Asher, Andrew D. (2017) Risk, Benefits, and User Privacy: Evaluating the Ethics of Library Data. In Protecting Patron Privacy: A LITA Guide. Bobbi Newman & Bonnie Tijerina, Eds. Pp. 43-56. Lanham, MD: Rowman & Littlefield.

# Risk, Benefits, and User Privacy: Evaluating the Ethics of Library Data

**Andrew D. Asher**
**Indiana University**

The collection and analysis of user-level library data is playing an increasingly prominent role in academic libraries. This is perhaps not surprising; the Association of College and Academic Libraries' (ACRL) influential 2010 report on the value of academic libraries emphasized the need for libraries to begin systematically collecting and incorporating large-scale data into their assessment activities (Oakleaf 2010), and many libraries have started actively seeking ways to utilize data about library services and collections in institutional efforts in learning analytics, with the goal of better understanding and measuring student learning, academic success, and engagement.

As institutional information hubs, libraries create a great deal of information about the activities of their users, and especially about the information resources they access—both academic and otherwise. The desire to utilize this "big data" to better understand student (and to a lesser extent, faculty) behavior fits within larger trends in higher education that are placing greater emphasis on analytical methods that seek to model students' educational experiences using data from universities' various systems for keeping track of their students, including (but certainly not limited to) data about student demographics, course outcomes, academic preparedness, attainment, persistence and retention, financial situation, and engagement, as well as behavioral information from learning management systems. Given the importance of library services and materials to the academic mission of the university, as well as the position of the library as a major cost center, it is likely that libraries and librarians will continue to encounter interest and pressure to link data about their users to other institutional datasets.

While the field of learning analytics in libraries is still relatively new, a growing number of studies have demonstrated the potential utility of these types of analyses by exploring links between library use and student attainment (Stone & Ramsden 2013; Stone, Pattern, & Ramsden 2012; Goodall & Pattern 2011) and retention (Allison 2015; Stemmer & Mahan 2016; Soria, Fransen & Nackerud 2015; Haddow 2013; Haddow & Joseph 2013; Soria, Fransen & Nackerud 2013), as well as the correlations between various types of library use and GPA (Allison 2015; Stemmer & Mahan 2016; Kot & Jones 2015; Soria, Fransen & Nackerud 2015; Nackerud, et. al 2013, Soria, Fransen & Nackerud 2013; Wong & Webb 2011 Cox & Jantti 2012). These studies have also helped spark a debate among librarians surrounding the analysis of library data and its implications on the privacy of library constituents. Utilizing library use data for analytical purposes presents an ethical dilemma for many librarians, who have traditionally considered this data to be confidential and sacrosanct, and that protecting users' right to privacy comprises one of the pillars of librarianship of a profession.

The American Library Association (ALA) enshrines the confidentiality in principal III of its code of ethics: "We protect each library user's right to privacy and confidentiality with respect to

information sought or received and resources consulted, borrowed, acquired or transmitted."[1] However, principal I of the ALA code of ethics obligates librarians to "provide the highest level of service to all library users through appropriate and usefully organized resources"--an obligation that is often aided by the analysis of library data.

This tension between to ethical imperatives of providing high-quality access and services and protecting the privacy and confidentiality of users is at the core of librarians' relationship with the analysis of user data. Collecting user data for research and analysis purposes rather than the necessity of administering library systems represents a shift in many libraries' stances toward user privacy, and is often a relatively new capability for many libraries and librarians. While libraries' technical capabilities to collect usage data have expanded rapidly with new digital systems and analytical tools, these abilities have sometimes seemed to outpace the ethical conversations surrounding this type of data collection.

In this chapter, I will address this debate by examining the ethical implications of tracking usage data across libraries' online systems, and will argue that librarians should approach the collection and analysis of library data from a research ethics point of view, and should utilize a human subjects research model for evaluating the efficacy of these types of "big data" studies.

