Abstract

The research project will conduct a student-centered, three-year research agenda into student perspectives of privacy issues associated with academic library participation in learning analytics (LA) initiatives. Led by the primary investigator at Indiana University-Indianapolis (IUPUI), the team consists of research collaborators at the University of Wisconsin-Madison, the University of Wisconsin-Milwaukee, the University of Illinois at Chicago, Northwestern University, Oregon State University, Indiana University-Bloomington, and a site facilitator at Linn-Benton Community College. Six scholars and practitioner experts in the areas of assessment, library analytics, diversity, and information ethics and policy will support the team as they develop research protocols and disseminate findings.

Learning analytics (LA) is the "measurement, collection, analysis, and reporting of [student and other data] for the purposes of understanding and optimizing learning and the environments in which it occurs." With LA, institutions are more prepared to describe (what is happening?), diagnose (why did it happen?), predict (what is likely to happen?), and prescribe (what should we do about it?) student learning by identifying factors that impede or promote success. Libraries are pursuing LA insights to evaluate the impact of library services, collections, and spaces on student learning. The success of LA depends in part on an institution's ability to connect campus information systems—including those under the purview of libraries—to aggregate and analyze student data. But as institutions continue to surface granular data and information about student life, the risk to student privacy grows. It is unclear what rights to students have in relation to the data, and there is little evidence regarding student perceptions of LA—especially when libraries are involved.

Very little research has addressed LA and student privacy issues from a student perspective, and extant research suggests that the student voice is missing from LA conversations. To the team's knowledge at the time of this writing, no scholarship currently exists that specifically considers student perceptions of their privacy when libraries are actively leading or contributing to LA initiatives. In fact, in Connaway et al.'s OCLC-sponsored study, the authors argue that "this topic is particularly fraught in the areas of assessment and academic libraries since there is a lack of established effective practices and standards addressing the methods and contexts that may threaten the privacy of students." Because of these indicators, the team believes there is a national need to study library LA and the privacy issues from a student perspective.

The team seeks to answer this overarching research question: How do LA initiatives align with and run counter to student expectations of privacy; and with these insights, how might libraries maximize the benefits of LA while respecting student expectations?

Three iterative research phases structure this project. During phase one, the research team will conduct preliminary interviews with students to identify themes about library participation in LA and LA generally with regard to privacy. During phase two, the research team will deploy a survey to undergraduate and graduate students at each researcher's respective institution. In the third and final phase, each team member will run a series of scenario-based focus groups with students to explore possible applications of LA that respect and break expectations of privacy. All three phases will lead to peer-reviewed scholarship, practitioner-focused conference presentations, workshop materials, and a toolkit for informing library practitioners about student privacy and LA.

Full Proposal Narrative

1. STATEMENT OF NATIONAL NEED

Introduction

The research team requests $514,484 to conduct a student-centered, three-year research project into student perspectives of privacy issues associated with learning analytics (LA) initiatives and academic library participation in LA. The research team consists of collaborators at the University of Wisconsin-Madison, the University of Wisconsin-Milwaukee, The University of Illinois at Chicago, Northwestern University, Oregon State University, Indiana University-Bloomington, and Indiana University-Indianapolis (IUPUI). Each research team member's respective institution will serve as a sampling site, along with Linn-Benton Community College. The team will begin phase one of three by conducting an environmental scan of the professional and academic literature related to LA. Next, the team will conduct student interviews to identify their perceptions of privacy issues regarding data sources, uses, protections, and rights. From this data, the team will create a taxonomy of privacy problems and then, in phase two, deploy a multi-institutional survey at the eight institutions to investigate further these issues and how students with diverse demographic and academic backgrounds perceive the problems. Building off of the first two phases, phase three uses expert-vetted scenarios to explore how students negotiate their privacy values and expectations in relation to institutional stakeholder needs. The driving goal of the project is to build up a comprehensive understanding of student privacy perspectives, rights, and normative expectations to inform librarians and LA proponents as they build capacity to deploy the technology throughout higher education. The team will disseminate just-in-time findings, practitioner-centered presentations, a toolkit for practitioners to investigate student privacy on their campuses, institutional workshops, and peer-reviewed research—all of which will be available at a public project website and broadcast in practitioner social media networks.

The Need to Study Student Privacy and Learning Analytics

Higher education actors and the learning analytics (LA) literature generally acknowledge that student privacy is valuable. However, simply arguing that student privacy has value and should be defended fails to push forward 1) arguments that carefully explain *why* it has value and 2) *what* goods it reaps for students. We argue that higher education has historically supported student privacy because of its instrumental value in promoting autonomy and human flourishing, academic freedom, and relationship making, all things commonly understood to be beneficial in the educational experience.[1] American librarians have long argued that user privacy is worthy of protection, putting the value at the core of its code of ethics.[2] Librarian practitioners and scholars alike situate privacy as a core value because of its instrumental role in support of intellectual freedom. As a moral good, intellectual freedom contributes to an individual's ability to seek information in support of idea generation and speech—things that contribute to building one's identity and support one's participation in democratic societies.[3] To maximize intellectual freedom, external limitations, such as actual and potential invasions of privacy, should be limited, which is the approach librarians have historically supported. But as we discuss below after introducing the reader to LA, student privacy is a core moral problem with LA. If librarians are to maximize the benefits of LA while accounting for its potential harms, they need a better understanding of student privacy expectations.

