The University of Michigan

The Impact of the Academic Library on *Learning* in the University

We request $499,750 from the IMLS for a proposed project at the University of Michigan (U-M) to study the links between library interactions and learning. Our project aligns with the NLG program's focus on libraries as community anchors. Like other academic libraries, the U-M library anchors learning for multiple communities including: a) U-M's faculty, staff, and fellows; b) residents of Ann Arbor and the State of Michigan – members of the public and academic visitors are welcome to use resources in-building; c) reciprocal borrowing with Ann Arbor area academic institutions; d) peer institutions in the U.S. e.g. the Big Ten Academic Alliance; and e) professions e.g. the U-M Law Library is open to all Michigan Bar members. Focusing on community (a), we will address the following research question: *How does the academic library impact learning in the university community, specifically in the areas of course instruction, funded research, and publications?* We will consult a set of 12 diverse institutions to prepare a report that explores the study's replication across institutional settings.

We define learning as the process of acquisition of knowledge and/or skills through formal study or instruction (course instruction), or experientially through laboratory research or clinical activities (funded research). However, learning is also the outcome of this knowledge acquisition process, including the ways or means by which this knowledge is disseminated to society (academic publishing). Our study will use both **deidentified** and **identifiable** library use data starting from 2016 when the university revised its privacy policy to make possible the collection of individual identifiers. Library data sources include: website server logs, MIRLYN (U-M Library Catalog) server logs, proxy server logs, Aleph circulation history and related data, and campus status & affiliation data. These datasets are linkable through both strong and unambiguous identifiers that are unique to each individual user, as well as IP addresses and timestamps. The size of these datasets would be in the order of a few gigabytes a day assuming that most of the data are retained on the servers, plus additional database transaction logs. *Identifiable* data are those generated by authenticated users while *deidentified* data are those that are unauthenticated but originate from networks managed by U-M as opposed to the world at large. For example, of the just over 1 million "sessions" on the library catalog for calendar year 2016, about 300,000 sessions were originated from U-M networks, and roughly 100,000 sessions were authenticated (or identifiable). The identifiable data will also be merged with U-M Data Warehouse[1] (e.g. human resources, grants, course instruction, etc.) and publications datasets. Our analysis will be carried out in three major steps as follows.
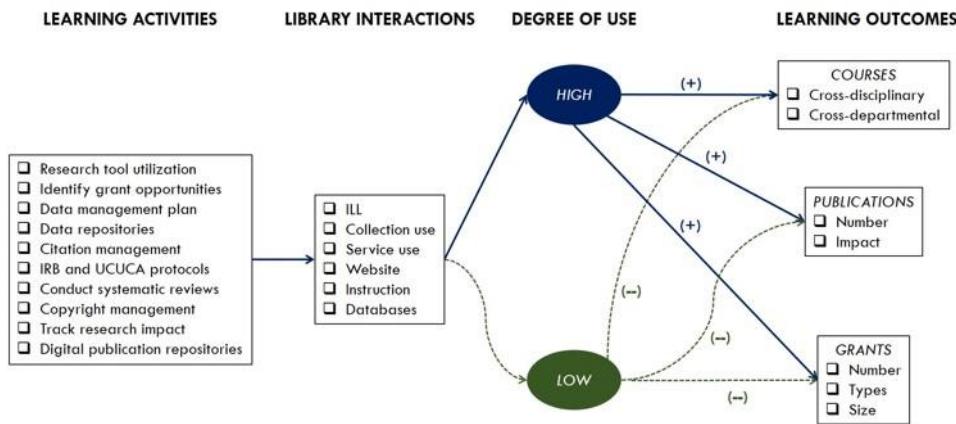

Figure 1. Impact of the academic library on learning

First, using deidentified data, we will apply clustering algorithms to identify typologies of library users on the basis of library interactions (Figure 1) e.g. user type A (high degree of use), user type B (low degree of use), etc. Second, we will replicate the clustering analysis using the identifiable data and use clustering findings from the deidentifiable analysis for robustness checks. Third, clusters from identifiable data will be used for statistical modeling of the associations between learning outcomes and library user types with appropriate controls e.g. individual-level (such as demographics), organizational-level (such as departmental affiliation), and spatial (such as office distance to library which matters for physical resources). Our proposed study is a holistic, longitudinal, and multi-method (e.g., regression modeling[2]) empirical analysis informed by patterns of library resource use (Figure 1) and models of information behavior.[3]