**"Big" usage data in libraries**

As library collections and usage have shifted from principally physical to principally digital materials, the systems that enabled online access to information simultaneously increased libraries' ability to observe and collect user data. Libraries now have the practical ability to systematically collect transaction-level information that link individuals with their item-level usage. Creating a detailed dataset containing every online resource used by every library user is probably within the technical capacity of most libraries and much of this data may already exist--even if it is not presently used--since it is prerequisite for authentication systems (see see Dixon 2008 and Coombs 2005). For example, EZProxy logs are set up by default to collect IP addresses and usernames.[2] With some effort, it is conceivable that a library could expand such a dataset to include essentially every interaction that a user has with library systems: every set of search terms, every record viewed, every article browsed or downloaded, and every book checked out by every library user.

Once individual identifiers such as a username are collected, linking library data to other institutional data such as demographic information, financial aid, measures of student and faculty success (engagement, retention, GPA, grant funding, publication records), and anything else that a university might track at an individual level is fairly trivial, and likely more a matter of institutional policy rather than technical capacity. Indeed, attempting to answer research

[1]See http://www.ala.org/advocacy/proethics/codeofethics/codeethics. The International Federation of Library Associations and Institutions (IFLA) Code of ethics (Part 3) takes an even more restrictive stance to user data stating: Librarians and other information workers respect personal privacy, and the protection of personal data, necessarily shared between individuals and institutions. The relationship between the library and the user is one of confidentiality and librarians and other information workers will take appropriate measures to ensure that user data is not shared beyond the original transaction. However the same tension between privacy and access that can be observed in the ALA code exists between Part 1 and Part 3 of the IFLA code. See http://www.ifla.org/news/ifla-code-of-ethics-for-librarians-and-other-information-workers-full-version#accesstoinformation

[2] See https://www.oclc.org/support/services/ezproxy/documentation/cfg/logformat.en.html

questions about the impact of libraries on student or faculty outcomes depends on utilizing unique identifiers and collecting data in a way to allow it to be associated with individuals.

A library dataset of this type would obviously contain extensive personal information about individual users and the topics that they are interested in. In addition to collecting user data in in accordance with ethical standards librarianship, in my opinion librarians also have a responsibility to collect this data in accordance with standards for human subjects research, and are therefore obligated to conduct ethical due diligence before datasets are created. At educational institutions in the United States, ethical review of human subject research is usually conducted by Institutional Review Boards (IRB), which are charged with assessing whether or not data collection practices and research protocols meet the requirements of three guiding principles: respect for persons, which establishes the requirements of disclosure of the data being collected and informed consent, beneficence, which asserts the research should first do no harm and second should maximize possible benefits while minimizing possible risks, and justice, which states that the selection of research subjects should be equitable (see Schrag 2010:87-88).[3]

While not all library data collection will fall under the auspices of IRBs,[4] the ethical principles on which IRB oversight is based provide a useful framework for discussing the efficacy of data collection in libraries. Unfortunately, the creation and utilization of the type library usage data described above potentially fails to meet all three of the ethical standards of beneficence, respect for persons, and justice

**Considering Risk and Benefits**

The potential benefits to analyzing large library usage datasets are numerous. Along with better understanding the relationship between student success measures and personal library use, these data may also help libraries provide better services to students, faculty, and public users, make more efficient use of funds, improve collections management, demonstrate the role of libraries in student learning and faculty research, and identify students at risk of dropping out or in need of additional help. All of these are important and laudable goals driven by a desire to improve educational experiences and practices.

