View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Caltech Authors - Main

JLSC

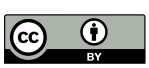
ISSN 2162-3309 | JLSC is published by the Pacific University Libraries | http://jlsc-pub.org

Volume 7, General Issue (2019)

Data Management Practices in Academic Library Learning Analytics: A Critical Review

Kristin A. Briney

Briney, Kristin A. (2019). Data Management Practices in Academic Library Learning Analytics: A Critical Review. *Journal of Librarianship and Scholarly Communication, 7*(General Issue), eP2268. https://doi.org/10.7710/2162-3309.2268



© 2018 Briney. This open access article is distributed under a Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0/)

ISSN 2162-3309 10.7710/2162-3309.2268

RESEARCH

Data Management Practices in Academic Library Learning Analytics: A Critical Review

Kristin A. Briney
Data Services Librarian, University of Wisconsin-Milwaukee

INTRODUCTION Data handling in library learning analytics plays a pivotal role in protecting patron privacy, yet the landscape of data management by librarians is poorly understood. **METHODS** This critical review examines data-handling practices from 54 learning analytics studies in academic libraries and compares them against the NISO Consensus Principles on User's Digital Privacy in Library, Publisher, and Software-Provider Systems and data management best practices. **RESULTS** A number of the published research projects demonstrate inadequate data protection practices including incomplete anonymization, prolonged data retention, collection of a broad scope of sensitive information, lack of informed consent, and sharing of patron-identified information. **DISCUSSION** As with researchers more generally, libraries should improve their data management practices. No studies aligned with the NISO Principles in all evaluated areas, but several studies provide specific exemplars of good practice. **CONCLUSION** Libraries can better protect patron privacy by improving data management practices in learning analytics research.

Received: 06/05/2018 Accepted: 01/03/2019

Correspondence: Kristin A. Briney, Golda Meir Library, 2311 East Hartford Avenue, Milwaukee, WI 53211, briney@uwm.edu

© 2019 Briney. This open access article is distributed under a Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0/)

IMPLICATIONS FOR PRACTICE

1. Data management practices by library researchers are not well understood.
2. Libraries desiring to protect patron privacy in learning analytics studies must consider the practical aspects of securing data in addition to the ethical considerations.
3. Where data handling practices are reported in the area of library learning analytics, many do not meet best practices and may risk patron privacy.
4. Libraries should improve data practices in the following areas: limiting data retention, reducing scope of data collection, following best practices for anonymization, gathering informed consent, and protecting patron-identified data. Additionally, libraries should be more transparent about data practices such as working through legal requirements, documenting data policy, securing data, and deleting data.
5. Library researchers can take advantage of peer expertise in data management to improve their data practices.

INTRODUCTION

Learning analytics is defined as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (Siemens, 2012). Within academic libraries, the use of learning analytics is becoming common in response to administrative demands to demonstrate library value and improve student experience (Cullen, 2005; Oakleaf et al., 2010; Palmer, 2012; Showers & Stone, 2014; Varnum, 2015). This creates a tension in that libraries espouse patron privacy (ALA, 2008; IFLA, 2012) yet learning analytics fundamentally require looking at individual-level data to draw conclusions (Oakleaf et al., 2010). These projects can weaken or eliminate patron privacy if personal data is not handled properly, over and above the broader ethical considerations around privacy in library learning analytics (K. M. L. Jones & Salo, 2018).

Libraries have many patron data protection standards to ground their practice and these research projects. These include ethical codes from the American Library Association (ALA) and International Federation of Library Associations (IFLA) (ALA, 2008; IFLA, 2012); research data management best practices (Briney, 2015; Corti, Van den Eynden, Bishop, & Woollard, 2014); privacy principles such as those from NISO (NISO, 2015); broader learning analytics codes of practice such as those from Jisc (Jisc, 2015); and local and federal privacy laws. Given the growth in the area of data management within

libraries (Briney, Goben, & Zilinski, 2015; Tenopir, Birch, & Allard, 2012), there is also an opportunity to apply to library-captured data the data management best practices librarians teach to researchers. Despite these resources, practices such as insufficient anonymization, poor data security, and inadequate data governance increase the risk of personal data exposure even for studies that have been approved by an Internal Review Board (IRB) and which strive to follow privacy laws like the Family Educational Rights and Privacy Act (FERPA) (1974).

The everyday handling of patron data greatly impacts privacy, yet little is known about how data is managed in library learning analytics projects. Poor data practices by those conducting analytics work increase the risk of data exposure and loss of privacy, making the evaluation of data handling crucial to assessing privacy in this sphere. The central question is, therefore, To what extent do data-handling practices in the area of library learning analytics match established data guidelines and best practices?

This critical review evaluates data practices as described in published academic library learning analytics projects and compares them to current library data privacy principles and research data management best practices. The intent is to identify gaps in current library learning analytics data practices so that libraries can improve their data management and perform analytics research that better protects patron privacy.

LITERATURE REVIEW

This review sits at the intersection of research data management, academic libraries, and learning analytics, yet few resources exist at the convergence of all three topics. Asher (2017) describes some data management best practices for library learning analytics in a chapter in the book *Protecting Patron Privacy*, couching risk, consent, justice, and data best practices for library analytics in the framework of research ethics. Yoose (2017a) provides a case study on balancing data security with the assessment needs of the Seattle Public Library, describing the many protection layers put into place for the library's data warehouse. Yoose further examines one of these data protection practices, deidentification, in a presentation describing deidentifying patron data for longitudinal library assessment projects (Yoose & Halsey, 2016) and in a post for Choose Privacy Week describing the two types of personally identifiable information (PII) and explaining how deidentification is not foolproof (Yoose, 2017b). Nicholson and Smith (2007) take a different approach to deidentification by proposing a framework for library data based on the stipulations from the U.S. Health Insurance Portability and Accountability Act (HIPAA). Similar work has been done on deidentifying learning analytics data more broadly, such as by Khalil and Ebner (2016), who review deidentification strategies and

limitations for learning analytics data. The larger field of learning analytics also has a number of resources that—while not specifically labeled *data management*—cover data handling best practices. These include Jisc's Code of Practice for Learning Analytics (Jisc, 2015) and U.S. Department of Education guidelines on securing student data (U.S. Department of Education, 2017).

To date, no studies have examined the data management practices of librarians. However, a case study exists in which Goben and Raszewski (2015) describe applying data management best practices to library data using the data life cycle as a framework. Comparatively, the data management practices of researchers are well studied, particularly by librarians (Goben & Griffin, in press; Perrier et al., 2017). Studies most often focused on topics not addressed in this review (such as data format, data size, and storage and backup) (Goben & Griffin, in press), yet some information is known about privacy and security, which are more relevant to this review. Carlson and Stowell-Bracke (2013) found that "interviewed students were generally not proactive or very much engaged in taking action to ensure that good management or security practices were being followed" (p. 355). Johnston & Jeffryes (2014) similarly found that "students did not consider data security an issue and felt that they had adequate protections in place" (p. 4) and that the students need to develop skills to "understand privacy issues associated with data" (p. 7). In a study of 416 researchers across many disciplines at Virginia Tech, 15% of respondents struggled with "dealing with sensitive data, data transmission, and encryption" (Shen, 2016, p. 510). An excellent synopsis of the current state of data management is provided by Toups and Hughes (2013), who said that researchers "believe they are managing data well, but their research is not without management issues" (p. 227).

METHODS

This research, categorized as a critical review (Grant & Booth, 2009), gathered studies from two sources. First, citations were harvested from two recent library learning analytics review articles (K. M. L. Jones & Salo, 2018; Kogut, 2016), with several further citations gathered from within those citations. Sixty-nine studies were identified using this method. A second batch of studies was harvested using a methodology based on that described in the review by Kogut (2016). Similar to that review, this author searched two databases, *Education Resources Information Center* (EBSCO) and *Library, Information Science, & Technology Abstracts with Full Text* (EBSCO), using the subject terms *academic libraries* and *academic achievement*; results were limited to journal articles in the English language. The search, completed in April 2018, yielded 377 articles.

In order to reduce methodological heterogeneity, only quantitative learning analytics projects in academic libraries (i.e., studies that correlate library use with student outcomes) that used patron-identified data during data collection or analysis were included. Studies were included whether they retained identifying data or aggregated/deidentified later. Learning analytics projects included in this review used data from bulk or automated harvesting of existing library, and potentially university, digital data sets or data generated by joining multiple digital data sets using a common patron identifier. Studies about analytics projects across multiple universities were included so long as they used campus-specific patron-identified information for analysis. Studies originated from several countries (including Australia, Chile, Hong Kong, Jordan, South Africa, Turkey, United Kingdom, and the United States) but were limited to the English language. Research conducted using only aggregate data (e.g., data summarizing all patrons) or studying nonacademic libraries was removed from consideration. There was no date limit. This selection method narrowed the group of studies for analysis down to 54, which are listed in Table 1.

The selected group of 54 studies were reviewed for their data handling methods, specifically seeking to answer questions related to

- How the project accounted for legal requirements
- Whether researchers planned for data handling issues and/or created a data management plan
- How long data was retained and in what format
- What security practices were being used
- What method of anonymization was being used
- What the scope of the data collected was
- Whether informed consent was collected
- Who was allowed access to the data

This methodology is limited, as it is usually not possible to glean comprehensive data handling methods from published study results. This is partly because of the format of a published article—as articles do not universally describe data-handling practices—but it is also symptomatic of a larger problem of reproducibility within scholarly publishing (Baker, 2016; Freedman, Cockburn, Simcoe, Parkes, & Gelber, 2015; Iqbal et al., 2016; Open Science Collaboration, 2015). This research cannot evaluate what is not reported, so a review of data handling practices can provide only a snapshot, not a complete picture. However, enough information about data handling practices is present across the 54 included articles to serve as a foundation for evaluation.

The critique of data handling methods was structured against the "NISO Consensus Principles on User's Digital Privacy in Library, Publisher, and Software-Provider Systems" (NISO, 2015) because it is specific to privacy and responsible data practices within the context of library ethics. Data handling methods were grouped under one of the NISO principles for evaluation. Eight of the twelve principles were used in this review:

- Shared Privacy Responsibility
- Transparency and Facilitating Privacy Awareness
- Security
- Data Collection and Use
- Anonymization
- Options and Informed Consent
- Sharing Data with Others
- Access to One's Own Data

Four other NISO principles are not evaluated, as they are not data practices made explicit in publications:

- Notification of Privacy Policies and Practices
- Supporting Anonymous Use
- Continuous Improvement
- Accountability

Table 1 lists the sections of this review in which each of the 54 evaluated studies are referenced. This provides an index of the studies where data practices are described or can be implied relevant to the eight evaluated NISO categories. Studies not indexed in a particular section indicate that no information about the relevant data practice is available in the published study. Quality evaluations of the data practices are not included in Table 1 and are instead enumerated in detail in the Results section of the article.