Understanding Learning Analytics

Learning analytics (LA) is the "measurement, collection, analysis, and reporting of [student and other data] for the purposes of understanding and optimizing learning and the environments in which it

---

[1] DeNicola, 2012; Rachels, 1975; Richards, 2008; Rubel & Jones, 2016

[2] American Library Association, 2017

[3] Dresang, 2006; Richards, 2008

occurs."[4] It is a socio-technical form of surveillance that monitors student behaviors and measures learning outcomes by flowing data from information systems to central data warehouses for downstream analysis. LA practices commonly analyze:

1. learners' individual characteristics (e.g., prior knowledge, academic performance),
2. activities in the learning environment (e.g., user pathways, download activity),
3. curricular benchmarks (e.g., learning outcomes, historical course information),
4. and interactions with peers and teachers (e.g., social network activity).[5]

With LA, institutions are more prepared to describe (what is happening?), diagnose (why did it happen?), predict (what is likely to happen?), and prescribe (what should we do about it?) student learning by identifying factors that impede or promote success. Effectively, LA systems treat students as data "test subjects" to be statistically analyzed and acted upon with social and academic intervention strategies to influence student behaviors.[6] LA proponents argue that analyzing student data and acting on analytic insights will, inter alia, enhance pedagogy, reinforce student learning outcomes, shore up retention rates, and improve institutional efficiencies in part to combat limiting accountability measures in a time of decreased funding for higher education.[7] According to some LA proponents, not acting on this data would be unethical and an abdication of institutional responsibility.[8]

Academic Libraries and Learning Analytics

Libraries are pursuing LA insights to evaluate the impact of library services, collections, and spaces on student learning.[9] Such efforts are not new, of course. Academic librarians have been strategically using assessment strategies for more than ten years, especially in the area of information literacy.[10] But recent research suggests that institutional administrators expect their libraries to tie efforts and resource expenditures to student success, which forces these initiatives from the "nice to do" to the "must do" category.[11] As a result, it is increasingly the case that libraries want access to more data in order "speak to the value" they offer to their institutions in order to justify rising costs, especially considering the immense pressures universities are encountering regarding fiscal accountability.[12]

Initial library LA research has sought to understand correlations between student success (e.g., GPA, retention, degree attainment) and particular types of library use (e.g., time in the library, material use, database access, librarian interactions).[13] 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.[14] Ardent proponents of library participation in LA argue that it is only after aggregating granular (i.e., identifiable) student data, including library interaction data, that data-driven insights yield the most useful returns.[15]

A Student's Data Double

The success of LA depends in part on an institution's ability to connect campus information systems —including those under the purview of libraries—to aggregate and analyze "static and dynamic"

---

[4] Siemens, 2012
[5] Ifenthaler & Schumacher, 2016, p. 924
[6] Bienkowski, Feng, & Means, 2012
[7] Brown, 2011; Long & Siemens, 2011; van Harmelen & Workman, 2012; Watters, 2011
[8] Campbell, DeBlois, & Oblinger, 2007; Willis III, 2013
[9] ACRL Research Planning and Review Committee, 2016
[10] Project SAILS, 2017; RAILS, n.d.
[11] Connaway, Harvey, Kitzie, & Mikitish, 2017
[12] Jantti & Cox, 2013
[13] Collins & Stone, 2014; Jantti, 2016; Nackerud, Fransen, Peterson, & Mastel, 2013
[14] Jantti, 2016; Magi, 2010
[15] Oakleaf, Walter, & Brown, 2017

student data.[16] Haggerty and Ericson's concept of the "data double" helps to explain the value and power of these data practices.[17] According to their analysis, data doubles emerge when assemblages of informational networks can abstract humans into data subjects for analysis and intervention. To understand how data doubles are emerging, readers need to understand how data sources and information flows in of support of LA and some of the ends to which institutions put LA practices. Specifically, the team will briefly outline data and informational practices that create student profiles, track behaviors, surface affective states, and analyze physiological markers.

Institutions begin developing data doubles during the admissions process. Students reveal among other things personal traits and experiences, demographic details, socioeconomic information, and professional dreams in their application that drive analytic practices.[18] Once admitted, institutions codify much of this student information in central student information systems (SISs) or archive it in admissions databases—both of which are augmented over time with academic information.[19] Often times, these student profiles serve up the variables LA projects correlate with learning behaviors and outcomes, including gender, age, ethnicity, first-generation status, etc.[20]

Once students are engaged in campus life, they usually have to identify themselves to get access to campus resources and information systems. In doing so, they leave "digital traces" or "breadcrumbs" that can reveal their physical and digital behaviors, including what they click on, with whom they communicate, and their information searching activities.[21] RFID-embedded student IDs, ID card swipes, IP tracking, and WiFi logs also enable institutions to track student movements on campus.[22] LA can then measure these behaviors in comparison with student peers and attempt to predict student success rates in a given course or program, and create personalized interventions to nudge students to take action.[23]

With LA, things once "unseen, unnoticed, and therefore unactionable" in the learning experience become datafied and actionable—including a student's emotional state and physical experiences.[24] Automated discourse analysis of student texts (e.g., discussion posts, e-mails) holds the potential to build a more "holistic picture" of student life by revealing their affective states, such as their level of self-confidence in a subject, a general level of happiness, or even depression.[25] In addition to improving learning outcomes, monitoring student communications and employing sentiment analysis may hold the potential to identify students on the verge of a violent meltdown, which could effectively improve campus security and increase the community's sense of safety.[26] Students' data doubles may also be augmented by data about their *actual* body as LA proponents build up capacity for biometrics. Analyzing facial features and heart rates, as some researchers have, shows potential for better understanding how students' physical behaviors correlate with classroom engagement and learning.[27]

Privacy Problems

As data doubles have emerged, the risk to student privacy has grown. Students have legal privacy rights as codified in the Family Educational Rights and Privacy Act (FERPA), but it is not clear

---

[16] Ifenthaler, 2015; Long & Siemens, 2011
[17] Haggerty & Ericson, 2000, p. 606
[18] Goff & Shaffer, 2014
[19] Dawson, Gašević, Siemens, & Joksimovic, 2014
[20] Nackerud et al., 2013
[21] Almosallam & Ouertani, 2014
[22] Bradberry, Ray, Wayman, Dhami, Charnock, & Pittges, 2017; Brazy, 2010
[23] Wildavsky, 2013, para. 16
[24] Bienkowski et al., 2012, p. ix
[25] Chen, Vorvoreanu, & Madhavan, 2014; Ferguson, 2012; Ferguson & Buckingham Shum, 2012; Siemens, 2012
[26] Crow, 2012
[27] Loewus, 2012, para. 1; Simon, 2012

whether or not students have a right to limit access to or amend data their institutions create about them for LA purposes.[28] Of course, expressing these rights first assumes that students are aware of their rights and the existence of LA practices; such things, however, can be very opaque to students, even though student lives are very transparent to institutional actors.[29] Empirical findings still have yet to prove that the benefits of LA justify the systemic surveillance, measurement, and classification of student life.[30] Until verified evidence emerges, students will plausibly perceive LA as intrusive and unfair.