Nevertheless, the push to collect ever more detailed and fine-grained educational data over longer periods of time makes weighing benefits and risks more difficult, since so much of how

[3] See 45 CFR 46 (Protection of Human Subjects 2009); http://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/

[4] The rules and regulations covering library data collection are multiple and complex. Depending on a particular university's interpretation of the regulations, data collection via library systems may not be considered human subjects research and therefore not subject to IRB review. This determination will depend on a particular institution's stance on whether or not such data is "private" (i.e. the person has a reasonable expectation of privacy), and whether or not the collection and analysis of this data contributes to "generalizable" knowledge. In the case of student users, library data in the US may also subject to The Family Educational Rights and Privacy Act (FERPA) regulations if the institution considers them educational records (See Jones & Arnold 2014 for a review of data ownership at universities). In my opinion, the collection and analysis of library use data is a type of social and behavioral research that is very similar to other research methods that would normally subject to the review and oversight of IRBs. Regardless of local review practices, a lack of an ethical review requirement does not relieve librarians and other researchers' obligation to collect data ethically, responsibly and respectfully (See Schrag 2010 for a detailed discussion of the history and controversy surrounding IRB regulations).

the data might be used (or misused) is unknown at the time of collection. Indeed, this potential is one of the reasons for the present excitement about learning analytics and "big" educational data.

Given that educational outcomes must usually be known in order to draw meaningful conclusions from library use data, the time horizons required mean that individuals who are providing the data rarely benefit directly from their findings. Although an absence of direct benefit to a research participant certainly does not preclude conducting research, when weighing risk and reward, researchers should carefully consider designs in which the people bearing the risk do not reap the benefits, particularly with regard to whose interests are actually being served.

As Rubel & Jones point out, when aggregated, data obtained from individuals is typically far more valuable to institutions than the subject themselves (2006:147). Particularly in the context of libraries, researchers must also be very aware of the financial and ideological interests at play (Watters 2013). In addition to universities, many of the parties likely to incur the benefits of library data analyses are third party service providers, including for-profit publishers and other corporations, whose primary responsibility is to their shareholders rather than students.

The use of data by a variety of institutional actors may, in fact, result in improved educational services, systems, and experiences for students and faculty in general, but the ethical question at hand is whether or not individuals have a right to some level of control and autonomy over their personal data from which this value is derived. Furthermore, it is important to question whether outcomes justify the data collected and if benefits outweigh students' loss of autonomy over how data about them is collected (Rubel & Jones 2016:148).

While continuing to evaluate the ways multiple parties benefit from the collection of library data is critical to conducting ethical research, many libraries, universities, researchers have taken the position that the indirect benefit to students of improving educational approaches justifies data collection--as demonstrated by current investment in learning analytics systems across the educational sector. However, it is equally important to evaluate the long-term risks that transaction-level datasets containing large numbers of individuals represent, and how it might be reused or misused by unexpected actors, including commercial, governmental, or law enforcement interests. One principal risk to library use data is the disclosure of identifying information.

Perhaps the most probable disclosure scenario for library use data is at the request of state actors such as law enforcement or regulatory agencies.[5] Once a dataset exists it is subject to subpoena by law enforcement (IRB consent forms usually require a warning about legally required disclosure). There is no researcher-subject privilege and we should not assume that universities will be willing or able to resist a subpoena, or that researchers can guarantee the confidentiality of research subject in such a scenario. Beyond criminal proceedings, it is not difficult to imagine cases in which library data might become of significant interest to a civil suit or other proceeding (e.g. copyright infringement proceedings), making an investigatory route to data disclosure fairly likely.

[5] While it is important not to overlook other potentially malicious actors, library data is likely difficult to monetize when compared to information such as account numbers or social security information (although there is a black market in library credentials). This probably makes data theft relatively unlikely.

For example, during the Boston College Belfast Project, oral history interviews were collected from former members of the Irish Republican Army who fought in Northern Ireland. These recordings were intended to remain secret until after a research participant's death, but after their existence was mentioned in a secondary source, they were subpoenaed in 2011 by the US Department of Justice at the request of the British Government as part of a murder investigation (McMurtrie 2014) The ensuing legal proceedings amply demonstrated the potential risks of archived datasets to institutions, the people providing the data, and to third parties referenced by the data.[6]

Although the Belfast Project was not a "big data" type dataset, and represents information that is probably much more sensitive and risky than the user data that libraries typically create and handle on a day-to-day basis, the FBI and other investigative agencies have shown a long-term and well-documented interest in obtaining library usage data (see ALA n.d.; Starr 2004). It is probably safe to assume that these practices will continue, and that any data created and held by libraries might become the target of an investigatory request. More chillingly, the post-9/11 practice using national security letters potentially prevents libraries and librarians from even disclosing the existence of a request (Peterson 2014).