	Study Evaluated	Shared Privacy Responsibility	Transparency and Facilitating Privacy Awareness	Security	Data Collection and Use	Anonymization	Options and Informed Consent	Sharing Data with Others	Access to One's Own Data	*TOTAL*
1	Allison, 2015		X	X		X				*3*
2	Black & Murphy, 2017	X		X				X		*3*
3	Bowles-Terry, 2012	X		X			X			*3*
4	Çetin & Howard, 2016			X	X	X				*3*
5	Cherry et al., 2013	X	X			X	X			*4*
6	Cook, 2014		X		X					*2*
7	Collins & Stone, 2014			X		X	X			*3*
8	Coulter et al., 2007			X						*1*
9	Cox & Jantti, 2012a		X	X	X	X	X			*5*
10	Cox & Jantti, 2012b		X	X	X					*3*
11	de Jager, 2018			X	X		X	X		*4*
12	Fransen & Peterson, 2016		X	X				X		*3*
13	Gariepy et al., 2017									*0*
14	Goodall & Pattern, 2011			X		X	X		X	*4*
15	Haddow, 2013			X	X					*2*
16	Haddow & Joseph, 2010	X		X	X					*3*
17	Jantti, 2016		X	X				X		*3*
18	Jantti & Cox, 2013		X	X		X				*3*
19	Jones, 2010		X	X			X			*3*
20	Kome, 2017		X		X		X			*3*
21	Kot & Jones, 2015		X	X	X	X		X	X	*6*
22	LeMaistre, 2015		X	X				X		*3*
23	LeMaistre et al., 2018			X		X	X			*3*
24	Massengale et al. 2016			X	X		X		X	*4*
25	McCarthy, 2017			X		X				*2*
26	Montenegro et al., 2016			X	X	X				*3*
27	Murray et al., 2016			X	X	X	X			*4*
28	Nackerud et al., 2012			X						*1*
29	Nackerud et al., 2013		X	X	X	X	X			*5*
30	Nackerud et al., 2015							X		*1*

	Study Evaluated	Shared Privacy Responsibility	Transparency and Facilitating Privacy Awareness	Security	Data Collection and Use	Anonymization	Options and Informed Consent	Sharing Data with Others	Access to One's Own Data	*TOTAL*
31	O'Kelly, 2016			X	X					*2*
32	Odeh, 2012			X						*1*
33	Pepper & Jantti, 2015		X	X	X	X		X		*5*
34	Renaud et al., 2015		X	X		X	X			*4*
35	Samson, 2014	X			X	X	X			*4*
36	Scarletto et al., 2013	X		X		X	X			*4*
37	Scott, 2014			X		X				*2*
38	Showers & Stone, 2014			X		X				*2*
39	Soria et al., 2013		X	X	X		X			*4*
40	Soria et al., 2014		X	X	X	X	X			*5*
41	Soria et al., 2015		X	X	X		X			*4*
42	Soria et al., 2017a		X	X	X					*3*
43	Soria et al., 2017b		X	X		X	X			*4*
44	Squibb & Mikkelsen, 2016			X		X		X		*3*
45	Stemmer & Mahan, 2015		X	X			X			*3*
46	Stemmer & Mahan, 2016		X	X			X			*3*
47	Stone et al., 2012			X		X				*2*
48	Stone, Pattern, et al., 2011	X	X	X	X	X	X			*6*
49	Stone & Ramsden, 2012	X	X	X		X	X	X	X	*7*
50	Stone, Ramsden, et al., 2011	X		X				X		*3*
51	Thorpe et al., 2016	X		X		X	X			*4*
52	White & Stone, 2010	X		X		X				*3*
53	Wong & Cmor, 2011			X						*1*
54	Wong & Webb, 2011	X	X	X		X				*4*
	TOTAL	*12*	*25*	*48*	*21*	*28*	*24*	*11*	*4*	

Table 1. Index of studies included under each of the eight NISO Principles evaluations. Indexing in this table is not an indicator of data handling quality (see the Results section for this information). Similarly, if studies are not included under a particular NISO Principle evaluation, this only denotes that the specified data practice was not discussed in the published study and therefore could not be evaluated.

The NISO Consensus Principles document was chosen as the evaluative framework over a data management–specific document due to the lack of peer-reviewed or consensus research on data management best practices. Prescriptive data management best practices are more often found in books (Borgman, 2016; Briney, 2015; Corti et al., 2014) and on library websites (Yoon & Schultz, 2017) than in peer-reviewed articles, meaning that there is no broadly agreed-upon framework against which to evaluate data handling practices. The NISO Consensus Principles are a good proxy for standardized data management principles in library learning analytics because they provide focused recommendations for data best practices within libraries. In this respect, the principles are also a superior evaluative framework for this research to learning analytics specific documents, such as Jisc's Code of Practice for Learning Analytics (Jisc, 2015), as they are more specifically structured around data best practices and account for libraries' additional ethical considerations around patron privacy. Using the NISO Principles to evaluate data-handling practices in library learning analytics does have some limitations, however. First, the NISO Consensus Principles focus primarily on electronic systems that hold library data, particularly those created by external content and software providers. Second, some of the studies being evaluated in this review are older than the NISO Consensus Principles, meaning that the authors could not have consulted the document in designing their research. Third, NISO is a national organization based in the United States, while this research includes studies from around the world. The Principles are broad enough, however, to apply to many countries and country-specific differences, such as for retention requirements in the United Kingdom, which are noted in the text. Finally, the NISO Consensus Principles are not comprehensive in considering all data management practices, such as quality control, that apply to learning analytics. To supplement the NISO Consensus Principles' limitations, this review's evaluations are therefore augmented with ethical codes from the ALA and IFLA (ALA, 2008; IFLA, 2012); Jisc's Code of Practice for Learning Analytics (Jisc, 2015), which applies to learning analytics broadly and is situated within the United Kingdom's legal framework; and research data management best practices (Briney, 2015; Corti et al., 2014). These documents are representative of best practices within the domains of library ethics, learning analytics, and research data management.

CRITICAL ANALYSIS

Shared Privacy Responsibility

The first NISO principle states that libraries operate under ethical requirements, such as from the ALA Code of Ethics and the IFLA Professional Code of Ethics (ALA, 2008; IFLA, 2012), and additionally may often come under legal requirements that dictate how

to handle patron data. This section evaluates mentions of legal considerations and Internal Review Board (IRB) within the text of published library learning analytics studies. References to consent, while under the purview of IRB, are addressed in the "Options and Informed Consent" section. Mentions of ethical considerations within the published studies are out of scope for this review.

Six studies mention legal considerations (Black & Murphy, 2017; Stone, Pattern, & Ramsden, 2011; Stone & Ramsden, 2012; Stone, Ramsden, & Pattern, 2011; White & Stone, 2010; Wong & Webb, 2011). Black and Murphy (2017) state that "students who elected to withhold directory information via the Family Educational Rights and Privacy Act (FERPA) and students who were under 18 years of age when first enrolled were excluded from [data collection]" (p. 410). The five other studies describe collaborating with institutional leadership or legal departments in planning their projects, beyond simply gaining permission from the institutional research office to access university data. A study by the Library Impact Data Project (LIDP) specifically stated that "a major issue identified at the very beginning of the project was the need to abide by legal regulations and restrictions, such as data protection" (p. 9). To work within this constraint, "the team liaised with Jisc Legal at the outset of the project and subsequent further discussion with the University of Huddersfield Legal and Data Protection Officers have helped to ensure that there is complete anonymization" (Stone, Pattern, et al., 2011, p. 9). More conversations about legal requirements may be happening locally, but are not being reported via formal publication.

The nature of the evaluated research mandates gaining IRB approval, yet only six library learning analytics studies specifically reference the IRB process (or their country's equivalent) in the published text (Bowles-Terry, 2012; Cherry, Rollins, & Evans, 2013; Haddow & Joseph, 2010; Samson, 2014; Scarletto, Burhanna, & Richardson, 2013; Thorpe, Lukes, Bever, & He, 2016). Haddow and Joseph (2010) is one of these; they describe gaining approval from their University's Human Research Ethics Committee on the condition of ensuring that "individual students were not identified or identifiable and the secure storage of data" (p. 236). In the case of Cherry et al. (2013), the authors state that "in order to ensure students' privacy, the authors applied for and received Institutional Review Board approval to perform this study" (p. 389). It is worth noting that the IRB does not necessarily use the same framework as libraries for evaluating harm to study participants, nor do they consider library principles in their approvals. Therefore, libraries should use IRB approval in combination with recommendations from documents such as the NISO Consensus Principles in order to fully protect patron privacy under library ethical principles.

Transparency and Facilitating Privacy Awareness

While this NISO principle centers on communication rather than explicit data handling, the principle recommends documenting data-handling policies and practices, including "what data are collected, why data is collected, who has access to the data, how the data are stored and secured, when that data might be disclosed and to whom, and what the organization's data retention and/or deletion policies are" (NISO, 2015, "2. Transparency and Facilitating Privacy Awareness Library," para. 1). These decisions correspond to the information that is traditionally covered in a research data management plan (DMP) (Thoegersen, 2015). Documentation on data handling decisions, either as a DMP or through policy, is the vehicle for promoting the transparency advocated for in the NISO principle. Because of the alignment with a DMP, this section of the review examines descriptions of activities like data management planning. It also reviews a major data practice not fully covered by other NISO principles—data retention.

Only Stone et al. explicitly talk about planning, stating that "due to the short timescale of the project, potential issues with data were anticipated at the proposal stage" (Stone & Ramsden, 2012, p. 551; Stone, Pattern, et al., 2011). This mirrors the preferred recommendation for the timing of data management planning, which is early in the project, as data and privacy decisions will affect how a project is designed and carried out. The limited published information on planning is not surprising, as DMPs are frequently developed early in the research process, such as for grant applications, rather than for published articles.

Data retention is not addressed adequately within the NISO Consensus Principles. Retention is covered briefly under the "Transparency" principle (stating that libraries should be transparent with retention and deletion policies) and under the NISO "Anonymization" principle (in the context of retaining anonymized data). The latter principle acknowledges that "anonymization may not completely eliminate the risk of re-identification" (NISO, 2015, "5. Anonymization," para. 1), so it's best to revert to the data management best practice to minimize retention of sensitive data when anonymization is either not done or not robust (Briney, 2015; Corti et al., 2014). With respect to limited data retention, recommendations in the Jisc "Code of Practice for Learning Analytics" are stronger, calling for minimized retention "only for appropriate and clearly defined periods" (Jisc, 2015, p. 4). This is likely due to limitations on retention imposed under the UK's Data Protection Act, which states that personal data should be kept for no longer than is absolutely necessary (Data Protection Act 1998, 1998).

Two trends emerge when examining data retention in library learning analytics projects: projects retaining data for analytics that might be otherwise regularly discarded, and projects that are retaining large analytics data sets for an unspecified or open amount of time.

In the first category, there are nine studies in which, for the purposes of analytics, the library appears to be retaining transactional data that should otherwise be deleted (Allison, 2015; Cook, 2014; B. Cox & Jantti, 2012a, 2012b; Kome, 2017; Renaud, Britton, Wang, & Ogihara, 2015; Stemmer & Mahan, 2015, 2016; Wong & Webb, 2011). Renaud et al. (2015) harvested transactional data every week for an academic year as a workaround for the fact that the library's integrated library system (ILS) did not maintain historical circulation data. Similarly, Cox and Jantti (2012a, 2012b) worked around the fact that their ILS only exports the total number of items borrowed to date for each unique student identifier. They exported data every two weeks and calculated the change from the previous data set in order to measure circulation over time. Other libraries kept data from several years to draw comparisons across cohorts or as students progressed through their degrees; such analyses have been done over two years (Allison, 2015), three years (Wong & Webb, 2011), four years (Stemmer & Mahan, 2015), six years (Stemmer & Mahan, 2016), and twelve years (Cook, 2014). Nominally, transactional library records would be used and then discarded under a relevant retention schedule, yet analytics projects can conflict with previously established retention practices. One published example of this is Stone, Pattern, et al. (2011); in amassing a multi-institution analytics platform, they found that two-thirds of identifying information for their target data was deleted under institutional policy in 2010. This caused the team to "put processes in place in order to be able to capture the data from 2011 onwards" (p. 11).

