for Staff Education

The investigation team identified a number of resources that other library staff might find helpful in understanding the issues and challenges in implementing library analytics:

- [U-M Library Privacy Statement](#) (updated March 2016)
- SPG Institutional Data Resource Management Policy (601.12)  
<http://spg.umich.edu/policy/601.12>
- Abelardo Pardo and George Siemens. [Ethical and privacy principles for learning analytics](#). <http://spg.umich.edu/policy/601.11> Article first published online: 1 APR 2014  
DOI: 10.1111/bjet.12152
- Institutional Review Board:  
<http://research-compliance.umich.edu/irb-health-sciences-and-behavioral-sciences-hsbs>
- Loyd Gitari Mbabu, Albert Bertram, Ken Varnum, [Patterns of Undergraduates' Use of Scholarly Databases in a Large Research University](#), The Journal of Academic Librarianship, Volume 39, Issue 2, March 2013, Pages 189-193, ISSN 0099-1333, <http://dx.doi.org/10.1016/j.acalib.2012.10.004>.

## Appendix 1: Survey Instrument

# Tell us about Your Library Data Needs

Recent changes in the Library Privacy Policy provide us with new opportunities to understand users, how they interact with our services and our resources, and how the library contributes to student and faculty success.

The Library Analytics Investigation Team is soliciting ideas for questions that could be answered with library-generated data. Data might include items such as web site usage, circulation information from the catalog, library instruction sessions, access of online resources, or other items.

From this survey, the team will select 3-5 questions to pilot a process for requesting, reviewing, and implementing data collection to support your work. All of the questions in aggregate will help the team understand the range of depth of library-generated data needs of our staff. Please submit your ideas even if you think it would be an unlikely candidate for our pilot.

Your email address ([varnum@umich.edu](mailto:varnum@umich.edu)) will be recorded when you submit this form. Not **varnum**? [Sign out](#)

\* Required

1. **Tell us your name \***

.....

2. **Your primary reporting unit \***

*Mark only one oval.*

- Budgeting and Planning
- Collections
- HathiTrust
- Health Sciences
- Learning and Teaching
- Library Information Technology
- Operations
- Publishing
- Research

3. **What is your question? (Please be as detailed as possible). \***

View sample questions that could be informed by library-generated data here:

<http://umlib.us/libgendata>

.....  
.....  
.....  
.....  
.....

**4. What kind of library-generated data would help you answer your question?**

(examples: transaction time stamps, search log queries, patron head counts, qualitative data derived from library service desk staffers, use statistics)

.....  
.....  
.....  
.....  
.....

**5. What is the benefit of investigating this question? \***

Please tell us about the potential value to library users, library staff, the university, and/or the profession.

.....  
.....  
.....  
.....  
.....

**6. Where might the library-generated data that you seek be found?**

Categories and corresponding examples provided below are not exhaustive and not indicative of data needs that will be favored for selection. Not all data elements listed below are necessarily available to researchers. We would rather collect requests for data that we don't or can't offer than constrain your imagination.

*Check all that apply.*

- Budget and Planning (budget info/hiring statistics/gifts, etc).
- Collections (examples: funds spent/items received/items weeded, etc).
- Circulation (loans/renewals/ILL transactions/Special Collections usage, etc).
- Instruction (courses taught/session length/session dates/open workshop attendance, etc).
- Online Resources (library proxy server logs/site visits/library website search logs, etc).
- Service Points (Ask a Librarian/reference/information desk/research consults/technology consults, etc).
- Spaces (gate counts/head counts/room reservations/event bookings/library workstation logins, etc).
- Other: .....

**7. Please tell us anything else you'd like us to consider regarding your data needs.**

.....  
.....  
.....  
.....  
.....



Send me a copy of my responses.

---

Powered by  
 Google Forms

## Appendix 2: Criteria for Evaluating the Survey

## Proposed Evaluation Rubric (highly scored proposals favored)

| <b>Criterion</b>   | <b>1 point</b><br>Unclear/<br>no | <b>2 points</b><br>Somewhat | <b>3 points</b><br>Mostly | <b>4 points</b><br>Yes |
|--|----------------------------------|-----------------------------|---------------------------|------------------------|
| Is request well-defined?   |                                  |                             |                           |                        |
| Does question tie in with institutional goals/mission/priorities?  |                                  |                             |                           |                        |
| Do workflows exist (or are future workflows identified) for accessing data (whether or not is has been collected yet)?             |                                  |                             |                           |                        |
| Is requested data readily available? (Is enough data available or will dataset only be useful when more data has been collected )? |                                  |                             |                           |                        |

## Notes about proposals:

Lots of proposals are about making sense of things

Many submitters want to know how their actions impact other services...

Submitters might really benefit from some dashboards that they can explore so that they can do some sense making before they ask specific questions.

Other possible criteria: (Can be used as tie-breakers)

- How much impact would knowing the answer have (low, low-middle, middle-high, high)
- How much effort to wrangle the data (low, low-middle, middle-high, high)
- How many divisions (1-8) could use the results? (one, a few, most, all)
- Will library generated data be enough to answer the question (or is it needed in conjunction with data that LIT cannot provide)?
- Is this a question that can be answered in other ways that workflows already exist (don't want to create new workflows unless workflows can be improved)
- A "wow" factor?

- Post-pilot notes: we should ask people to submit their hypotheses in addition to their question and the data that they want to use to investigate it.
- We need to design a request form that will ask people to scope their questions better so that we can know whether workflows exist...
- Future requests should include a consultation (possibly with LIT or someone who can do a type of “reference interview” to help figure out correlations, whether answering some but not all pieces of a request would be valuable, etc.