In the case of actors with subpoena powers, there is little effective way to prevent disclosure of information except not to create the data in the first place, since law enforcement agencies can also compel disclosure of data protection mechanisms such as encryption keys. Because of their size and the amount of information they contain, large and comprehensive datasets containing many individuals therefore represent a significant risk for unintended use and disclosure. Given that individuals bear most of this disclosure risk, while institutions gain most of the benefit from the data's use, it is paramount to consider whether the potential outcomes derived from library use data really justify creating a dataset that is so inherently risky.

## Obtaining Consent

The difficulty in predicting the long-term uses of transaction-level or user-level library data create significant problems for the informed consent process, especially surrounding data retention and reuse.

Currently, many library practices for disclosing the data they collect and obtaining informed consent are insufficient from a research ethics standpoint, and are often written more with an eye toward ensuring efficient functioning of library systems than research and data analysis activities. Libraries routinely obtain passive consent for user data collection through mechanisms such as by posting a library privacy policy. These policies are sometimes out of date or incomplete, or do not adequately and fully explain what data is collected, how it will be used, and what other data it might be linked to. These types of policies are an insufficient consent instrument from a research ethics perspective not only for these reasons, but also because they are not widely read or understood by library users. Libraries also rarely have an effective opt-out procedure to their privacy and data collection policies except for users to refrain from using the library systems that require authentication. While this is the form of passive consent most web services use for their data collection (via their terms of service agreements), human subjects regulations in the United States require that researchers ensure not only that

[6] See https://bostoncollegesubpoena.wordpress.com/ for a comprehensive history.

consent is sought from each participant, but also that a prospective research subject understands the information. Passive consent is inadequate for meeting this requirement.

Moreover, in order to meet the ethical standard of respect for persons, consent must be voluntary, and potential subjects must have the opportunity to opt out. When data collection is systemic and potentially built into library systems, there is essentially no effective opt out option. Since using library systems is (at least in theory) vital to a successful academic career, forcing users into a situation whereby the only way to avoid participation in data collection and research is to refrain from using library systems and resources, a coercive situation is created because there is no adequate substitute. In this scenario, one might argue that opting out of participation actually does harm to the individual.

**Supporting Justice**

At first glance, comprehensive and systematic collection of library data might seem to meet the ethical standard of justice that requires equitable selection of participants since all library users would be included in the data collection. Nevertheless, there are a number of potential pitfalls in this area as well.

Due to their smaller numbers, minority groups of any type (e.g. race, ethnicity, gender expression, LGBT identity, religious affiliation) will always be more visible within a systematically collected dataset, and by extension, individual members of these groups will be easier to potentially identify. As discussed above, since the interests of individual students and their institutions are not necessarily aligned, imperatives to increase measures of student success may create incentives for institutions to utilize behavioral, performance, and demographic data to guide decisions about recruitment or advising—perhaps even consciously or inadvertently recruiting students whose demographic profile appears more likely to succeed or steering students in a particular direction of study where they appear likely to perform well (Rubel & Jones 2016:143). Since such data and decisions are often related to student demographics in complex and subtle ways, these processes can produce the effect of discrimination against particular groups, and there is a growing body of evidence suggesting that already socially and economically underprivileged and marginalized groups are much more likely to bear the brunt of this type of data-led discrimination (for a small selection of examples see Schrage 2014; Croll 2012; Borocas & Selbst 2015; Tene and Polonetsky 2013; Robertson & Travaglia 2016).