---

[1] "Data Sets in the U-M Data Warehouse." (http://www.mais.umich.edu/reporting/datasets.html)

[2] "An Introduction to Regression Analysis." (http://www.law.uchicago.edu/files/files/20.Sykes_.Regression.pdf)

[3] "Information Behavior." (https://pages.gseis.ucla.edu/faculty/bates/articles/information-behavior.html)

The multidisciplinary (e.g. library science, information science, network science, spatial science, etc.) project team is led by Dr. **Felichism Kabo**, Institute for Social Research (ISR), and the Michigan Institute for Clinical & Health Research (MICHR) network science evaluation faculty. Co-investigators are: Dr. **Stephanie Teasley**, Research Professor, School of Information; **Laurie Alexander**, Associate University Librarian for Learning and Teaching; **Doreen Bradley**, Director of Learning Programs and Initiatives, U-M Library; and **Bryan Skib**, Associate University Librarian for Collections. The project team also includes expert consultants and staff from various U-M units, including the U-M Consulting for Statistics, Computing & Analytics Research (CSCAR).

The proposed project will join a growing learning analytics (LA) studies ecosystem at U-M. The project will use U-M's cyberinfrastructure and organizational capacities in LA e.g. the Michigan Institute for Data Science (MIDAS),[4] and the centralized U-M Data Warehouse. Critically, the University Library has revised its privacy policy such that data from 2016-present are available with identifiers. These resources and developments at U-M make our proposed project feasible within the three-year study period, and enhance our collaborative efforts. For example, we will share methods and findings with the project advisory group (PAG), a set of 12 diverse institutions (four of which are based in the State of Michigan), and iteratively solicit their feedback on the ease and feasibility of replicating our study. This will entail sharing with them resources, such as data dictionaries, that enable linkages across multiple datasets. This would pinpoint, for example, which identifiers to use to link library-use data to course instruction or grants data. U-M has the infrastructure to enable the types of data merges and analyses critical to the project. This is essential to accomplishing the goal of developing procedures and protocols that can be replicated at PAG and other institutions. Moreover, U-M's institutional review board has streamlined its protocols due to its increasing experience of working with learning analytics researchers.

Workplan, Goals, and Outcomes

We plan to analyze how the library impacts courses, research grants, and scholarly publications, thereby serving the needs of the community of faculty, fellows, and staff. That is, the project will focus on how the academic library anchors learning in the university. This will entail merging datasets from the university library and the U-M Data Warehouse. University library data is available in deidentified and identifiable forms, and will be used for the identification of typologies of library users using clustering algorithms, and subsequently for advanced statistical modeling of the associations between library user types and learning outcomes. Access to identifiable data enables us to crossmatch individuals across library use and learning outcomes datasets. It will also enable us to develop individual-level models of the links between specific types of library behaviors and interactions, and the learning outcomes of funded research, scholarly publications, and courses. We will share and discuss findings, methods, and protocols with PAG institutions (both electronically and face-to-face) over the entire project timeline (PAG letters of support available on request). PAG members are accomplished library faculty and administrators from the following institutions: DePaul University, Association of College & Research Libraries (ACRL), Big Ten Academic Alliance, New England College, Saint Mary's College of California, University of Illinois at Urbana-Champaign, Wayne State University, University of Toronto, Ferris State University, Western Michigan University, Ohio State University, and University of Detroit Mercy.

Dissemination Plan

The development of robust empirical measures of the academic library's contributions to learning is of high interest to individual universities and academic consortia. Starting Fall 2018, we will share our research findings at leading conferences including the Library Assessment Conference, the Coalition for Networked Information (CNI) Membership Meeting, and the ACRL Conference. We will also consult with the PAG to develop a report that explores how the project can be replicated at different types and sizes of institutions.

Budget

We are requesting $499,750 from IMLS for a three-year project period: $245,697 in salaries and wages, and $66,338 in fringe benefits to support the 12-person team of investigators, staff, graduate students, and consultants; $4,000 for dissemination travel; $5,000 for using the Flux cluster; and $178,715 in indirect costs.

---

[4] "About MIDAS." (http://midas.umich.edu/about/)