The second category, projects that retain data sets for an unspecified or open amount of time, differs from the first in the scope and purpose of data collection. Here, many transactional data sets are combined with student demographic and outcome information from the institution into one large analytics database to build longitudinal projects. These projects give no indication of a finite retention period, with databases continually growing and resulting in an expanding series of published analyses. This category thus represents sustained analytics programs. Two libraries fall into this category, including the University of Minnesota with seven studies (Fransen & Peterson, 2016; Nackerud, Fransen, Peterson, & Mastel, 2013; Soria, Fransen, & Nackerud, 2014, 2017a, 2017b; Soria, Fransen, Nackerud, & Kross, 2013; Soria, Nackerud, & Peterson, 2015), and the University of Wollongong with six studies (Cox & Jantti, 2012a, 2012b; Jantti, 2016; Jantti & Cox, 2010, 2013; Pepper & Jantti, 2015). These institutions have built and published on large library analytics data sets over the course of several years and offer no indication

of a defined data retention period. The University of Wollongong started collecting student data into a learning analytics database in 2010 with the express purpose of building longitudinal time-series data on student resource use (Cox & Jantti, 2012a). Four studies at other institutions also expressed a desire to maintain analytics data for an extended period for further analysis (Cherry et al., 2013; J. L. Jones, 2010; Kot & Jones, 2015; LeMaistre, 2015). Troublingly, Jones stated that "the most exciting aspect of the project is that there is no endpoint; visitors continue to swipe in every day, and the data are collected continuously" (J. L. Jones, 2010).

Given U.S. public libraries' history with PATRIOT Act requests (Goodman & Goodman, 2008; Peterson, 2014), academic libraries with open-ended data retention strategies raise significant concerns about unintended reuse. The longer libraries hold on to sensitive information, the greater the risk that it can be accessed via legal (through venues such as warrants and open records requests) or illegal methods (such as hacking or disclosure by a disgruntled employee). The easiest way to reduce risk to patrons is to discard their data under the pertinent records retention schedule or once the relevant transaction is complete.

Only two studies explicitly discuss deleting data at the end of a project (Stone, Pattern, et al., 2011; Stone & Ramsden, 2012). The studies mention data retention and deletion practices as problematic to obtaining data for a multiuniversity project. There is no other published evidence that data destruction is occurring, suggesting a lack of transparency on the part of academic libraries.

Security

Very few library learning analytics studies provide information on data security, defined by the NISO Consensus Principles as including

> encryption of personal data while they are at-rest and in-motion; prompt updates of systems and software to address vulnerabilities; systems, procedures, and policies for access control of sensitive data; a procedure for security training for those with access to data; and documented procedures for breach reporting, incident response, and system, software, and network security configuration and auditing. (NISO, 2015, "3. Security," para. 1)

Of the studies that cover this information, four have campus units that routinely deal with sensitive information running their assessment infrastructure (de Jager, Nassimbeni, Daniels, & D'Angelo, 2018; Jantti & Cox, 2013; Renaud et al., 2015; Thorpe et al.,

2016). In the United Kingdom, the LAMP project is following a consortial model that leverages central security expertise (Showers & Stone, 2014). LeMaistre, Shi, and Thanki (2018) handled security differently, using several layers including encrypting raw data, limiting accessed to data even after anonymization, and reporting results only for groups larger than 10 students. Grand Valley State University demonstrated a strong model in which the library never held any patron-identified information during their analysis of the impact of information literacy instruction. Rather, course numbers were sent to GVSU's institutional research department, which performed student-level calculations in their secure environment and sent back aggregate statistics. This means that the library never needed to maintain a secure environment for analytics data (O'Kelly, 2016).

Few studies provide security information, but many raise security concerns. The chief concern is that data has the potential to be exposed as it crosses between systems unless it is encrypted in transit. Several studies pull data from multiple library silos for analysis, such as Soria et al. (2013, 2014, 2017a) pulling assessment data from 10 different library service points and Murray, Ireland, and Hackathorn (2016) pulling data from eight different library sources. Even moving analytics data from a central secure environment and into Microsoft Excel (Haddow & Joseph, 2010; Pepper & Jantti, 2015) can raise security issues, as an individual password-protected computer used for a wide variety of tasks cannot have the same level of protection as a single-purpose secure server. A significant number of studies, 46 of the 54 examined, imply transferring individual-level data between the library and a central university office (Allison, 2015; Black & Murphy, 2017; Bowles-Terry, 2012; Çetin & Howard, 2016; Collins & Stone, 2014; Coulter, Clarke, & Scamman, 2007; Cox & Jantti, 2012a, 2012b; de Jager et al., 2018; Fransen & Peterson, 2016; Goodall & Pattern, 2011; Haddow, 2013; Haddow & Joseph, 2010; Jantti, 2016; Jantti & Cox, 2013; J. L. Jones, 2010; Kot & Jones, 2015; LeMaistre, 2015; LeMaistre et al., 2018; Massengale, Piotrowski, & Savage, 2016; McCarthy, 2017; Montenegro et al., 2016; Murray et al., 2016; Nackerud et al., 2012, 2013; Odeh, 2012; Pepper & Jantti, 2015; Renaud et al., 2015; Scarletto et al., 2013; Scott, 2014; Soria et al., 2014, 2017a, 2017b, 2013, 2015; Squibb & Mikkelsen, 2016; Stemmer & Mahan, 2015, 2016; Stone, Pattern, & Ramsden, 2012; Stone, Pattern, et al., 2011; Stone & Ramsden, 2012; Stone, Ramsden, et al.. 2011; Thorpe et al., 2016; White & Stone, 2010; Wong & Cmor, 2011; Wong & Webb, 2011). Of these, only LeMaistre et al. (2018) explicitly states that the transfer is being done "securely." This information often contains very sensitive information like patron IDs, demographics, and grades, which mandate extra protections under statutes such as FERPA (Family Educational Rights and Privacy Act, 1974). This data must be protected at all times—not only when it is being stored but also when it is being transferred—as moving data comes with unique vectors of security risk such as insecure email, WiFi sniffing, etc. Wong mitigates some of the transfer risk by preventing student name, ID

number, and GPA from being sent together in the same file: the library sends names and IDs to the Academic Registry, which replaces these two data points with GPA, shuffles rows in the spreadsheet, and returns the data (Wong & Cmor, 2011; Wong & Webb, 2011). Data transfers represent a potential hole in security coverage for sensitive information, yet are present in a large number of library learning analytics projects. Security practices, such as encryption, need to extend across these transfers to better protect patron privacy.

Security practices must also be extended to vendors providing assessment services or data to libraries, such as those recommended in ACRL's "Value of Academic Libraries" report (Oakleaf et al., 2010). While many libraries use vendor tools for data collection, none of the reviewed studies appear to be using a library vendor to conduct their analysis. Vendors are starting to provide service in the area of assessment (Cullen, 2005; Enis, 2014; OrangeBoy, Inc., 2017) but also have a troubling history with maintaining the privacy of patron data (Hellman, 2014a, 2014b, 2016; Lambert, Parker, & Bashir, 2015; Magi, 2010; Reidsma, 2016). Even where vendors do not provide assessment specifically, they hold a significant amount of data that can impact assessment projects. For example, one institution in the LIDP "ran into problems . . . when they found out that although their gate entry system did keep historical data it was stored by the system supplier and was therefore not readily available" (Stone, Pattern, et al., 2011, p. 11). The LIDP also expressed a need to establish a "liaison with publishers about linking the results of the project with their usage data" (Stone, Ramsden, et al., 2011, "Evaluation and Exit Strategy (Month 6)," para. 3). Vendor contracts should therefore be negotiated with security in mind, particularly as to what vendors are allowed to do with data residing in their systems.

Beyond storage, transfer, and vendor security, a number of security practices can and should be put into place to protect library learning analytics data. Yoose, in her public library analytics case study, describes the layers of security the Seattle Public Library uses during an assessment project, including obscuring the most identifying information in the library's data set; storing limited sensitive data points together; strictly controlling access to the raw data; and other practices (Yoose, 2017a).

Security layers should be combined with regular security reviews, as "security is a process, not a static condition" (Charlton, 2017, para. 7). Additionally, library staff also need training in order to be able to implement up-to-date security practices. Libraries need to engage in a culture of security around learning analytics projects, providing local support, and training and being proactive about putting a system of measures in place to protect patron data.

Data Collection and Use

This NISO principle speaks to the practice of "data minimization," which means minimizing the amount of sensitive data by only collecting what is absolutely necessary. Specifically, it states that "the potential benefit to the user, the library, content-, or software-provider derived from the collection and use of users' personal data must be balanced against the impact of that collection and use on users and their right to privacy" (NISO, 2015, "4. Data Collection and Use," para. 1). Data minimization is done to reduce the security burden of keeping lots of information safe and to minimize risk to patrons (Schneier, 2016).

Library analytics projects span a range of practices with respect to data minimization. Eight projects encompass a particularly wide range of data collection (Kot & Jones, 2015; Massengale, Piotrowski, & Savage, 2016; Murray et al., 2016; Nackerud et al., 2013; Soria et al., 2014, 2017a, 2013, 2015). One expansive example is Massengale et al. (2016), in which library data was collected on "students physically entering the building, students receiving instruction from a librarian, students visiting the Research Help Office (RHO), students checking out laptops, students using interlibrary loan, students reserving study rooms, students placing requests for 3-D printer usage, and students accessing our online resources" (p. 230). Conversely, Kot and Jones analyze use of only three library service points (workstations, study rooms, and research clinics) against a wide range of university data including: students' demographic characteristics such as "student's sex, race/ethnicity, citizenship, age at matriculation, and the matriculation term;" academic preparation including "the student's high school GPA, SAT math score, and SAT verbal score, as well as an indicator of whether the student transferred any Advanced Placement (AP) credits;" and other variables such as "the student's college or school, the number of credits taken in the first term, whether the student lived on campus, whether the student participated in a Freshman Learning Community (FLC), and the student's level of unmet financial need" (Kot & Jones, 2015, p. 571). In other cases, such as that warned about by Kome, the library may be collecting more data than necessary or anticipated through things like WiFi logs and proxy servers, which allow the library to track patron locations and resource usage at the individual level. Even information such as HTTP headers can contain sensitive information (Kome, 2017). On the low end of the scope spectrum is O'Kelly (2016), whose project was designed to analyze student-identified data even though the library never held any data at this granularity. Similarly, Cook (2014) received only GPAs, graduation rates, and test scores averaged across annual cohorts of students. Most projects are in between these two extremes, though many studies described a wish to have more specific data or overall more data for analysis (Çetin & Howard, 2016; B. Cox & Jantti, 2012a; Haddow, 2013; Murray et al., 2016; Pepper & Jantti, 2015; Samson, 2014; Stone, Pattern, et al., 2011).