Such privacy issues would not be problematic if colleges and universities knew how to address them, but the literature signals that they do not. In fact, the "existing legal and ethical maps" have little applicability to emerging data mining practices.[31] Institutions are unsure of the right ethical paths to take, in part because they have historically had a difficult time developing policies for privacy in the first place.[32] A lack of guiding policy and an unfamiliarity with unique ethical issues associated with LA has left colleges and universities scrambling for some form of ethical guidance.[33] According to Berg, without such guidance, institutional actors may act in haste, creating harmful downstream effects to student privacy.[34]

The Knowledge Gap

Very little research has addressed LA and student privacy issues from a student perspective, and extant research suggests that the student voice is missing from LA conversations.[35] Questions addressed include: legal rights and limitations, autonomy and information justice problems, and professional ethics questions regarding library participation in LA initiatives.[36] Some work that does explicitly address student perceptions asked leading questions and cannot be trusted.[37] Other work highlights that the privacy issues at play are varied and contextual.[38]

To the team's knowledge at the time of this writing, no scholarship currently exists that specifically considers student perceptions of their privacy when libraries are actively leading or contributing to LA initiatives. In fact, in Connaway et al.'s OCLC-sponsored study, the authors argue that "this topic is particularly fraught in the areas of assessment and academic libraries since there is a lack of established effective practices and standards addressing the methods and contexts that may threaten the privacy of students."[39] Because of these indicators, the team believes there is a national need to study library LA and the privacy issues from a student perspective.

2. PROJECT DESIGN

Theoretical Framework

Privacy is a "chameleon-like" concept: it can represent different values and perspectives depending on contextual characteristics.[40] So, we adopt Helen Nissenbaum's theory of contextual integrity (CI) as our theoretical framework.[41] CI "accounts for a right to privacy in personal information... in terms of

---

[28] Rubel & Jones, 2016
[29] Richards & King, 2013
[30] Dringus, 2012; Swenson, 2014
[31] King & Richards, 2014, para. 3
[32] Ferguson, 2012; Pardo & Siemens, 2014
[33] Sclater, 2016
[34] Berg, 2013
[35] Roberts, Howell, Seaman, & Gibson, 2016
[36] Johnson, 2017; Jones & Salo, forthcoming - 2018; Rubel & Jones, 2016; Showers, 2015; Zeide, 2017
[37] see Arnold & Sclater, 2017
[38] Ifenthaler & Schumacher, 2016; Roberts et al., 2016
[39] Connaway et al., 2017, p. 30
[40] BeVier, 1995; Nissenbaum, 2004
[41] Nissenbaum, 2004; Nissenbaum, 2010

appropriate flow."[42] Put differently, CI posits that there is no universal right to or value of privacy: such rights and values are rooted in particular contexts and are regulated by context-specific informational norms, or expectations regarding how information flows, to whom, and under what conditions. CI is a good fit for this project's research questions (RQs) and goals. The RQs drive our design of research protocols that will surface student expectations regarding their privacy and how they believe identifiable data and information should flow for LA initiatives. Using these contextual expectations, we ultimately aim to develop useful insights and recommendations for librarians and other institutional stakeholders to aid them in their design of LA technologies, services, and practices such that they are in alignment with context-specific informational norms students identify in the research.

Research Questions

We have four general research questions:

RQ 1: What privacy issues do students identify when informed about library LA initiatives, practices, data types, and data sources?
RQ 2: How do the identified privacy issues map to particular goals of LA initiatives by specific stakeholders (e.g., librarians, instructors, advisors, etc.)?
RQ 3: How do privacy perceptions change according to relevant student demographics and academic experiences?
RQ 4: With regard to their privacy expectations, what library and non-library LA scenarios are acceptable to students, how do they explain the variations in acceptability, and what recommendations would students make to resolve existing privacy problems?

Project Structure

We frame the project with three phases focused on data collection and analysis using a variety of methods. During phase 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 phase two, building off of the interview data, the research team will construct, test and deploy a survey to undergraduate and graduate students at each researcher's respective institution (n=7 institutions). In the third and final phase, each team member will run a series of scenario-based focus groups with students to explore possible applications of LA that respect and break expectations of privacy. We fully describe each phase in the following subsections after providing the project goals. Please see the Gantt chart in the schedule of completion document for more information on each phase's schedule and intermediate outcomes.

Project Goals

The research team seeks to achieve the following overarching goals:

1. Produce a rigorous account of student perceptions of academic library participation in LA initiatives;
2. create an empirically-driven and theoretically-sound definition of student privacy;
3. develop recommendations for practice, policy, and technological design to account for student privacy expectations;
4. increase practitioner awareness of student privacy issues related to LA;
5. engage the practitioner community in conversations about LA and student privacy;
6. and empower academic librarians to add to conversations about LA on their campuses.

Phase One: Environmental Scanning (May, 2018 through April, 2019)

Environmental scans provide a snapshot of an emerging technology and its socio-political context. At the same time, they identify actual practices, identify future opportunities, and reveal moral, ethical, and political tensions.[43] The team will begin its work in phase one with an environmental scan by developing a comprehensive literature review and completing interview-based research with student

---

[42] Conley, Datta, Nissenbaum, & Sharma, 2012, p. 772

[43] Amanatidou et al., 2012

participants to establish a baseline concerning student privacy and academic library participation in LA initiatives. It is necessary to develop an understanding of the extant literature and curate initial data specifically on student privacy perceptions before developing the multi-institutional survey that we will deploy in phase two; without such a baseline, the survey may be misinformed. The team expects to interview at minimum 15 students per institution, resulting in a sample size of at least 120 students. After transcribing the interviews, the PI will lead the team as they collaboratively code the data in Dedoose, a web-based qualitative research tool, to develop thematically-sound categories and inductive, grounded theoretical concepts (i.e., not concepts designed into an *a priori* codebook).[44]

Questions Driving Empirical Research

- How is LA defined and by whom?
- What are the educational and socio-political motivations for library participation in LA?
- What are the exemplary library LA practices, what are the ends they are trying to accomplish, and what methods (e.g., data sources, statistics) do they use to achieve those ends?
- What are the outstanding moral and ethical issues—specifically as they relate to student privacy issues—associated with LA and how do they interact with library participation?
- What is the student perception of LA?
- How do students define student privacy?
- How do students expect their institutions and libraries to use identifiable student data?