This reality makes a strong case for limiting the extent of information that is collected and how it is linked together. However, since it is difficult to identify in advance what variables will be related to student outcomes, many institutions are taking a much more expansive view to data collection. Rubel & Jones note "because we cannot know *a priori* what information will shed light on learning environments and outcomes, any information would seem fair game on the relevance view" (2016:151). In response, they suggest that "a good principle for whether it is justifiable to collect information is whether the universities would be justified in intervening to change the behavior captured in the data" (Rubel & Jones 2016:152). In the case of libraries, in would probably be justifiable for a university to intervene and suggest that a student who has

never used the library might benefit from doing so, but it probably would not be justifiable in intervening in what specific materials and resources the student has used.

**Data Practices for Libraries**

Given that there is risk associated both with collecting and with not collecting library data, how should libraries respond to increasing pressures to participate in learning analytics initiatives and the collection and analysis of usage data?

First, consent procedures should be reviewed before data collection, and whenever possible libraries should move to obtaining explicit rather than passive consent and opt-in rather than opt-out models for data collection. Before collecting data, libraries and librarians should also carefully consider the associated risks and the "worst case scenario" of the data's release. Once data is created, the assumption should be that it will eventually be compromised, and at a minimum standard data security procedures should be put in place, such as physically separating subject identification data from analysis data sets, and locally encrypting data files. Libraries should apply the social science adage not to collect data that you do not have a specific plan to analyze and use, and should adopt data collection strategies the help minimize the risk to participants should the data be disclosed.

In general, transaction-level data that uniquely identifies both a user and an item should be avoided unless required for a specific and limited purpose. Because of its inherent risk to privacy, systematic collection of this data should also be avoided. Data collected should be aggregated at a level that balances analytical specificity with user privacy. For example, electronic usage data might be collected at the database level rather than the item level, or circulation data at the LC classification level. User groups can also be aggregated—for example grouping sets of similar majors--to help minimize the risk of identifying individuals based on demographic information. Hanson, Nackerud & Jensen (2008) suggest aggregating to groups no smaller than 15-20 members, but given that reidentification techniques are likely to continue to become more sophisticated, researchers might consider using an even higher threshold for aggregation. In this way, aggregating at the highest level necessary to conduct a particular analysis can help mitigate risk if the data is eventually exposed.

Researchers should also consider the level of demographic data that is actually required for their analyses since each additional data point aides in attempts to identify participants. For example, demographic information such as gender and ethnicity that is commonly included in datasets might be spurious to analysis of relationships between library use and measures of student success, and might be dropped if no relationship is observed.

Removing identifying information from stored datasets is often used as a strategy to mitigate or eliminate risks from unintended disclosure of this type of data. Unfortunately, even de-identifying a dataset by destroying the links between individually identifying information and other data is possibly insufficient to protect research subjects' privacy. Re-identification of research subjects from information contained in datasets is decreasing in difficulty, a task that would be made even easier in the case of a library user dataset since the source population would be known (i.e. the students enrolled at a university during particular dates). For this reason, a dataset containing even rudimentary demographic data about students, (such as major, graduation year, sex, ethnicity, etc.), might be impossible to de-identify ((see also Tene and

Polonetsky's (2003:11) discussion of "obscurity" and Khalil & Ebner's (2016) analysis of de-identification as applied to learning analytics).

Due to the difficulty of securing "big data" and the legal and ethical risks of disclosure, Bruce Schneier (2016) argues that it should be thought of as a "toxic asset" and treated accordingly. The best way to protect risky data is not to create it. The second best way is to destroy it as soon as possible after analysis is complete.

Data destruction plans are therefore an integral component to securely managing privacy and minimizing risks to research subjects. All studies should include a data destruction plan, or a justification of why the data should be retained and for what purpose. Any datasets containing user demographic data or other identifying information should not be retained indefinitely and should be destroyed after a reasonable period following the completion of data analysis.

Finally, libraries should advocate for their universities to adopt a code of practice for data related to learning analytics.[7] In the absence of an institutional code of practice, libraries should develop their own. In this way, libraries have the opportunity to become centers of expertise and best practices for collecting and using institutional data, much as they have in other areas of scholarly communication such as copyright and research data management. Given librarianship's ethnical emphasis on ensuring the privacy of its users, this is a natural—and indeed vital—role for librarians to play as universities continue to develop their learning analytics programs.