Libraries should carefully consider the scope of data collected for analytics projects, collecting only what is absolutely necessary in order to reduce security requirements and lower the cost of a potential data breach. Two 2017 breaches at public libraries resulted in the loss of sensitive patron information, including driver's license numbers and birthdates (Alameda County Library, 2017; Week.com, 2017); without these data points, the breaches would have been less damaging. Data security experts are starting to shift focus from the task of total breach prevention toward minimizing damage when an inevitable breach happens (Morgan, 2017; Schneier, 2016). This is because it is impossible to totally prevent data breaches—institutions with better security than libraries still get breached (Department for Business Innovation, 2013)—and breaches are expensive. Ponemon Institute's "2017 Cost of Data Breach Study" estimated that the average total cost of a data breach is $3.62 million. The study found that "the average cost for each lost or stolen record containing sensitive and confidential information" was $141 in 2017, with educational records costing an average of $200 each (Ponemon Institute LLC, 2017, p. 1). Many universities have already experienced data breaches (Garg, 2016; O'Neil, 2014) and universities continue to be a target, as shown in a Gemalto report finding that education breaches doubled in the first half of 2017 (Gemalto, 2017). Academic libraries will not escape costs (and nonpecuniary damages) from an inevitable data breach and should therefore focus on minimizing the collection of sensitive data.

The other portion of this NISO principle recognizes that some types of personal data are particularly sensitive, "e.g., regarding race, gender, socioeconomic class, ability, etc." (NISO, 2015, "4. Data Collection and Use," para. 2). This data is sensitive even if it is not a direct identifier, and it is not usually thought of by libraries as PII. Such data requires a higher level of scrutiny and justification to use (Jisc, 2015), requires extra protection, and is often legally protected under statutes such as FERPA. In a review of literature on libraries' contributions to academic success, Kogut (2016) found that GPA, a sensitive data point, was evaluated in over 40% of studies. This reflects trends in broader learning analytics practices where the analysis of GPA is prevalent (York, Gibson, & Rankin, 2015). Three other studies have been conducted specifically looking at sensitive socioeconomic status compared to library use (Haddow, 2013; Haddow & Joseph, 2010; Soria et al., 2015) and another pair used a related proxy for the socioeconomic status, such as funding aid (de Jager et al., 2018; Montenegro et al., 2016). In addition to libraries holding individual sensitive data points, there is a significant issue that by combining multiple data sets for analytics, anonymity decreases and the sheer amount of sensitive information available about an individual increases (Ohm, 2009). Several library learning analytics publications miss this privacy concern. Oakleaf, Brown, Nackerud, Jantti, and Abel (2017) describe the main data difficulties in library learning analytics as not having enough specific enough data or having siloed data that is difficult to combine, but don't

recognize the privacy and security concerns raised by amassing such data. Similarly, Cox and Jantti (2012b) argue that "the personally identifiable data that users can glean from the [learning analytics system] is significantly less than that which can already be ethically and legally obtained through the library management system (LMS), logs, and access to student management systems" ("Meeting the Challenges," para. 4). However, this cannot be the case, as learning analytics systems combine a range of sensitive data points such as demographics, library use, and GPA for an individual student all in one place. This makes it easy for someone, including a bad actor, to access all of this sensitive information at once, whereas that person might otherwise not have access to each of the individual siloes from which the data originated. All of this evidence points to the need for further discussion on how library learning analytics data has the potential to do harm (K. M. L. Jones & Salo, 2018; Rubel & Jones, 2016).

Anonymization

This NISO principle calls for removing personally identifying information (PII) from data sets, a process called "anonymization" or "deidentification," when such information is no longer needed for analysis purposes and when retaining analytics data after a project ends. Anonymization is appealing for library learning analytics projects because it suggests a method to address the inherent tension between needing individual-level data to run analytics and library ethics that mandate individual privacy. However, libraries should be aware that the NISO Consensus Principles recommend anonymization as one of many information privacy controls and that anonymization "may not completely eliminate the risk of re-identification" (NISO, 2015, "5. Anonymization," para. 1). In a different approach, the Jisc Code of Practice for Learning Analytics focuses more on minimizing retention of learning analytics data, partly due to legal requirements of the United Kingdom's Data Protection Act (Jisc, 2015).

Unfortunately, published library learning analytics projects are rife with examples of inadequate anonymization practices that demonstrate a flawed understanding of how to properly anonymize data (Garfinkel, 2015). One central issue is that demographic information, a key type of information leveraged for many learning analytics projects, can still be identifying even in the absence of a name or student ID number. A classic example of this problem comes from security researcher LaTanya Sweeney, who proved that 87% of Americans are uniquely identifiable by the combination of their birthdate, zip code, and gender (Sweeney, 2000); while none of these data points singly exposes the individual, they are powerful in combination. This research found that libraries routinely remove "direct identifiers"—such as user name, ID number, or e-mail address—but often fail to account for the fact that "indirect identifiers"—such as gender, ethnicity, year in

school, major, veteran status, etc.—can be combined within a data set or to external data to identify patrons. Data sets may contain demographic combinations such as "African American female physics major" that are identifying. Alternatively, data can be combined with external social media information to reidentify individuals, such as when a study of Facebook profiles for students at an unnamed university was reidentified due to the unique majors offered at that institution (Parry, 2011).

Seven studies remove only direct identifiers in their anonymization process, leaving behind data sets that likely still violate patron privacy (Allison, 2015; McCarthy, 2017; Montenegro et al., 2016; Nackerud et al., 2013; Renaud et al., 2015; Scarletto et al., 2013; Thorpe et al., 2016). Of these, one library described anonymizing data that contained patron library usage indicators and demographics (such as major, gender, and ethnicity) by saying that "Internet IDs were removed from the data, creating an anonymized set" (Nackerud et al., 2013, p. 138); this, by definition, is not an anonymized set. Similarly, Thorpe et al. (2016) "removed the identifying data by replacing usernames with randomly generated numeric IDs," (p. 379) yet still retained the potentially identifying attributes age, sex, ethnicity, major, and year in. Scarletto et al. (2013) received data "without identifying information" (p. 373) from the university, even though the data contains "department, major, grade point average, class standing, international status, home campus, ethnicity, and gender" (p. 373). An additional five library projects say that they anonymize data but offer no information on how they actually do so (Cherry et al., 2013; Collins & Stone, 2014; Goodall & Pattern, 2011; Samson, 2014; Squibb & Mikkelsen, 2016).

Six studies describe more complex strategies for suppressing identifiable information (LeMaistre et al., 2018; Stone et al., 2012; Stone, Pattern, et al., 2011; Stone & Ramsden, 2012; White & Stone, 2010; Wong & Webb, 2011). LeMaistre et al. (2018) state that their university's Office of Institutional Research anonymized the data yet only describe the process of swapping the user ID with an alternate surrogate ID. However, this office also recognized the potential for reidentification and consequently practiced aggregation, such as by limited reported results to groups of 10 or larger. Wong and Webb (2011) excluded majors that contained fewer than 30 people in their analysis because these groups could not produce statistically significant results. The other four studies were from Stone et al., which excluded groups including: courses smaller than 35 students; degrees with fewer than five students; and "distance learners, post graduates, part-time students, sandwich courses, short courses and courses with low numbers where anonymity could not be guaranteed" (Stone & Ramsden, 2012, p. 550; Stone, Pattern, et al., 2011; White & Stone, 2010). It should be noted that even Stone falls into the trap of claiming that removing direct identifiers equates with anonymization, stating that "once the data have been combined this identifier is removed, thus ensuring anonymity" (Stone, Pattern,

et al., 2011, p. 9). The data also underwent a second deidentification process for public release—which included obscuring the names of schools and departments and generalizing some course information (Stone et al., 2012; Stone & Ramsden, 2012)—meaning that there was still a chance for reidentification after the initial suppression of the various subpopulations.

Improper anonymization can harm already marginalized populations. Simply removing names and ID does not protect small subpopulations of users that are easier to reidentify within a data set. Three studies have been published showing small ethnic populations of under 20 students, including American Indian or Native American (n=8) and Hawaiian or Pacific Islander (n=4) students (Soria et al., 2014); and Hawaiian (n=18) students (Soria et al., 2017b); and Native American (n = 3) and Pacific Islander (n = 2) students (citation omitted in order to protect students' identities). Alarmingly, the latter study further winnows demographic information down to examples of (n=1), drastically increasing the possibility of reidentification. A similar study by Samson (2014) stated that the study contained a small number of students who had a documented disability (n=20), another sensitive data point. Combining ethnicity or disability information about these students with other data points, such as major or gender, could easily result in reidentification of many of these individuals.

Outliers in patron demographic data can also identify individuals in the absence of names and ID numbers. For example, two studies list maximum age of study participants as 49.8 and 83 years old (citations omitted in order to protect students' identities). As students 35 and older represented less than one fifth of enrollments in U.S. postsecondary institutions in 2015 (Digest of Educational Statistics, 2017), these oldest students are likely to fall outside of the normal distribution of students' ages and thus may be at greater risk for reidentification. Small populations and outliers should therefore raise a red flag during the anonymization process and be generalized into larger categories or removed from analysis altogether.

Behavior tracking also allows for reidentification of personal information. For example, individuals have been identified by the web pages they visited and the movies they watched on Netflix, resulting in harm to both the individual creating the data trail and the corporation holding the data (Hern, 2017; Narayanan & Shmatikov, 2008). Within the library, resources consulted can paint a similar identifying picture. Most libraries characterize resource usage as either the student "did use" or "didn't use" that resource type during the study or group usage into several frequency levels. However, eight studies counted the total number of transactions in a specific category (Allison, 2015; Çetin & Howard, 2016; Collins & Stone, 2014; LeMaistre et al., 2018; Montenegro et al., 2016; Nackerud et

al., 2013; Scott, 2014; Thorpe et al., 2016). This method is touted as a way to protect patron privacy (Nackerud et al., 2013) when it may not, in fact, truly do so. Outliers, such as the individual who checked out/renewed 9,324 items (Nackerud et al., 2013), can identify patrons even in the absence of name or ID numbers. More concerning, due to higher risk of reidentification, are studies that track the time duration that a user engaged with an electronic resource. Montenegro et al. (2016) tracked both time, in 10-minute increments, and number of sessions of e-resource usage, broadly. Two studies from the University of Wollongong were more specific, following individuals' use of particular electronic resources in 10-minute increments rather than broad categories. The Pepper and Jantti study even publishes resources that were only used for time durations of only 20 or 30 minutes across the full group of patrons, increasing the probability that individual patrons could be identified from their resource use behavior even in the absence of reported names (B. Cox & Jantti, 2012a; Pepper & Jantti, 2015). Libraries must consider how patron resource use, in addition to demographic information, can be identifying even in the absence of a primary identifier like username.

Poor anonymization is not a problem unique to libraries (Barbaro & Zeller, 2006; J. Cox, 2016; Pandurangan, 2014). Anonymization is difficult and, as some experts argue, may even be impossible when using a data set based on information about people (Narayanan & Felten, 2014; Ohm, 2009). Having expert help with identified information is critical. The library learning analytics studies that acknowledge that demographic information is identifying both have a separate campus technology or consortial group running their servers (Jantti & Cox, 2013; Showers & Stone, 2014). They employ storage servers that actively prevent librarians from querying for patron-level information, such as by leveraging HIPAA health record technology (Renaud et al., 2015). Even these technologies are not foolproof, as acknowledged by this example from Jantti: "Hypothetically, if we only have five students from Botswana, then it may be possible to identify those individuals from manipulating various aggregated views filtered to citizenship" (Jantti & Cox, 2013, p. 167). Expert knowledge may also come from local records managers and the campus legal team, which should be versed in the national laws (such as FERPA in the United States and the Data Protection Act in the United Kingdom) that provide guidance on what constitutes student PII and how to handle this information. No matter the source, library learning analytics projects must consult experts to be sure that their anonymization practices protect patron privacy to the fullest extent and adjust their data collection and security practices to better protect analytics data that potentially cannot be fully anonymized.