Performance Categories and Outcomes

Research	Dissemination and Practitioner Outreach
• Write and publish comprehensive literature review to website; begin writing annotated bibliography for Toolkit • Complete interview-based research • Submit interview-based paper for peer review and publication	• Build and publish project website • Publish six project updates on project website • Publish a summary of phase one findings on project website • Draft phase one Toolkit materials • Submit two conference presentation proposals

Phase Two: Surveys (February, 2019 through April, 2020)

After developing an initial dataset from the interviews, the research team will use its findings to develop thematic survey questions concerning student privacy, especially with a focus on library participation in LA. The team will distribute the survey at each team member's institution using a sampling strategy inclusive of students with academically and personally diverse backgrounds. The survey will develop a large corpus of quantitative and qualitative data on student privacy that is currently lacking in the literature. Analysis of this data will better inform librarians and other institutional actors about how students define and value their privacy in relation to particular data sources, practices, and end goals. The team expects to develop a representative sample and invite around 3,000 participants at each institution to complete the survey. Assuming only a 10% response rate, the complete sample would amount to 2,400 students across eight institutions. The team will employ survey design standards, build and run the survey in Qualtrics, and pre-test the survey to limit survey fatigue effects.[45] We will analyze the survey's quantitative data using descriptive and inferential statistical methods and follow coding best practices for the qualitative data (e.g., open-ended questions), again using Dedoose as a team.[46]

Questions Driving Empirical Research

- How do students think their institution and library are *currently* using student data and information?
- What data uses do students find problematic for their privacy?

---

[44] Charmaz, 2014; Richards, 2015

[45] Ben-Nun, 2008; Dillman, Smyth, & Christian, 2014

[46] Richards, 2015

- What data practices do students find to be an unjustifiable use of their identifiable data and information?
- How might student behaviors change with particular LA uses of identifiable data and information?
- Do students expect to be informed about and/or given an opportunity to consent to LA practices?
- How do student demographics and academic profiles explain perceptions regarding their privacy?

Performance Categories and Outcomes

Research	Dissemination and Practitioner Outreach
• Work with advisors to develop questions for survey • Coordinate with institutional representatives to develop sample populations • Test protocol and complete survey research • Submit survey-based paper for peer review and publication	• Publish six project updates on project website • Publish a summary of phase two findings on project website • Draft phase two Toolkit materials • Submit two conference presentation proposals, one for a practitioner audience and one for a scholarly audience

Phase Three: Participatory Focus Groups (January, 2020 through April, 2021)

In phase three, the team will use the empirical findings from the past two phases to develop and conduct scenario-based discussions with groups of students. The overarching research goal for this phase is to engage students in thinking through a multiplicity of plausible "futures" for LA in ways that "aid the imagination, stimulate creativity, and reveal novel possibilities" for related practices in libraries.[47] When working through the scenarios, students will be asked to reflect on how common student perceptions of privacy limit or enable certain LA practices; additionally, we will engage students in conversations about how to develop solutions to seemingly intractable ethical problems. Data from this research will lead to student-created recommendations regarding library services, resources, and policy that reflect their data use expectations. Findings from this phase will aid responsible innovation and policy development by LA technologists and academics. The team expects to run at minimum three focus groups per institution with around five students per group, resulting in a total of 24 group sessions with at least 120 participants. We will video record and later transcribe focus groups. Either a research team member or student assistant will take comprehensive notes during focus groups. With the PI's lead, the team will analyze recordings, transcriptions, and research notes according to standard focus group research methods.[48]

Questions Driving Empirical Research

- How might we envision new library services built on student data?
- What about a given scenario creates privacy problems?
- How do stakeholder goals and values influence student responses to scenarios?
- How might problems in a scenario be resolved using education, policy, and technological design?

Performance Categories and Outcomes

Research	Dissemination and Practitioner Outreach
• Build scenarios for focus groups with input from library LA stakeholders • Test focus group protocols • Run focus groups at each institution • Submit focus group-based paper for peer review and publication	• Publish six project updates on project website • Publish a summary of phase two findings on project website • Complete Toolkit, publish to project website • Run professional development workshops • Submit two conference presentation proposals

[47] Kahn, 1971, p. 150; Rowland & Spaniol, 2015; Wiek, Keeler, Schweizer, & Lang, 2013

[48] Stewart & Shamdasani, 2014

Risks and Assumptions

Research and anecdotal experience suggests that students are mostly unaware of emerging data infrastructures in higher education and the extent to which such things enable student surveillance. Our conversations with students will bring this knowledge to their attention, and we are cognizant of the fact that some students may feel uncomfortable with such information. Discomfort could result in changed behaviors and affect relationships students have with librarians, instructors, and other institutional actors. In fact, we have seen this occur at other institutions when students become aware of sensitive institutional record-keeping practices.[49] Consequently, we will develop research protocols that introduce students to learning analytics (LA) in a fair and balanced manner without unnecessarily and prematurely raising privacy concerns; we will also direct students to their respective institution's registrar and chief information officers if they have questions about how their institution uses identifiable data and information. Regardless of the potential problems, the benefits of engaging students in privacy conversations outweigh the harms because of the positive educational effects and the potential to align LA practices with student privacy expectations in the future.

Dissemination Plan

The team has a three-pronged strategy for disseminating research findings speedily and widely, which involves developing a digital presence, communicating to practitioner communities, and seeking publication and presentation of scholarly work in respected and highly visible outlets.

Digital Presence

The team will develop a blog-based website to document project progress and facilitate the quick publication of project updates. We will use our digital presence to push updates to the practitioner and scholarly communities when we have reflections, insights, and emerging findings. The website will also automatically publish updates to a project Twitter account; individual team members will distribute these updates to their social networks using Twitter. The website will also serve as a clearinghouse to distribute all publications, presentations, and research artifacts available within limits set by copyright restrictions, institutional review boards, and the team's data management plan, which is outlined in digital product form document.