**Works Cited**


Allison, DeeAnn. "Measuring the Academic Impact of Libraries." *Portal: Libraries and the Academy* 15, no. 1 (2015): 29–40.

American Library Association (ALA). "ALA | FBI in Your Library." Accessed June 11, 2016. http://www.ala.org/Template.cfm?Section=Intellectual_Freedom_Issues&Template=/ContentManagement/ContentDisplay.cfm&ContentID=21662.

Barocas, Solon, and Andrew D. Selbst. "Big Data's Disparate Impact." SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, August 14, 2015. http://papers.ssrn.com/abstract=2477899.

Coombs, Karen A. "Protecting User Privacy in the Age of Digital Libraries." *Computers in Libraries* 25, no. 6 (2005): 16–20.

Cox, Brian, and Margie Jantti. "Discovering the Impact of Library Use and Student Performance." *Educause Review*, 2012.

Crawford, Gregory A. "The Academic Library and Student Retention and Graduation: An Exploratory Study." *Portal: Libraries and the Academy* 15, no. 1 (2015): 41–57.


[7] For example, see the JISC Code of Practice for Learning Analytics: https://www.jisc.ac.uk/guides/code-of-practice-for-learning-analytics

Croll, Alistair. “Big Data Is Our Generation’s Civil Rights Issue, and We Don’t Know It – Solve for Interesting.” Accessed June 11, 2016. http://solveforinteresting.com/big-data-is-our-generations-civil-rights-issue-and-we-dont-know-it/.

Dixon, Pam. “Ethical Issues Implicit in Library Authentication and Access Management: Risks and Best Practices.” *Journal of Library Administration* 47, no. 3–4 (2008): 141–162.

Emmons, Mark, and Frances C. Wilkinson. “The Academic Library Impact on Student Persistence.” *College & Research Libraries* 72, no. 2 (2011): 128–149.

Goodall, Deborah, and David Pattern. “Academic Library Non/low Use and Undergraduate Student Achievement: A Preliminary Report of Research in Progress.” *Library Management* 32, no. 3 (2011): 159–170.

Haddow, Gaby. “Academic Library Use and Student Retention: A Quantitative Analysis.” *Library & Information Science Research* 35, no. 2 (2013): 127–136.

Haddow, Gaby, and Jayanthi Joseph. “Loans, Logins, and Lasting the Course: Academic Library Use and Student Retention.” *Australian Academic & Research Libraries* 41, no. 4 (2010): 233–244.

Hanson, Cody, Shane Nackerud, and Kristi Jensen. “Affinity Strings: Enterprise Data for Resource Recommendations,” 2008. http://conservancy.umn.edu/handle/11299/46576.

Jerome, Joseph. “Buying and Selling Privacy: Big Data’s Different Burdens and Benefits.” *Stanford Law Review Online*, 2013. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2294996.

Johnson, Jeffrey Allen. “How Data Does Political Things: The Processes of Encoding and Decoding Data Are Never Neutral.” *Impact of Social Sciences*, October 7, 2015. http://blogs.lse.ac.uk/impactofsocialsciences/2015/10/07/how-data-does-political-things/.

Khalil, Mohammad, and Martin Ebner. “De-Identification in Learning Analytics.” *Journal of Learning Analytics* 3, no. 1 (2016): 129–138.

Kot, Felly Chiteng, and Jennifer L. Jones. “The Impact of Library Resource Utilization on Undergraduate Students’ Academic Performance: A Propensity Score Matching Design.” *College & Research Libraries* 76, no. 5 (July 1, 2015): 566–86. doi:10.5860/crl.76.5.566.

McMurtrie, Betrh. “Secrets from Belfast.” *The Chronicle of Higher Education*, (Jan. 26, 2014).