Options and Informed Consent

While not explicitly a data management practice, some libraries' broad data collec-

tion practices run into issues with informed consent. The need for informed consent is touched on in the NISO "Data Collection and Use" and "Options and Informed Consent" principles. The former principle states that "users' personal data should only be used for purposes disclosed to them and to which they consent," and the latter principle dictates how students should automatically be opted-out "when personal data are not required to provide services" and that students opting in should later be allowed to opt out and delete their data (NISO, 2015, "6. Options and Informed Consent," para. 1). Failings around opt-in/opt-out and informed consent are rampant within library learning analytics publications.

Of the studies examined for this review, just three explicitly perform analytics only on students who have opted in and only one study describes an opt-out mechanism. The best example of "opt in" is Thorpe et al., which harvested and analyzed library usage data for the 75 students opting into the study (Thorpe et al., 2016). Stemmer and Mahan also performed data harvesting and analytics only on students who opted to take the library's survey, though they did not publish any information describing their informed consent process (Stemmer & Mahan, 2015, 2016). LeMaistre et al. (2018) provide the sole example of "opt out," linking to their Terms of Use and Privacy Policy from the EZ-Proxy login web page (proxy logins served as the main library usage data collected for the study). This mechanism is not ideal for two reasons: first, opting out only stops the collection of new data about a student and does not result in the deletion of old data; and second, notification about the process may not be easy to discover given that none of the 3,530 students covered by the study opted out during the semester the study was conducted. Libraries need to start providing transparent mechanisms for students to opt in to studies—including the possibility for subsequent "opt out" and data deletion—instead defaulting to automatic student inclusion. Without allowing students to opt into studies and setting "opt out" as the default, libraries act in the same manner as the organizations against which they advocate for user privacy (Macrina, 2017).

Informed consent is a problematic area in library learning analytics studies. The Nuremberg Code defines informed consent for human subjects research as the following:

> The person involved should have legal capacity to give consent; should be so situated as to be able to exercise free power of choice, without the intervention of any element of force, fraud, deceit, duress, over-reaching, or other ulterior form of constraint or coercion; and should have sufficient knowledge and comprehension of the elements of the subject matter involved, as to enable him to make an understanding and enlightened decision. This latter element requires that, before the acceptance of an affirmative decision by the experimental subject, there should

> be made known to him the nature, duration, and purpose of the experiment; the method and means by which it is to be conducted; all inconveniences and hazards reasonably to be expected; and the effects upon his health or person, which may possibly come from his participation in the experiment. (*The Nuremberg Code*, 1949, item 1)

Five studies provide explicit information about informed consent for at least some portion of the study data (Bowles-Terry, 2012; B. Cox & Jantti, 2012a; Stone, Pattern, et al., 2011; Stone & Ramsden, 2012; Thorpe et al., 2016). Three of these studies describe obtaining informed consent for a focus group portion of their study but do not say anything about obtaining consent for broad data mining. An example of this disconnect appears in Bowles-Terry's (2012) study, in which phase one consisted of focus groups and used informed consent, while data mining made up phase two, with no published evidence that informed consent was obtained. Stone et al. followed up analysis of analytics data with focus groups (Stone, Pattern, et al., 2011; Stone & Ramsden, 2012); informed consent was conducted for these groups, yet the studies provided no information on consent processes for collection of the learning analytics data. Two other studies describe obtaining consent for all their analysis, though using different methods. Thorpe et al. (2016) offers the best example of gathering informed consent, asking for consent at the point of collecting usernames and allowing students to opt in to the study. The library also trained its staff on the consent form and username collection protocols. Cox and Jantti describe how the university students "are required as part of the enrollment process to provide consent to the university to use their personal information for certain purposes, and the library [learning analytics system] fell within the scope of one of those purposes" (B. Cox & Jantti, 2012a, p. 310). This consent, however, is not compliant with The Nuremberg Code's stipulation for the "free power of choice," as students cannot deny consent and still enroll.

Fifteen studies describe data collection strategies where novel analysis of patron data may warrant obtaining informed consent, but no information on the collection of consent is provided. Six libraries analyze student outcomes correlated to card swipes at a library entrance (Collins & Stone, 2014; de Jager et al., 2018; Goodall & Pattern, 2011; J. L. Jones, 2010; Renaud et al., 2015; Scarletto et al., 2013) but there is no published evidence that these patrons are informed of the full use of their ID data and its potential hazards. Another three studies use the strategy of collecting IDs by always requiring user authentication for access to electronic resources (Cherry et al., 2013; Massengale et al., 2016; Murray et al., 2016), but this does not equate with informed consent either when the data is being used for the secondary analysis purpose. Analyses of space usage raise similar issues (Kome, 2017; "Measure the Future – Libraries & Open Hardware," n.d.).

More problematic is the following strategy from Minnesota that attempts to cover a gap in their data collection: "Due to IP-based authentication, this method does not record all on-campus usage of databases, e-journals, and e-books. However, Internet ID can often be captured anyway, if the on-campus user has logged into another service, such as campus e-mail" (Nackerud et al., 2013, p. 135). Authenticating for one campus service cannot provide proper informed consent to mine data for another campus service. An earlier statement by Nackerud et al. (2012) on the fact that it is "difficult to gather Internet IDs if students don't give them to us" reinforces the lack of consent for gathering this type of information (Nackerud et al., 2012, p. 5). The other four articles published about this data set (Soria et al., 2014, 2017b, 2013, 2015) do not provide further clarity, stating only that this data is collected via a "click-through" script. Under the Contextual Integrity theory of privacy (Nissenbaum, 2010), all of these examples break the context for which data was originally collected and may therefore violate a patron's expectations of privacy. Even the act of librarians analyzing patron data can violate privacy, if patrons do not expect librarians to collect data and use it in this fashion. Patrons may expect librarians to be looking at their data or expect complete privacy, but this cannot be elucidated without engaging with an informed user or by conducting research into patron privacy expectations for library learning analytics, which has not been done. Libraries must think critically about how novel uses of patron data may operate outside of patron privacy expectations, thus requiring implementation of transparent policies and informed consent.

Another issue with informed consent is that, while NISO Consensus Principles advise for informed consent and opt-in participation, IRB approvals may not require informed consent for learning analytics projects due to possible "IRB-exempt" status. In its "Introduction to the Responsible Conduct of Research," the Office of Research Integrity at the U.S. Department of Health and Human Services defines one type for IRB-exempt research as "research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if unidentifiable or publicly available" (Steneck, 2007, p. 41). This exempt classification applies to many of the studies reviewed here due to their foundations in existing library and university data sets. Being categorized as "IRB-exempt" often, but not universally, means that a research project is not required to use the full range of standard human subject procedures such as informed consent. Libraries, however, must consider the second part of the Office of Research Integrity's definition that requires individuals in the data set to be unidentifiable. Given the scope of poor anonymization practices highlighted in this review, there are legitimate concerns of patron identifiability in IRB-exempt studies. Samson demonstrates this problem perfectly, stating that "the Institutional Review Board exempted this project due to the anonymization of the data" (p. 621) while also describing how the study's suppos-

edly anonymized data is full of indirect identifiers such as gender, major, and ethnicity (Samson, 2014); this study is only one of several with similar problems. Libraries and institutional review boards both need to be more critical of privacy practices in order to truly assure that library learning analytics research merits "exempt" status. Without a guarantee that data sets can be truly anonymized, libraries should follow the NISO Consensus Principles and human subject protocols to always have patrons opt into learning analytics studies via explicit informed consent.

Sharing Data with Others

This NISO principle dictates that libraries should carefully consider the privacy implications of access to user data before sharing and that "user activity data to be shared should be anonymized and aggregated to a level that minimizes privacy risks to individual users, unless the user has opted-in to a service" (NISO, 2015, "7. Sharing Data with Others," para. 2).

Six studies discuss sharing data and results apart from their publication (Black & Murphy, 2017; Kot & Jones, 2015; LeMaistre, 2015; Squibb & Mikkelsen, 2016; Stone & Ramsden, 2012; Stone, Ramsden, et al., 2011). Some take the approach of making aggregated data available, such as through a dashboard (LeMaistre, 2015) or by presenting their aggregate findings to campus stakeholders to advocate for increased library engagement. Stone et al.'s LIDP analytics project generated an anonymized, openly licensed data set (Pattern, 2011), though data release was performed in consultation with "Jisc Legal and the University of Huddersfield's Legal and Data Protection Officers" (Stone & Ramsden, 2012, p. 550). Anonymization also went beyond superficial removal of usernames (Stone & Ramsden, 2012; Stone, Ramsden, et al., 2011), the specifics of which are described in the Anonymization section of this review.

More concerning is the approach that goes directly against the NISO principle by sharing individual student data with other university authorities. The University of Wollongong's library suggested sharing student library resource usage patterns with instructors to help boost library use via instructor intervention (Pepper & Jantti, 2015). This is part of a broader effort to include library data into institution-level learning analytics systems (Jantti, 2016). Minnesota is piloting alerting advisors of an individual's low library use (Nackerud, Fransen, Peterson, & Mastel, 2015) and already allows for advisors to receive notification about introductory library workshop completion and for integrated library referrals from advisors (Fransen & Peterson, 2016; Oakleaf, Macintyre, et al., 2017). De Jager et al. (2018) place library usage data in the campus data warehouse, though the data is only viewable in the aggregate and only by library staff. Libraries appear to be at

the start of a movement arguing for the inclusion of library analytics data into broader institutional learning analytics platforms (M. Jones, 2015; "LG-98-17-0019-17 | Library Integration in Institutional Learning Analytics (LIILA)," 2017; Oakleaf, 2016; Oakleaf, Brown, et al., 2017; Oakleaf, Walter, & Brown, 2017), which is in direct conflict with the NISO principle on sharing data with others. While most libraries may be following this NISO principle by sharing learning analytics results only in the aggregate, the exceptions that share patron information with other institutional authorities raise significant concerns.

Access to One's Own Data

The final NISO principle evaluated in this review states that "users should have the right to access their own personal information or activity data" and that "so far as is feasible . . . users may request correction or deletion" (NISO, 2015, "10. Access to One's Own User Data," para. 1). This principle normally would not apply to learning analytics studies, as data is usually deidentified. However, given that some libraries are sharing patron-identified library data with other institutional offices, this review must evaluate data practices around access to one's own data.

None of the reviewed studies describes any mechanism for making raw data available to the patrons involved in library learning analytics projects. Four studies describe plans to interact with students about the aggregate results, either to promote library use or gain greater understanding of study results (Goodall & Pattern, 2011; Kot & Jones, 2015; Massengale et al., 2016; Stone & Ramsden, 2012). Massengale et al. (2016) shared high-level results with their institution's Student Library Advocates group for feedback. Stone and Ramsden (2012) describe a partner institution using project data to engage student interest, though they provide very few details on the mechanics. Such access to aggregate data is expected and in line with the NISO Consensus Principles.