Practitioner Outreach and Education

We recognize that the scholarly work we complete is only so useful if published just in traditional outlets (e.g., peer-reviewed journals). Therefore, we will actively seek to present and publish researching findings and insights in the following outlets:

Targeted Practitioner Conferences	Targeted Practitioner Publication Outlets
• ACRL Annual • ALA Annual • ARL Library Assessment • EDUCAUSE • Performance Measurement in Libraries	• *College and Research Library News* • *EDUCAUSE Review* • *Chronicle of Higher Education* • *Inside Higher Ed* • *Library Journal*

In addition to our broad practitioner outreach, we aim to have a local impact by creating research-informed training sessions for colleagues at our respective institutions. As a team, we will co-develop workshops, and each team member commits to teaching one workshop at her/his institution.

Finally, the team will develop "Libraries, Learning Analytics, and Student Privacy: A Toolkit," which we will distribute on the project's website and deposited into IUPUI's institutional repository in the final

[49] see Lifshits, 2015; Pérez-Peña, 2015

year.[50] Disseminating the Toolkit will empower practitioners to educate themselves on the subject area and replicate our studies if they so choose. The Toolkit will be released under a Creative Commons Attribution Non-Commercial 4.0 (CC-BY-NC) license and include the following elements:

- A white paper including an overview of the project, a comprehensive literature review, and a summary of major findings;
- an annotated, thematic bibliography of useful literature;
- all presentation slide decks created during the project's duration;
- interview protocols;
- the survey questionnaire;
- and scenario materials used in the focus groups.

Scholarly Publications

Finally, the research team will disseminate the findings in scholarly publications. Each phase lends itself to developing an empirically-based paper based on the method employed. The iterative and multiple method design of the project holds the potential to develop significant findings. Our target journals include, inter alia, *College & Research Libraries*, *Library & Information Science Research*, and *Library Quarterly*.

Project Team

Primary Investigator and Collaborative Researchers

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 and quantitative research methods (e.g., constructivist grounded theory, face-to-face and web-based interviews, focus groups, and surveys). Individually, researchers have presented in these areas and published in peer-reviewed journals; together, they are conducting an ARL SPEC Survey on learning analytics (LA). The primary investigator's research agenda, Kyle M. L. Jones (Indiana University-Indianapolis), focuses solely on LA and related information policy, information ethics, and social informatics concerns.[51,52] Collaborative researchers include: Andrew Asher (Indiana University-Bloomington), Kristin Briney (University of Wisconsin-Milwaukee), Abigail Goben (University of Illinois at Chicago), Michael Perry (Northwestern University), M. Brooke Robertshaw (Oregon State University), and Dorothea Salo (University of Wisconsin-Madison). See each researcher's two-page resume for more information.

Advisors

Each advisor has scholarly expertise and/or practitioner experience that will help optimize the research methods and maximize the insights born from the project. Advisors include: Ira Bennett (Arizona State University), Anne-Marie Deitering (Oregon State University), Jan Fransen (University of Minnesota-Twin Cities), Wayne Hilson, Jr. (Indiana University-Indianapolis), Alan Rubel (University of Wisconsin-Madison), and Michael Zimmer (University of Wisconsin-Milwaukee).

Project Management and Staff Roles

Primary Investigator

The primary investigator (PI), will lead all aspects of research protocol development, data analysis, and dissemination and is responsible for managing the project's overall progress in order to accomplish detailed and time-sensitive tasks. The PI will use Freedcamp, a project management system, to schedule major and minor deliverables, delegate tasks, and facilitate communication with

---

[50] Although it is difficult to project what the toolkit will ultimately look like, we will structure it in part on OCLC's "Sharing, Privacy and Trust in Our Networked World" report available here: http://www.oclc.org/content/dam/oclc/reports/pdfs/sharing.pdf

[51] Access the PI's presentations and publications at his professional research page: http://thecorkboard.org/research/

[52] Underlined names link to researcher and advisor profiles when the document is viewed as a PDF.

the team. As the team is already using Freedcamp for their collective work on another joint project (the ARL SPEC Survey), this tool is already a part of their workflow.

Research Collaborators

Each research collaborator, with support of a student research assistant, is responsible for running research protocols at her/his respective institution with the support of the PI. This work includes, among other things, receiving institutional review board support for research protocols, recruiting participants, facilitating on-site data collection, and running end-of-project workshops. Research collaborators have separated themselves into sub-teams to take ownership of thematic areas of work according to interests and expertise; the PI will lead all sub-teams. See the supplemental document titled “Sub-Team Details” (Supportingdoc1.pdf) for more information.

Advisors

The team will consult with advisors primarily on an ad hoc basis in relationship to their area of expertise. However, the team will strategically consult all advisors when drafting research protocols and sampling strategies, analyzing data and finalizing major publications (e.g., research papers), and drafting the Toolkit.

3. ASPECTS OF DIVERSITY AND INCLUSION

We will address diversity by strategically sampling across our student bodies. The intersectionality of personal characteristics (e.g., socioeconomic status, gender, sexuality, race, religion, ethnicity, educational background, etc.) creates diverse student populations with unique values and perspectives. These things often intersect with one’s definition and value of privacy—informational, personal, or otherwise. An advisor, Wayne Hilson, Jr., will aid our efforts to address diversity in the research. Hilson has expertise in addressing diversity issues and developing inclusive strategies as the senior director of the Division of Diversity, Equity, and Inclusion at IUPUI. With Hilson’s support, we will craft sampling strategies that reflect our student bodies and proactively seek out non-majority voices and experiences. See the supplementary section titled “A Comparison of Institutional Profiles” (Supportingdoc2.pdf) for student body data for each institution. We will also work with other advisory board members to consider theoretical, practical, and institutional perspectives and value sets that are not our own, in so doing we will account for our privilege and biases.[53]

4. NATIONAL IMPACT

Students are at the heart of higher education’s mission and drive its efforts. As such, they are a key community stakeholder, and institutions need to hear the student voice as they move towards systematic data-driven decision making—especially when students are the primary target of data mining and analysis practices. This work will highlight student expectations regarding their privacy at a scale not seen in the literature, and it holds the potential to inform library and non-library LA initiatives alike. While identifying privacy concerns will highlight areas where institutions and their libraries need to develop policy and change data practices, it will also highlight areas where privacy is *less* of concern to students in ways that can open up opportunities to create new user services and technologies. The outreach aspects of the project will inform librarians and develop their competencies around LA, which will enable them to lead their institutional peers in conversations about student privacy and work constructively with their administrators to develop LA tools and practices. This project also holds the potential to inform future work post-grant funding, including more research on LA data ethics and developing training workshops and educational programming. Finally, IMLS has funded work on on developing capacity for LA (see IMLS grant number LG-98-17-0019-17) and privacy generally (see IMLS grant number LG-73-17-0062-17)—but not the two areas combined. We believe the project complements and advances the work of these recently funded IMLS grants.

