

Project Title

Getting to know their data doubles: An inquiry into student perceptions of privacy issues associated with academic library participation in learning analytics

Principal Investigator

Kyle M. L. Jones (MLIS, PhD), Assistant Professor
Department of Library and Information Science
Indiana University-Indianapolis (IUPUI)

1. Introduction. With research team members at the University of Wisconsin-Madison, the University of Wisconsin-Milwaukee, the University of Illinois at Chicago, Northwestern University, Oregon State University, and Indiana University, the principal investigator requests \$492,550 (including \$179,820 for indirect costs) to conduct a three-year research project into student perspectives of academic library participation in learning analytics (LA) initiatives with a focus on issues of student privacy. No known research exists that addresses these issues in relationship to library participation in LA. To close this knowledge gap, this project will pursue student-focused research consisting of interviews, cross-institutional surveys, and focus groups to develop an understanding of the issues from a student point of view, which will help build library capacity for LA.

2. Background and National Need. Learning analytics (LA) is the “measurement, collection, analysis, and reporting of [student data] for the purposes of understanding and optimizing learning and the environments in which it occurs.”^[1]* It is a socio-technical form of surveillance that enables institutional actors to monitor student behaviors and intervene in student lives by connecting information systems and developing data warehouses. LA proponents argue that analyzing student data will, inter alia, enhance pedagogy, reinforce student learning outcomes, and improve institutional efficiency.^[2,3,4]

Libraries are pursuing LA insights to evaluate the impact of library services and spaces on student learning. But, it is increasingly the case that libraries also want access to more data in order “speak to the value” they add to student learning in order to justify rising expenditures, especially considering the immense pressures universities are encountering regarding fiscal accountability.^[5] To this end, initial work has sought to understand correlations between student success (e.g., GPA, retention, degree attainment) and particular types of library use.^[6,7,8] Cutting-edge library LA initiatives have shared real-time library use data with instructors to intervene in student learning behaviors, as well as track reading activities in eBooks.^[9,10]

Ardent proponents of library participation in LA argue that aggregating granular (i.e., identifiable) student data, including library interaction data, yields the most useful data-driven insights.^[11] By creating infrastructures that connect campus information systems for data aggregation purposes, universities, with library participation, are building surveillant assemblages.^[12] In so doing, they create “data doubles,” essentially comprehensive profiles that enable institutions to use algorithms to describe, diagnose, and predict student behaviors—learning or otherwise—that affect learning outcomes.^[13] Libraries have long argued that privacy and intellectual freedom—what some call a “basic American value”—should be protected, but these surveillance practices threaten this core professional principle and put students at risk.^[14,15]

Very little research has addressed learning analytics (LA) and student privacy issues from a student perspective, and extant research suggests that the student voice is missing from LA conversations.^[16] Questions addressed by the literature include: the role of the Family Educational Rights and Privacy Act ^[17], autonomy and information justice problems ^[18,19], and professional ethics questions regarding library participation in LA initiatives.^[20,21] Some work that does explicitly address student perceptions asked leading questions and cannot be trusted.^[22] Other work highlights that the privacy issues at play are varied and need contextual understanding.^[23,24] To the team’s knowledge at the time of this writing, no scholarship currently exists that specifically considers the contextual issues associated with library data, library participation in LA, and student perceptions of their privacy when libraries are actively a part of LA initiatives.

3. Research Questions. If student privacy is the right and ability of a student to control information that captures intellectual and behavioral data about one’s self, then emerging data-driven practices lead to the

* Underlined numbers in brackets are linked to references and related works when the document is viewed as a PDF.

following thematic research questions: how do students feel about libraries engaging in learning analytics (LA) initiatives that use data (identifiable or otherwise) to: 1) predict their intellectual behaviors, 2) intervene in their academic life, and 3) plausibly share data about them to unknown actors outside the library, within the university, and to third parties?

4. Project Description. The project has three yearly phases. During year one, the research team will conduct preliminary interviews with students to identify themes about library participation in learning analytics (LA) and LA generally with regard to privacy. During year two, the research team will use interview findings to construct, test and deploy a survey to undergraduate and graduate students at each researcher's respective institution (n=7). Following this, each researcher will run focus groups with students to further investigate findings from the survey responses. This work will yield data from a broad sample of diverse students spread across unique public and private universities. During year three, the research team will disseminate the findings in scholarship, which will also inform the development of workshops that each researcher will run at his or her respective institution to build librarian understanding of the privacy issues and develop practical paths forward (e.g., data management practices, policy) with library participation in LA. Should funding be available, there is also an opportunity to run these workshops at other institutions. Project updates will also be posted on a publicly accessible website.

5. Impact. The project holds strong potential to have a local and national impact given the interest in the subject area and the need for more research. Locally, the aforementioned training sessions will help develop awareness about and capacity for learning analytics (LA), as well as signal that libraries are leaders in areas of data ethics and student privacy. Nationally, the team's commitment to publishing important findings in respected peer-reviewed journals and at scholarly and practitioner conferences holds potential to inform a range of academics and practitioners alike. The project team also believes this project complements and advances the work of recently funded IMLS grants in the area of LA and privacy generally.[see [25,26,27](#)]

6. Project Team. The research team represents a collaboration between academics and practitioners who hold expertise in the areas of data ethics and privacy, data management, library assessment, qualitative research, and survey design. Individually, researchers have presented and published in these areas, and together they are conducting an ARL SPEC Survey on learning analytics (LA). Advisors are supporting the project with their own scholarly and practitioner experience. *PI:* [Kyle M. L. Jones](#); *Researchers:* [Andrew Asher](#), [Kristin Briney](#), [Abigail Goban](#), [Michael Perry](#), [M. Brooke Robertshaw](#), and [Dorothea Salo](#); *Advisors:* [Alan Rubel](#), [Michael Zimmer](#), [Jan Fransen](#), and [Anne-Marie Deitering](#).

7. Budget. The PI requests \$492,550 in total, which includes the following:

Category	Year One	Year Two	Year Three	TOTAL
PI salary and benefits (n=1)	\$34,676	\$35,716	\$36,788	\$107,180
PI research assistant (n=1)	\$3,900	\$3,900	\$3,900	\$11,700
Researcher collaboration stipends (n=6)	\$53,400	\$53,400	\$53,400	\$160,200
Conference fees and travel for research team	\$7,967	\$7,967	\$7,967	\$23,900
Research materials, software licenses, etc.	\$2,133	\$2,133	\$2,134	\$6,400
Participant incentives	\$583	\$583	\$583	\$1,750
Advisory board stipends (n=4)	\$533	\$533	\$533	\$1,600
YEARLY SUB-TOTAL	\$103,192	\$104,232	\$105,306	\$312,730
YEARLY INDIRECT COSTS (57.5%)	\$59,335	\$59,933	\$60,551	\$179,820
TOTAL	\$162,527	\$164,165	\$165,857	\$492,550