More problematic are the few cases, mentioned in the previous section, where university authorities gain access to library-use data on individual students, yet there is no published evidence that the individual students themselves can gain access to that data. The fact that a few libraries share individual student–level data with other campus units and none of them appear to share with patrons is concerning. Where libraries insist on sharing student data with university administrators, the imbalance of not making that data available to the people to which that data corresponds has the potential to erode patron trust in libraries. A common argument brought up at this point is that university officials are allowed access to educational records. For example, FERPA states that records may be accessed by school officials who the "institution has determined to have legitimate

educational interests" (Family Educational Rights and Privacy Act, 1974, §99.31 (1)(i)(A)) However, FERPA also states that a "student must be given the opportunity to inspect and review the student's education records" (§99.10 (a)). Access to one's own data is required by both NISO and FERPA guidelines when that data is held at the level of specific individuals. If libraries choose to share patron-identified data with other university officials—going against NISO principles and library codes of ethics—then they must also provide mechanisms for students to access their own data. This is more evidence for why libraries must seriously consider all of the consequences of sharing patron-identified data.

In the learning analytics context, NISO's recommendation for access to one's own data—and sharing individual patron data more broadly—conflicts with other NISO recommendations for good security, minimized data retention, and anonymization. The risks of the other principles far outweigh the benefits of following the "Access to One's Own Data" principle. It is better for libraries to aggregate data and achieve the transparency advocated for in this principle by informing patrons about the scope of data collection and the data's intended use. The result would be informed patrons, a lower data security burden, and better protection of patron privacy.

DISCUSSION

A literature review of published data management practices will always result in an incomplete understanding of how individual projects are managing their data due to the fact that data handling information is only intermittently present in published results. However, Table 1 demonstrates that there is enough data handling information present within published library learning analytics studies to form a basis for this evaluation. This review found many examples of inadequate data management practices, including extended data retention, a broad scope of data collection, insufficient anonymization, lack of informed consent, and sharing of patron-identified data. Further study and transparency is merited on how libraries work through legal requirements for data, document data handling in analytics projects (such as through policy or a data management plan), secure data, and delete analytics data. These findings should be used to both guide further research and to give libraries a starting point for improving their data practices in learning analytics.

With respect to the central question behind this review, we can draw an initial conclusion that academic libraries' actual data practices are not living up to data best practices in the area of learning analytics. Of the data practices that are reported, many fail to meet the NISO Principles guidelines. With respect to anonymization, 18 of the 54 studies

either mentioned anonymization and/or described using anonymization techniques, yet most of the 18 demonstrate incomplete anonymization: seven studies only removed direct identifiers; six described more complex data suppression strategies that do not necessarily equate to anonymization; and five did not describe their anonymization methods. Taken as a whole, there is little evidence that libraries are actually anonymizing data. This problem is made worse in relation to IRB and consent. IRB-exempt status depends on data being "unidentifiable," yet a large portion of library data is only seemingly anonymous. With only three studies letting students opt into data harvesting, one discussing "opting out," and five of the 54 studies explicitly mentioning consent (though not necessarily for all portions of a project), it is likely that many of the reviewed studies were conducted under an IRB exemption. This means that insufficient anonymization matters because it can negate the validity of IRB exemption.

Similar data management concerns exist at the boundary of security, scope of data collection, and data retention. Only six studies describe the management of their security systems, yet 46 of the 54 studies imply transferring patron-level data between the library and university. Adding to security concerns are the 22 studies describing extended retention of data, eight studies enumerating a large range of data collected, and five studies examining sensitive socioeconomic data or proxies for such. Large amounts of sensitive data and extended data retention increase the necessity of security, yet even good security cannot entirely prevent a data breach. This means that libraries need to take security much more seriously, of which a part is reducing both the scope and retention period of sensitive data.

There were no studies that matched the ideal outlined in the NISO Consensus Principles, yet there were studies that handled data well in certain areas. These include Thorpe et al.'s (2016) use of opt-in and consent, O'Kelly's (2016) neat avoidance of the burden of security by having the library hold only aggregate data, and Wong's suppression of data during transfer and analysis (Wong & Cmor, 2011; Wong & Webb, 2011). Additionally, academic libraries can look to the example set by Yoose (2017a), in a public library setting, on integrating security and assessment. It is likely that there are further examples of good data practices that are happening in academic libraries that are not being reported. However, it should be noted that researchers in general struggle with managing data well, and this review provides evidence that librarians are not entirely different in this area. Ideally, this review will prompt more transparency around data practices in library learning analytics projects—potentially using the underrepresented NISO principles in Table 1 as a guide—and the publication of further examples of good data practices.

CONCLUSION

This review provides evidence of a conflict between libraries' commitment to patron privacy and their current data handling practices in learning analytics projects, as learning analytics data may not be as protected as libraries believe it is. Libraries therefore need to invest in training staff and providing infrastructure in security, up-to-date anonymization procedures, and data privacy practices if they wish to conduct learning analytics research in a responsible manner. There is also the potential to involve library staff trained in data management in learning analytics projects. Libraries should also be more transparent about data practices, both via policy and when publishing. The desire to protect patron privacy must be followed up with proper data management practices. Anything else places patron privacy at risk.

ACKNOWLEDGMENTS

The author thanks Dorothea Salo and Abigail Goben for their guidance in developing this research idea and for their helpful feedback on this article's many drafts.

REFERENCES

American Library Association. (2008). *Code of ethics of the American Library Association*. Retrieved from http://www.ala.org/advocacy/sites/ala.org.advocacy/files/content/proethics/codeofethics/Code of Ethics of the American Library Association.pdf

Alameda County Library. (2017). Frequently asked questions. Retrieved November 20, 2017, from https://www.aclibrary.org/content/frequently-asked-questions

Allison, D. (2015). Measuring the academic impact of libraries. *Portal: Libraries and the Academy*, *15*(1), 29–40. https://doi.org/10.1353/pla.2015.0001

Asher, A. (2017). Use, security, and ethics of data collection: Risks, benefits, and user privacy: Evaluating the ethics of library data. In B. Newman & B. Tijerina (Eds.), *Protecting patron privacy* (pp. 43–56). Lanham, MD: Rowman & Littlefield.

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*(7604), 452–454. https://doi.org/10.1038/533452a

Barbaro, M., & Zeller, T. (2006, August 9). A face is exposed for AOL searcher No. 4417749. *New York Times*. Retrieved from http://www.nytimes.com/2006/08/09/technology/09aol.html

Black, E. L., & Murphy, S. A. (2017). The Out Loud assignment: Articulating library contributions to first-year student success. *Journal of Academic Librarianship*, *43*(5), 409–416. https://doi.org/10.1016/j.acalib.2017.06.008

Borgman, C. L. (2016). *Big data, little data, no data: Scholarship in the networked world*. Cambridge, MA: MIT Press.

Bowles-Terry, M. (2012). Library instruction and academic success: A mixed-methods assessment of a library instruction program. *Evidence Based Library and Information Practice*, *7*(1), 82. https://doi.org/10.18438/B8PS4D

Briney, K. (2015). *Data management for researchers: Organize, maintain and share your data for research success*. Exeter, UK: Pelagic Publishing.

Briney, K., Goben, A., & Zilinski, L. (2015). Do you have an institutional data policy? A review of the current landscape of library data services and institutional data policies. *Journal of Librarianship and Scholarly Communication*, *3*(2), eP1232. https://doi.org/10.7710/2162-3309.1232

Carlson, J., & Stowell-Bracke, M. (2013). Data management and sharing from the perspective of graduate students: An examination of the culture and practice at the water quality field station. *Portal: Libraries and the Academy*, *13*(4), 343–361. https://doi.org/10.1353/pla.2013.0034

Çetin, Y., & Howard, V. (2016). An exploration of the relationship between undergraduate students' library book borrowing and academic achievement. *Journal of Librarianship & Information Science*, *48*(4), 382–388. https://doi.org/10.1177/0961000615572404

Charlton, G. (2017). Data exchange and the art of iterating security checkups. Retrieved June 2, 2017, from https://chooseprivacyweek.org/data-exchange-and-the-art-of-iterating-security-checkups/

Cherry, E., Rollins, S. H., & Evans, T. (2013). Proving our worth: The impact of electronic resource usage on academic achievement. *College & Undergraduate Libraries*, *20*(3–4), 386–398. https://doi.org/10.1080/10691316.2013.829378

Collins, E., & Stone, G. (2014). Understanding patterns of library use among undergraduate students from different disciplines. *Evidence Based Library and Information Practice*, *9*(3), 51. https://doi.org/10.18438/B8930K

Cook, J. M. (2014). A library credit course and student success rates: A longitudinal study. *College & Research Libraries*, *75*(3), 272–283. https://doi.org/10.5860/crl12-424

Corti, L., Van den Eynden, V., Bishop, L., & Woollard, M. (2014). *Managing and sharing research data: A guide to good practice*. Los Angeles: SAGE Publications.

Coulter, P., Clarke, S., & Scamman, C. (2007). Course grade as a measure of the effectiveness of one-shot information literacy instruction. *Public Services Quarterly*, *3*(1), 147–163. https://doi.org/10.1300/J295v03n01_08

Cox, B., & Jantti, M. (2012a). Capturing business intelligence required for targeted marketing, demonstrating value, and driving process improvement. *Library & Information Science Research*, *34*(4), 308–316. https://doi.org/10.1016/j.lisr.2012.06.002

Cox, B., & Jantti, M. (2012b). Discovering the impact of library use and student performance. *EDUCAUSE Review*. Retrieved from http://www.educause.edu/ero/article/discovering-impact-library-use-and-student-performance

Cox, J. (2016, May 12). 70,000 OkCupid users just had their data published. *Motherboard*. Retrieved from https://motherboard.vice.com/en_us/article/8q88nx/70000-okcupid-users-just-had-their-data-published

Cullen, K. (2005). Delving into data. Retrieved May 31, 2017, from http://lj.libraryjournal.com/2005/08/technology/delving-into-data/#_

Data Protection Act 1998 (1998). UK: Statute Law Database. Retrieved from http://www.legislation.gov.uk/ukpga/1998/29/contents

de Jager, K., Nassimbeni, M., Daniels, W., & D'Angelo, A. (2018). The use of academic libraries in turbulent times: Student library behaviour and academic performance at the University of Cape Town. *Performance Measurement & Metrics*, *19*(1), 40–52. https://doi.org/10.1108/PMM-09-2017-0037

Department for Business Innovation. (2013). *2013 Information Security Breaches Survey*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/200455/bis-13-p184-2013-information-security-breaches-survey-technical-report.pdf

Digest of Educational Statistics. (2017). Table 303.40. Total fall enrollment in degree-granting postsecondary institutions, by attendance status, sex, and age: Selected years, 1970 through 2026. Retrieved from https://nces.ed.gov/programs/digest/d16/tables/dt16_303.40.asp?current=yes

Enis, M. (2014). Gale releases analytics on demand, a demographic GIS for libraries. Retrieved August 10, 2017, from http://www.thedigitalshift.com/2014/04/research/gale-releases-analytics-demand-demographic-gis-libraries/

Family Educational Rights and Privacy Act, Pub. L. No. 20 U.S.C. § 1232g (1974). United States: United States Congress.

Fransen, J., & Peterson, K. (2016, November). Graduate in four years? Yes, the library can help with that! Paper presented at the Library Assessment Conference, Arlington, VA. Retrieved from http://old.libraryassessment.org/bm~doc/77-fransen-2016.pdf

Freedman, L. P., Cockburn, I. M., Simcoe, T. S., Parkes, H., & Gelber, C. (2015). The economics of reproducibility in preclinical research. *PLoS Biology*, *13*(6), e1002165. https://doi.org/10.1371/journal.pbio.1002165

Garfinkel, S. L. (2015). *NISTIR 8053: De-identification of personal information*. https://doi.org/10.6028/NIST.IR.8053

Garg, R. (2016). Top breaches in higher education 2015–2016. Retrieved from http://zecurion.com/2016/05/24/top-breaches-in-higher-education-in-2015-2016/

Gemalto. (2017). *First half 2017 breach level index report: Identity theft and poor internal security practices take a toll*. Retrieved from https://www.gemalto.com/press/Pages/First-Half-2017-Breach-Level-Index-Report-Identity-Theft-and-Poor-Internal-Security-Practices-Take-a-Toll.aspx

Goben, A., & Griffin, T. (in press). In Aggregate: Trends, Needs, and Opportunities from Research Data Management Surveys. *College & Research Libraries*.