---

[53] Walter & Andersen, 2013; Zuberi & Bonilla-Silva, 2008

Schedule of Completion - Grant Year One

		GRANT YEAR: ONE	2018								2019			
TASK NAME	**CATEGORY**	**PHASE**	**May**	**June**	**July**	**August**	**September**	**October**	**November**	**December**	**January**	**February**	**March**	**April**
Hire research assistants for phase one	Project Administration	One												
Comprehensive literature review	Research: Literature	One												
Build and publish project website	Outreach: Development	One												
Develop interview protocol	Research: Design	One												
Test interview protocol	Research: Design	One												
Indiana University IRB approval	Research: Design	One												
Research team institutional IRB approval	Research: Design	One												
Recruit interview participants	Research: Design	One												
Conduct interviews	Research: Data Collection	One												
Analyze interviews	Research: Analysis	One												
Write phase one scholarly paper	Research: Writing	One												
Edit and submit phase one scholarly paper for peer review and publication	Research: Submission for Publication	One												
Draft phase one toolkit materials	Outreach: Artifact Development	One												
Publish phase one findings summary to project website	Outreach: Artifact Dissemination	One												
Submit year one grant report to IMLS	Grant Administration	One												
Submit conference proposal - Library Assessment Conference (2019)	Research: Presentation	One												
Submit conference proposal - ACRL (2019)	Research: Presentation	One												
Develop survey	Research: Design	Two												
Test and validate survey	Research: Design	Two												
Indiana University IRB approval	Research: Design	Two												

Schedule of Completion - Grant Year Two

		GRANT YEAR: TWO	2019								2020			
TASK NAME	**CATEGORY**	**PHASE**	**May**	**June**	**July**	**August**	**September**	**October**	**November**	**December**	**January**	**February**	**March**	**April**
Hire research assistants for phase two	Project Administration	Two												
Research team institutional IRB approval	Research: Design	Two												
Develop recruitment lists with respective institutional research offices	Research: Design	Two												
Run survey	Research: Data Collection	Two												
Analyze survey results	Research: Analysis	Two												
Write phase two scholarly paper	Research: Writing	Two												
Edit and submit phase two scholarly paper for peer review and publication	Research: Submission for Publication	Two												
Draft phase two toolkit materials	Outreach: Artifact Development	Two												
Publish phase two findings summary to project website	Outreach: Artifact Dissemination	Two												
Submit year two grant report to IMLS	Grant Administration	Two												
Submit conference proposal - EDUCAUSE (2020)	Research: Presentation	Two												
Submit conference proposal - ALA Annual (2020)	Research: Presentation	Two												
Develop focus group scenario protocol	Research: Design	Three												
Test focus group scenario protocol	Research: Design	Three												
Indiana University IRB approval	Research: Design	Three												
Research team institutional IRB approval	Research: Design	Three												

Schedule of Completion - Grant Year Three

		GRANT YEAR: THREE	2020								2021			
TASK NAME	**CATEGORY**	**PHASE**	**May**	**June**	**July**	**August**	**September**	**October**	**November**	**December**	**January**	**February**	**March**	**April**
Research team institutional IRB approval (continued)	Research: Design	Three												
Recruit focus group participants	Research: Design	Three												
Run focus groups	Research: Data Collection	Three												
Analyze focus group data	Research: Analysis	Three												
Write phase three scholarly paper	Research: Writing	Three												
Edit and submit phase three scholarly paper for peer review and publication	Research: Submission for Publication	Three												
Develop institutional professional development workshops	Outreach: Artifact Development	Three												
Run professional development workshops	Outreach: Artifact Dissemination	Three												
Draft phase three toolkit materials	Outreach: Artifact Development	Three												
Publish phase three findings summary to project website	Outreach: Artifact Dissemination	Three												
Edit and publish toolkit materials into one cohesive toolkit to project website	Outreach: Artifact Dissemination	Three												
Submit year three grant report to IMLS	Grant Administration	Three												
Submit conference proposal - AERA (2021)	Research: Presentation	Three												
Submit conference proposal - ACRL (2021)	Research: Presentation	Three												

DIGITAL PRODUCT FORM

Introduction
The Institute of Museum and Library Services (IMLS) is committed to expanding public access to federally funded digital products (i.e., digital content, resources, assets, software, and datasets). The products you create with IMLS funding require careful stewardship to protect and enhance their value, and they should be freely and readily available for use and re-use by libraries, archives, museums, and the public. However, applying these principles to the development and management of digital products can be challenging. Because technology is dynamic and because we do not want to inhibit innovation, we do not want to prescribe set standards and practices that could become quickly outdated. Instead, we ask that you answer questions that address specific aspects of creating and managing digital products. Like all components of your IMLS application, your answers will be used by IMLS staff and by expert peer reviewers to evaluate your application, and they will be important in determining whether your project will be funded.

Instructions

☐ Please check here if you have reviewed Parts I, II, III, and IV below and you have determined that your proposal does NOT involve the creation of digital products (i.e., digital content, resources, assets, software, or datasets). You must still submit this Digital Product Form with your proposal even if you check this box, because this Digital Product Form is a Required Document.

If you ARE creating digital products, you must provide answers to the questions in Part I. In addition, you must also complete at least one of the subsequent sections. If you intend to create or collect digital content, resources, or assets, complete Part II. If you intend to develop software, complete Part III. If you intend to create a dataset, complete Part IV.

Part I: Intellectual Property Rights and Permissions

A.1 What will be the intellectual property status of the digital products (content, resources, assets, software, or datasets) you intend to create? Who will hold the copyright(s)? How will you explain property rights and permissions to potential users (for example, by assigning a non-restrictive license such as BSD, GNU, MIT, or Creative Commons to the product)? Explain and justify your licensing selections.