Mezick, Elizabeth M. “Return on Investment: Libraries and Student Retention.” *The Journal of Academic Librarianship* 33, no. 5 (2007): 561–566.

Nackerud, Shane, Jan Fransen, Kate Peterson, and Kristen Mastel. “Analyzing Demographics: Assessing Library Use across the Institution.” *Portal: Libraries and the Academy* 13, no. 2 (2013): 131–145.

Oakleaf, Megan. “Value of Academic Libraries: A Comprehensive Research Review and Report.” Association of College and Research Libraries. 2010. http://www.ala.org/acrl/sites/ala.org.acrl/files/content/issues/value/val_report.pdf

Peterson, Andrea. “Librarians Won’t Stay Quiet about Government Surveillance.” *Washington Post*, 2014. https://www.washingtonpost.com/news/the-switch/wp/2014/10/03/librarians-wont-stay-quiet-about-government-surveillance/.

Robertson, Hamish, and Joanne Travaglia. “Big Data Problems We Face Today Can Be Traced to the Social Ordering Practices of the 19th Century.” *Impact of Social Sciences*, October 13, 2015. http://blogs.lse.ac.uk/impactofsocialsciences/2015/10/13/ideological-inheritances-in-the-data-revolution/.

Rubel, Alan, and Kyle M. L. Jones. “Student Privacy in Learning Analytics: An Information Ethics Perspective.” *The Information Society* 32, no. 2 (March 14, 2016): 143–59. doi:10.1080/01972243.2016.1130502.

Schneier, Bruce. “Data Is a Toxic Asset - Schneier on Security.” Accessed March 16, 2016. https://www.schneier.com/blog/archives/2016/03/data_is_a_toxic.html.

Schrag, Zachary M. *Ethical Imperialism: Institutional Review Boards and the Social Sciences, 1965–2009*. JHU Press, 2010.

Schrage, Michael. “Big Data’s Dangerous New Era of Discrimination.” *Harvard Business Review*, January 29, 2014. https://hbr.org/2014/01/big-datas-dangerous-new-era-of-discrimination.

Soria, Krista M., Jan Fransen, and Shane Nackerud. “Library Use and Undergraduate Student Outcomes: New Evidence for Students’ Retention and Academic Success.” *Portal: Libraries and the Academy* 13, no. 2 (2013): 147–164.

———. “Stacks, Serials, Search Engines, and Students’ Success: First-Year Undergraduate Students’ Library Use, Academic Achievement, and Retention.” *The Journal of Academic Librarianship* 40, no. 1 (2014): 84–91.

Starr, Joan. “Libraries and National Security: An Historical Review.” *First Monday* 9, no. 12 (December 6, 2004). http://firstmonday.org/ojs/index.php/fm/article/view/1198.

Stemmer, John K., and David M. Mahan. “Investigating the Relationship of Library Usage to Student Outcomes.” *College & Research Libraries* 77, no. 3 (May 1, 2016): 359–75. doi:10.5860/crl.77.3.359.

Stone, Graham, David Pattern, and Bryony Ramsden. “Library Impact Data Project.” *Sconul Focus*, no. 54 (2012): 25–28.

Stone, Graham, and Bryony Ramsden. “Library Impact Data Project: Looking for the Link between Library Usage and Student Attainment.” *College & Research Libraries* 74, no. 6 (November 1, 2013): 546–59. doi:10.5860/crl12-406.

Tene, Omer, and Jules Polonetsky. “Big Data for All: Privacy and User Control in the Age of Analytics.” *Nw. J. Tech. & Intell. Prop.* 11 (2012): xxvii.

Watters, Audrey. “Student Data Is the New Oil: MOOCs, Metaphor, and Money.” *Hack Education*, October 17, 2013. http://hackeducation.com/2013/10/17/student-data-is-the-new-oil.

Wong, Shun Han Rebekah, and T. D. Webb. “Uncovering Meaningful Correlation between Student Academic Performance and Library Material Usage.” *College and Research Libraries* 72, no. 4 (2011): 361.