Goben, A., & Raszewski, R. (2015). The data life cycle applied to our own data. *Journal of the Medical Library Association*, *103*(1), 40–44. https://doi.org/10.3163/1536-5050.103.1.008

Goodall, D., & Pattern, D. (2011). Academic library non/low use and undergraduate student achievement. *Library Management*, *32*(3), 159–170. https://doi.org/10.1108/01435121111112871

Goodman, A., & Goodman, D. (2008). America's most dangerous librarians. *Mother Jones*. Retrieved November 22, 2017, from http://www.motherjones.com/politics/2008/09/americas-most-dangerous-librarians/

Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, *26*(2), 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

Haddow, G. (2013). Academic library use and student retention: A quantitative analysis. *Library & Information Science Research*, *35*(2), 127–136. https://doi.org/10.1016/j.lisr.2012.12.002

Haddow, G., & Joseph, J. (2010). Loans, logins, and lasting the course: Academic library use and student retention. *Australian Academic & Research Libraries*, *41*(4), 233–244. https://doi.org/10.1080/00048623.2010.10721478

Hellman, E. (2014a). Analysis of privacy leakage on a library catalog webpage. Retrieved July 19, 2017, from https://go-to-hellman.blogspot.com/2014/09/analysis-of-privacy-leakage-on-library.html

Hellman, E. (2014b). Go To Hellman: Libraries are giving away the user-privacy store [Blog post.] Retrieved June 1, 2017, from https://go-to-hellman.blogspot.com/2014/08/libraries-are-giving-away-user-privacy.html

Hellman, E. (2016). 97% of Research library searches leak privacy . . . and other disappointing statistics [Blog post.] Retrieved June 9, 2017, from https://go-to-hellman.blogspot.com/2016/05/97-of-research-library-searches-leak.html

Hern, A. (2017, August 1). "Anonymous" browsing data can be easily exposed, researchers reveal. *Guardian*. Retrieved from https://www.theguardian.com/technology/2017/aug/01/data-browsing-habits-brokers

International Federation of Library Associations. (2012). Professional codes of ethics for librarians. Retrieved July 28, 2017, from https://www.ifla.org/faife/professional-codes-of-ethics-for-librarians

Iqbal, S. A., Wallach, J. D., Khoury, M. J., Schully, S. D., Ioannidis, J. P. A., & Ubel, P. (2016). Reproducible research practices and transparency across the biomedical literature. *PLoS Biology*, *14*(1), e1002333. https://doi.org/10.1371/journal.pbio.1002333

Jantti, M. (2016). Libraries and big data: A new view on impact and affect. In J. Atkinson (Ed.), *Quality and the academic library: Reviewing, assessing and enhancing service provision* (pp. 267–274). Amsterdam: Chandos. https://doi.org/10.1016/B978-0-12-802105-7.00026-9

Jantti, M., & Cox, B. (2010). Measuring the value of library resources and student academic performance through relational datasets. In *Proceedings of the Library Assessment Conference: Building effective, sustainable, practical assessment* (pp. 525–532). http://old.libraryassessment.org/bm~doc/proceedings-lac-2010.pdf

Jantti, M., & Cox, B. (2013). Measuring the value of library resources and student academic performance through relational datasets. *Evidence Based Library and Information Practice*, *8*(2), 163. https://doi.org/10.18438/B8Q89F

Jisc. (2015). *Code of practice for learning analytics*. Retrieved from https://www.jisc.ac.uk/guides/code-of-practice-for-learning-analytics

Johnston, L., & Jeffryes, J. (2014). Data management skills needed by structural engineering students: Case study at the University of Minnesota. *Journal of Professional Issues in Engineering Education and Practice*, *140*(2), 5013002. https://doi.org/10.1061/(ASCE)EI.1943-5541.0000154

Jones, J. L. (2010, October). Using library swipe-card data to inform decision making. Paper presented at the Georgia Council of Media Organizations Conference, Athens, GA. Retrieved from https://works.bepress.com/jenniferlinkjones/2/

Jones, K. M. L., & Salo, D. (2018). Learning analytics and the academic library: Professional ethics commitments at a crossroads. *College & Research Libraries*, *79*(3), 304. https://doi.org/10.5860/crl.79.3.304

Jones, M. (2015). LAMP to be integrated into Jisc's learner and business analytics & activities. Retrieved August 10, 2017, from http://jisclamp.mimas.ac.uk/

Khalil, M., & Ebner, M. (2016). De-identification in learning analytics. *Journal of Learning Analytics*, *3*(1), 129–138. https://doi.org/10.18608/jla.2016.31.8

Kogut, A. (2016, October). Academic library services and undergraduate academic success: Trends in research literature. Paper presented at the 2016 *Library Assessment Conference*, Arlington, VA. Retrieved from http://libraryassessment.org/bm~doc/59-kogut-2016.pdf

Kome, S. (2017, April). Protect researcher privacy in the surveillance era. Paper presented at the Coalition for Networked Information Membership Meeting, Albuquerque, NM. Retrieved from https://www.youtube.com/watch?v=xjAcCbUdUkI&feature=youtu.be

Kot, F. C., & Jones, J. L. (2015). The impact of library resource utilization on undergraduate students' academic performance: A propensity score matching design. *College & Research Libraries*, *76*(5), 566–586. https://doi.org/10.5860/crl.76.5.566

Lambert, A. D., Parker, M., & Bashir, M. (2015). Library patron privacy in jeopardy an analysis of the privacy policies of digital content vendors. *Proceedings of the Association for Information Science and Technology*, *52*(1), 1–9. https://doi.org/10.1002/pra2.2015.145052010044

LeMaistre, T. (2015). Cost per user: Analyzing EZProxy logs for collection development. *Proceedings of the Charleston Library Conference*. Retrieved from https://docs.lib.purdue.edu/charleston/2015/endusers/9/

LeMaistre, T., Shi, Q., & Thanki, S. (2018). Connecting library use to student success. *Portal: Libraries & the Academy*, *18*(1), 117–140. https://doi.org/10.1353/pla.2018.0006

LG-98-17-0019-17 | Library Integration in Institutional Learning Analytics. (2017). Retrieved November 3, 2017, from https://www.imls.gov/grants/awarded/lg-98-17-0019-17

Macrina, A. (2017). Library Freedom Project – Making real the promise of intellectual freedom in libraries. Retrieved November 24, 2017, from https://libraryfreedomproject.org/

Magi, T. (2010). A content analysis of library vendor privacy policies: Do they meet our standards? *University Libraries Faculty and Staff Publications*. https://doi.org/10.5860/0710254

Massengale, L., Piotrowski, P., & Savage, D. (2016). Identifying and articulating library connections to student success. *College & Research Libraries*, *77*(2), 227–235. https://doi.org/10.5860/crl.77.2.227

McCarthy, S. C. (2017). At issue: Exploring library usage by online learners with student success. *Community College Enterprise*, *23*(2), 27–31.

Measure the Future – Libraries & Open Hardware. (n.d.). Retrieved June 1, 2017, from http://measurethefuture.net/

Montenegro, M., Clasing, P., Kelly, N., Gonzalez, C., Jara, M., Alarcón, R., . . . Saurina, E. (2016). Library resources and students' learning outcomes: Do all the resources have the same impact on learning? *Journal of Academic Librarianship*, *42*(5), 551–556. https://doi.org/10.1016/j.acalib.2016.06.020

Morgan, S. (2017, December). Why incident response is the best cybersecurity ROI. Retrieved December 18, 2017, from https://www.csoonline.com/article/3243246/leadership-management/why-incident-response-is-the-best-cybersecurity-roi.html

Murray, A., Ireland, A., & Hackathorn, J. (2016). The value of academic libraries: Library services as a predictor of student retention. *College & Research Libraries*, *77*(5), 631–642. https://doi.org/10.5860/crl.77.5.631

Nackerud, S., Fransen, J., Peterson, K., & Mastel, K. (2013). Analyzing demographics: Assessing library use across the institution. *Portal: Libraries and the Academy*, *13*(2), 131–145. https://doi.org/10.1353/pla.2013.0017

Nackerud, S., Fransen, J., Peterson, K., & Mastel, K. (2015). Retention, student success and academic engagement at Minnesota (University of Minnesota). In B. Showers (Ed.), *Library analytics and metrics: Using data to drive decisions and services* (pp. 58–66). London: Facet Publishing.

Nackerud, S., Fransen, J., Peterson, K., Mastel, K., Soria, K., & Peterson, D. (2012). Library data and student success. In *Library Technology Conference*. Retrieved from http://digitalcommons.macalester.edu/libtech_conf/2012/sessions/28

Narayanan, A., & Felten, E. W. (2014, July 9). *No silver bullet: De-identification still doesn't work*. Retrieved from http://randomwalker.info/publications/no-silver-bullet-de-identification.pdf

Narayanan, A., & Shmatikov, V. (2008, May). Robust de-anonymization of large sparse datasets (How to break anonymity of the Netflix Prize Dataset). Paper presented at the 2008 IEEE Symposium on Security and Privacy, Oakland, CA. https://doi.org/10.1109/SP.2008.33

Nicholson, S., & Smith, C. A. (2007). Using lessons from health care to protect the privacy of library users: Guidelines for the de-identification of library data based on HIPAA. *Journal of the American Society for Information Science and Technology*, *58*(8), 1198–1206. https://doi.org/10.1002/asi.20600

NISO. (2015). *NISO Consensus Principles on User's Digital Privacy in Library, Publisher, and Software-Provider Systems (NISO Privacy Principles)*. Retrieved from http://www.niso.org/apps/group_public/download.php/16064/NISO Privacy Principles.pdf

Nissenbaum, H. F. (2010). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford, CA: Stanford Law Books.