For the sake of clarity throughout this document, “digital products” include artifacts that will be made available to the general public and include the following:

- Website project updates (i.e., blog posts);
- published articles;
- and all documents included in the Toolkit, which will be made up of:
    - A white paper including an overview of the project, a comprehensive literature review, and a summary of the project’s major findings;
    - an annotated, thematic bibliography of useful literature;
    - a complete package of all presentation slide decks created during the project’s duration;
    - interview protocols;
    - the survey questionnaire;
    - and scenario materials used in the focus groups.

All digital products will be licensed under a Creative Commons Attribution Non-Commercial 4.0 (CC-BY-NC) license with exceptions. Digital products will include the CC-BY-NC designation and content consumers will be directed to the appropriate Creative Commons webpage for more information (https://creativecommons.org/licenses/by-nc/4.0/). It is the team’s intent to make our products as widely distributable and reusable as possible, and the CC-BY-NC helps us to accomplish these goals. Regarding exceptions, we may experience restrictions on the use of the CC-BY-NC license when publishing peer-reviewed work in journals. We will work with journal editors to try to negotiate for the application of the CC-BY-NC license.

Research artifacts (e.g., project notes, communications, etc.) and empirical data sets are *not* considered digital products because they may include identifiable information about participants that raise privacy concerns. Therefore, we will not make this items publicly available. We address these concerns in A.3. and Part III.

A.2 What ownership rights will your organization assert over the new digital products and what conditions will you impose on access and use? Explain and justify any terms of access and conditions of use and detail how you will notify potential users about relevant terms or conditions.

The organizations for whom the research team works assert no ownership rights over digital products except where research data is concerned. Digital products will be made openly accessible on either the project website or in

ScholarWorks, an institutional repository (https://scholarworks.iupui.edu/). Our employers may assert ownership rights over the research data—we are the stewards of the data.

A.3 If you will create any products that may involve privacy concerns, require obtaining permissions or rights, or raise any cultural sensitivities, describe the issues and how you plan to address them.

Empirical data (e.g., sampling lists, interview recordings, survey responses, and focus group recordings) creates privacy concerns for participants. The team will take three precautions to protect participants. First, the team will collaboratively develop and agree to a protocol for the proper handling of sensitive data, which will account for things such as access limitations, versioning, copy restrictions, encryption responsibilities, and locked file workflows. Second, all digitally recorded empirical data and research notes will be stored on encrypted servers using the Box cloud service. Box is an Indiana University approved third-party provider of secure document storage services. Only research collaborators will have access to a shared folder inclusive of research data and research notes. Third, all empirical aspects of the project will undergo institutional review board (IRB) review and approval. We will follow their guidelines for obtaining informed consent and protecting participant privacy strictly.

Part II: Projects Creating or Collecting Digital Content, Resources, or Assets

A. Creating or Collecting New Digital Content, Resources, or Assets

A.1 Describe the digital content, resources, or assets you will create or collect, the quantities of each type, and format you will use.

Digital Artifact	Quantities	Format
Audio recordings	~105	.mp3
Blog posts	18 (6 per year over three years)	Combination of HTML, CSS, PHP, and multimedia assets
Conference presentations	6	.pptx
Qualitative data	Unknown; includes researcher notes which cannot be predicted	.docx, .pdf
Quantitative data	Unknown; depends on amount of versioning which cannot be predicted	.xslx, .spss
Research articles	3	.docx, .pdf
Scenario documents	~21	.pdf
Survey responses	7 institutional data sets	.xslx
Toolkit	1	.pdf
Video recordings	~21	.mp4
Website	1	Combination of HTML, CSS, PHP, and multimedia assets

A.2 List the equipment, software, and supplies that you will use to create the content, resources, or assets, or the name of the service provider that will perform the work.

Equipment, software, supplies, service provider
Adobe Acrobat

Equipment, software, supplies, service provider
Apple desktop publishing applications
Business-grade computing devices
Dedoose
Digital audio recorders
Handheld video recorders
Microsoft desktop publishing applications
Mindnode
Qualtrics
Quicktime
SPSS
WordPress (and related server infrastructure)

A.3 List all the digital file formats (e.g., XML, TIFF, MPEG) you plan to use, along with the relevant information about the appropriate quality standards (e.g., resolution, sampling rate, or pixel dimensions).

Digital File Formats	Quality Standards
HTML	At least HTML4
CSS	At least CSS3
PHP	At least PHP 5.2.4
.mp3	Device default
.mp4	Device default
.xslx	Not applicable; application default
.pdf	Not applicable; application default
.docx	Not applicable; application default
.spss	Not applicable; application default
.pptx	Not applicable; application default

B. Workflow and Asset Maintenance/Preservation

B.1 Describe your quality control plan (i.e., how you will monitor and evaluate your workflow and products).

The research team will review and oversee the workflow and product creation. Kyle Jones, the PI, will maintain responsibility for artifact organization and fidelity to workflow standards.

B.2 Describe your plan for preserving and maintaining digital assets during and after the award period of performance. Your plan may address storage systems, shared repositories, technical documentation, migration planning, and commitment of organizational funding for these purposes. Please note: You may charge the federal award before closeout for the costs of publication or sharing of research results if the costs are not incurred during the period of performance of the federal award (see 2 C.F.R. § 200.461).

Data will be retained for a minimum of three years after the end of the project, as mandated by federal guidelines. Sensitive data will be stored securely and deleted at the end of the retention period. The project website will be made available for at least five years after the end of the project. The Toolkit will be maintained in ScholarWorks with no expected retention limit. Responsibility for maintaining published articles rests with the journals.

C. Metadata

C.1 Describe how you will produce any and all technical, descriptive, administrative, or preservation metadata. Specify which standards you will use for the metadata structure (e.g., MARC, Dublin Core, Encoded Archival Description, PBCore, PREMIS) and metadata content (e.g., thesauri).

Any applicable metadata will be associated with documents deposited in ScholarWorks according to its metadata schema.

C.2 Explain your strategy for preserving and maintaining metadata created or collected during and after the award period of performance.

Not applicable.

C.3 Explain what metadata sharing and/or other strategies you will use to facilitate widespread discovery and use of the digital content, resources, or assets created during your project (e.g., an API [Application Programming Interface], contributions to a digital platform, or other ways you might enable batch queries and retrieval of metadata).

Not applicable.

D. Access and Use

D.1 Describe how you will make the digital content, resources, or assets available to the public. Include details such as the delivery strategy (e.g., openly available online, available to specified audiences) and underlying hardware/software platforms and infrastructure (e.g., specific digital repository software or leased services, accessibility via standard web browsers, requirements for special software tools in order to use the content).

Digital products will be made available to the general public using the project website, ScholarWorks, and any journal dissemination targets (e.g., print journals, databases, etc.).

D.2 Provide the name(s) and URL(s) (Uniform Resource Locator) for any examples of previous digital content, resources, or assets your organization has created.

The project team has not created any previous digital artifacts as a team. ScholarWorks will provide DOIs for digital artifacts we will deposit in the digital repository.

Part III. Projects Developing Software

A. General Information

A.1 Describe the software you intend to create, including a summary of the major functions it will perform and the intended primary audience(s) it will serve.

Not applicable.

A.2 List other existing software that wholly or partially performs the same functions, and explain how the software you intend to create is different, and justify why those differences are significant and necessary.

Not applicable.

B. Technical Information

B.1 List the programming languages, platforms, software, or other applications you will use to create your software and explain why you chose them.

Not applicable.

B.2 Describe how the software you intend to create will extend or interoperate with relevant existing software.

Not applicable.

B.3 Describe any underlying additional software or system dependencies necessary to run the software you intend to create.

Not applicable.

B.4 Describe the processes you will use for development, documentation, and for maintaining and updating documentation for users of the software.

Not applicable.

B.5 Provide the name(s) and URL(s) for examples of any previous software your organization has created.

Not applicable.

C. Access and Use

C.1 We expect applicants seeking federal funds for software to develop and release these products under open-source licenses to maximize access and promote reuse. What ownership rights will your organization assert over the software you intend to create, and what conditions will you impose on its access and use? Identify and explain the license under which you will release source code for the software you develop (e.g., BSD, GNU, or MIT software licenses). Explain and justify any prohibitive terms or conditions of use or access and detail how you will notify potential users about relevant terms and conditions.

Not applicable.

C.2 Describe how you will make the software and source code available to the public and/or its intended users.

Not applicable.

C.3 Identify where you will deposit the source code for the software you intend to develop:

Name of publicly accessible source code repository: Not applicable.

URL: Not applicable.

Part IV: Projects Creating Datasets

A.1 Identify the type of data you plan to collect or generate, and the purpose or intended use to which you expect it to be put. Describe the method(s) you will use and the approximate dates or intervals at which you will collect or generate it.

Type of Data	Purpose/Intended Use	Intervals for Collection
Audio recordings	For transcription and qualitative analysis (coding, etc.)	October, 2018 through January, 2019
Survey responses	For qualitative analysis (coding, etc.) and quantitative analysis (descriptive and inferential statistics)	August, 2019 through October, 2019
Video recordings	For transcription and qualitative analysis (coding, etc.)	August, 2020 through November, 2020

A.2 Does the proposed data collection or research activity require approval by any internal review panel or institutional review board (IRB)? If so, has the proposed research activity been approved? If not, what is your plan for securing approval?

IRB is required but has not been approved. We will seek primarily approval at Indiana University for all three phases of research; we will seek secondary approval at institutions with whom we are working to sample student participants. The schedule for seeking IRB approval for each phase is available in the schedule of completion document.

A.3 Will you collect any personally identifiable information (PII), confidential information (e.g., trade secrets), or proprietary information? If so, detail the specific steps you will take to protect such information while you prepare the data files for public release (e.g., data anonymization, data suppression PII, or synthetic data).

Due to the sensitive nature of the raw data, only aggregate results will be shared. We will not publicly release data files that could include any PII, confidential, or proprietary information.

A.4 If you will collect additional documentation, such as consent agreements, along with the data, describe plans for preserving the documentation and ensuring that its relationship to the collected data is maintained.

Any additionally research documentation will be stored alongside the data—except in the case of identifiable keys—within the secure Box folder. We will maintain the documentation for the same amount of time as the data, which will be three years after the end of the project, as mandated by federal guidelines.

A.5 What methods will you use to collect or generate the data? Provide details about any technical requirements or dependencies that would be necessary for understanding, retrieving, displaying, or processing the dataset(s).

Type of Data	Method	Technical Requirements/ Dependecies
Audio recordings	Digital audio recording devices (e.g., handheld recorders, smartphones)	An ability to export the file from the device for analysis on a computer
Survey responses	Qualtrics web survey software	A basic understanding of the software
Video recordings	Digital video recording devices (e.g., handheld recorders)	An ability to export the file from the device for analysis on a computer
Qualitative data	Textual data	An ability to code textual data using Dedoose
Quantitative data	Numerical data	An ability to use SPSS and Excel to run statistical tests

A.6 What documentation (e.g., data documentation, codebooks) will you capture or create along with the dataset(s)? Where will the documentation be stored and in what format(s)? How will you permanently associate and manage the documentation with the dataset(s) it describes?

The PI will develop an index of the data, linking each datum to a specific folder in Box. When necessary, the PI will create a key linking participants to their pseudonym; only the PI will have access to the key. The PI will be responsible for creating codebooks in sharable formats (i.e. for use in Dedoose). The Box system will automatically archive revisions of documents.

A.7 What is your plan for archiving, managing, and disseminating data after the completion of the award-funded project?

Due to the sensitive nature of the raw data, only aggregate results will be shared in research results. Data will be retained for a minimum of three years after the end of the project, as mandated by federal guidelines. Sensitive data will be stored securely and deleted at the end of the retention period.

A.8 Identify where you will deposit the dataset(s):

Name of repository: Not applicable, raw data will not be deposited.

URL:

A.9 When and how frequently will you review this data management plan? How will the implementation be monitored?

Kristin Briney and Abigail Goben, data management experts, will actively work with the PI to maintain data and uphold data management expectations. This will be an ongoing process. If unforeseen changes occur during the project's duration, the PI will consult with IMLS to ensure compliance with all rules and requirements.