Oakleaf, M. (2016). Getting ready & getting started: Academic librarian involvement in institutional learning analytics initiatives. *Journal of Academic Librarianship*, *42*(4), 472–475. https://doi.org/10.1016/j.acalib.2016.05.013

Oakleaf, M., Brown, M., Nackerud, S., Jantti, M., & Abel, R. (2017). Closing the data gap: Integrating library data into institutional learning analytics. In *Educause Annual Conference*. Retrieved from https://events.educause.edu/~/media/files/events/user-uploads-folder/e17/sesso21/closing-the-data-gap--presentation-part-1.pdf

Oakleaf, M., Gilchrist, D., Kingma, B., Kyrillidou, M., Kuh, G., Owen, P. L., … Barnes, T. (2010). *Value of academic libraries: A comprehensive research review and report*. Retrieved from www.acrl.ala.org/value

Oakleaf, M., Macintyre, R., Sharma, A., Krieb, D., Fransen, J., Nackerud, S., & Peterson, K. (2017). Data in the library is safe, but that's not what data is meant for: Exploring the longitudinal, responsible use of library and institutional data to understand and increase student success #acrlcorrelate. In *ACRL*. Retrieved from https://files.zotero.net/13408627639/1298_Data in the Library is Safe ACRL 2017.pdf

Oakleaf, M., Walter, S., & Brown, M. (2017, August 14). The academic library and the promise of NGDLE. *Educause Review*. Retrieved from http://er.educause.edu/articles/2017/8/the-academic-library-and-the-promise-of-ngdle

Odeh, A. Y. (2012). Use of information resources by undergraduate students and its relationship with academic achievement. *Libri*. https://doi.org/10.1515/libri-2012-0018

Ohm, P. (2009, August 13). Broken promises of privacy: Responding to the surprising failure of anonymization. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450006

O'Kelly, M. (2017). Academic libraries and student retention: The implications for higher education. 2016 Library Assessment Conference: Building Effective, Sustainable, Practical Assessment (pp. 485–490). Arlington, VA: Association of Research Libraries. Retrieved from http://old.libraryassessment.org/bm~doc/78-okelly-2016.pdf

O'Neil, M. (2014). Data breaches put a dent in colleges' finances as well as reputations. Retrieved December 19, 2017, from https://www.chronicle.com/article/Data-Breaches-Put-a-Dent-in/145341

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251). Retrieved from http://science.sciencemag.org/content/349/6251/aac4716.abstract

OrangeBoy, Inc. (2017). Retrieved December 15, 2017, from https://www.orangeboyinc.com/

Palmer, J. (2012). Library analytics - Community survey results. Retrieved from https://www.slideshare.net/joypalmer/survey-library-analyticsfindings

Pandurangan, V. (2014). On taxis and rainbows. Retrieved June 30, 2017, from https://tech.vijayp.ca/of-taxis-and-rainbows-f6bc289679a1

Parry, M. (2011, July 10). Harvard researchers accused of breaching students' privacy. *Chronicle of Higher Education*. Retrieved from https://www.chronicle.com/article/Harvards-Privacy-Meltdown/128166

Pattern, D. (2011). Library impact data project data. Retrieved from http://eprints.hud.ac.uk/id/eprint/11543/

Pepper, A., & Jantti, M. (2015). The tipping point: How granular statistics can make a big difference in understanding and demonstrating value. *Australian Library and Information Association Information Online*. Sydney, Australia. Retrieved from https://works.bepress.com/mjantti/34/

Perrier, L., Blondal, E., Ayala, A. P., Dearborn, D., Kenny, T., Lightfoot, D., . . . MacDonald, H. (2017). Research data management in academic institutions: A scoping review. *PLoS ONE*, *12*(5), e0178261. https://doi.org/10.1371/journal.pone.0178261

Peterson, A. (2014, October 3). Librarians won't stay quiet about government surveillance. *Washington Post*. https://www.washingtonpost.com/news/the-switch/wp/2014/10/03/librarians-wont-stay-quiet-about-government-surveillance/

Ponemon Institute LLC. (2017). *2017 Cost of Data Breach Study*. Retrieved from https://www.ibm.com/security/infographics/data-breach/

Reidsma, M. (2016). The unbearable lightness of vendor security [Blog post]. Retrieved from https://matthew.reidsrow.com/worknotes/177

Renaud, J., Britton, S., Wang, D., & Ogihara, M. (2015). Mining library and university data to understand library use patterns. *Electronic Library*, *33*(3), 355–372. https://doi.org/10.1108/EL-07-2013-0136

Rubel, A., & Jones, K. M. L. (2016). Student privacy in learning analytics: An information ethics perspective. *Information Society*, *32*(2), 143–159. https://doi.org/10.1080/01972243.2016.1130502

Samson, S. (2014). Usage of e-resources: Virtual value of demographics. *Journal of Academic Librarianship*, *40*(6), 620–625.

Scarletto, E. A., Burhanna, K. J., & Richardson, E. (2013). Wide awake at 4AM: A study of late night user behavior, perceptions and performance at an academic library. *Journal of Academic Librarianship*, *39*(5), 371–377. https://doi.org/10.1016/j.acalib.2013.02.006

Schneier, B. (2016, March 4). Data is a toxic asset [Blog post.] Retrieved from https://www.schneier.com/blog/archives/2016/03/data_is_a_toxic.html

Scott, M. (2014). Interlibrary loan article use and user GPA: Findings and implications for library services. *Journal of Access Services*, *11*(4), 229–238. https://doi.org/10.1080/15367967.2014.945116

Shen, Y. (2016). Strategic planning for a data-driven, shared-access research enterprise: Virginia Tech research data assessment and landscape study. *College & Research Libraries*, *77*(4), 500–519. https://doi.org/10.5860/crl.77.4.500

Showers, B., & Stone, G. (2014). Safety in numbers: Developing a shared analytics service for academic libraries. *Performance Measurement and Metrics*, *15*(1/2), 13–22. https://doi.org/10.1108/PMM-03-2014-0008

Siemens, G. (2012). Learning analytics. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12* (p. 4). New York: ACM Press. https://doi.org/10.1145/2330601.2330605

Soria, K. M., Fransen, J., & Nackerud, S. (2014). Stacks, serials, search engines, and students' success: First-year undergraduate students' library use, academic achievement, and retention. *Journal of Academic Librarianship*, *40*(1), 84–91. https://doi.org/10.1016/j.acalib.2013.12.002

Soria, K. M., Fransen, J., & Nackerud, S. (2017a). Beyond books: The extended academic benefits of library use for first-year college students. *College & Research Libraries*, *78*(1). https://doi.org/10.5860/crl.v78i1.16564

Soria, K. M., Fransen, J., & Nackerud, S. (2017b). The impact of academic library resources on undergraduates' degree completion. *College & Research Libraries*, *78*(6). https://doi.org/10.5860/crl.0.0.16626

Soria, K. M., Fransen, J., Nackerud, S., & Kross, A. (2013). Library use and undergraduate student outcomes: New evidence for students' retention and academic success. *Portal: Libraries and the Academy*, *13*(2), 147–164. https://doi.org/10.1353/pla.2013.0010

Soria, K. M., Nackerud, S., & Peterson, K. (2015). Socioeconomic indicators associated with first-year college students' use of academic libraries. *Journal of Academic Librarianship*, *41*(5), 636–643. https://doi.org/10.1016/j.acalib.2015.06.011

Squibb, S. D., & Mikkelsen, S. (2016). Assessing the value of course-embedded information literacy on student learning and achievement. *College & Research Libraries*, *77*(2), 164–183. Retrieved from http://dx.doi.org/10.5860/crl.77.2.164

Stemmer, J. K., & Mahan, D. M. (2015). Assessing the library's influence on freshman and senior level outcomes with user surveys. *Evidence Based Library and Information Practice*, *10*(2), 8. https://doi.org/10.18438/B8PG62

Stemmer, J. K., & Mahan, D. M. (2016). Investigating the relationship of library usage to student outcomes. *College & Research Libraries*, *77*(3), 359–375. https://doi.org/10.5860/crl.77.3.359

Steneck, N. H. (2007). *Introduction to the responsible conduct of research*. Retrieved from https://ori.hhs.gov/sites/default/files/rcrintro.pdf

Stone, G., Pattern, D., & Ramsden, B. (2011). Does library use affect student attainment? A preliminary report on the Library Impact Data Project. *LIBER Quarterly*, *21*(1), 5. https://doi.org/10.18352/lq.8005

Stone, G., Pattern, D., & Ramsden, B. (2012). Library Impact Data Project. *SCONUL Focus*, *54*. Retrieved from https://www.sconul.ac.uk/sites/default/files/documents/8_0.pdf

Stone, G., & Ramsden, B. (2012). Library Impact Data Project: Looking for the link between library usage and student attainment. *College & Research Libraries*, *74*(6), crl12-406. https://doi.org/10.5860/crl12-406

Stone, G., Ramsden, B., & Pattern, D. (2011). Looking for the link between library usage and student attainment. *Ariadne*, (67). Retrieved from http://www.ariadne.ac.uk/issue67/stone-et-al/

Sweeney, L. (2000). *Simple demographics often identify people uniquely* [Working paper]. Retrieved from http://dataprivacylab.org/projects/identifiability/paper1.pdf

Tenopir, C., Birch, B., & Allard, S. (2012). *Academic libraries and research data services: Current practices and plans for the future* [White paper]. Retrieved from http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf

The Nuremberg Code. (1949). Retrieved from https://history.nih.gov/research/downloads/nuremberg.pdf

Thoegersen, J. (2015). Examination of Federal Data Management Plan Guidelines. *Journal of eScience Librarianship*, *4*(1). https://doi.org/10.7191/jeslib.2015.1072

Thorpe, A., Lukes, R., Bever, D. J., & He, Y. (2016). The impact of the academic library on student success: Connecting the dots. *Portal: Libraries and the Academy*, *16*(2), 373–392. Retrieved from https://muse.jhu.edu/article/613847/summary

Toups, M., & Hughes, M. (2013). When data curation isn't: A redefinition for liberal arts universities. *Journal of Library Administration*, *53*(4), 223–233. https://doi.org/10.1080/01930826.2013.865386

U.S. Department of Education. (2017). Data security checklist. Retrieved July 24, 2017, from https://studentprivacy.ed.gov/resources/data-security-checklist

Varnum, K. (2015). Editorial board thoughts: Library analytics and patron privacy. *Information Technology and Libraries*, *34*(4), 2. https://doi.org/10.6017/ital.v34i4.9151

Week.com. (2017, October 12). ALERT: Library patrons in 10 Western Wisconsin counties affected. *News 25*. Retrieved from https://web.archive.org/web/20171027115805/http://www.week.com/story/36583282/2017/10/Thursday/alert-library-patrons-in-10-western-wisconsin-counties-affected-in-data-breach

White, S., & Stone, G. (2010). Maximizing use of library resources at the University of Huddersfield. *Serials: The Journal for the Serials Community*, *23*(2), 83–90. https://doi.org/10.1629/2383

Wong, S. H. R., & Cmor, D. (2011). Measuring association between library instruction and graduation GPA. *College & Research Libraries*, *72*(5), 464–473. https://doi.org/10.5860/crl-151

Wong, S. H. R., & Webb, T. D. (2011). Uncovering meaningful correlation between student academic performance and library material usage. *College and Research Libraries 72*(4). Retrieved from http://crl.acrl.org/index.php/crl/article/view/16168

Yoon, A., & Schultz, T. (2017). Research data management services in academic libraries in the US: A content analysis of libraries' websites. *College & Research Libraries*, *78*(7), 920–933. Retrieved from http://crl.acrl.org/index.php/crl/article/view/16788/18346

Yoose, B. (2017a). Balancing privacy and strategic planning needs: A case study in de-identification of patron data. *Journal of Intellectual Freedom and Privacy*, *2*(1), 15–22. https://doi.org/10.5860/JIFP.V2I1.6250

Yoose, B. (2017b). De-identification and patron data [Blog post]. Retrieved March 8, 2018, from https://chooseprivacyweek.org/de-identification-and-patron-data/

Yoose, B., & Halsey, S. (2016, April). De-identifying patron data to balance privacy and insight. Program session at the *Public Library Association 2016 Conference*, Denver, CO. Retrieved from http://2016.placonference.org/program/de-identifying-patron-data-to-balance-privacy-and-insight/

York, T. T., Gibson, C., & Rankin, S. (2015). Defining and measuring academic success: Practical assessment, research & evaluation, *20*(5). Retrieved from http://www.pareonline.net/getvn.asp?v=20&n=5