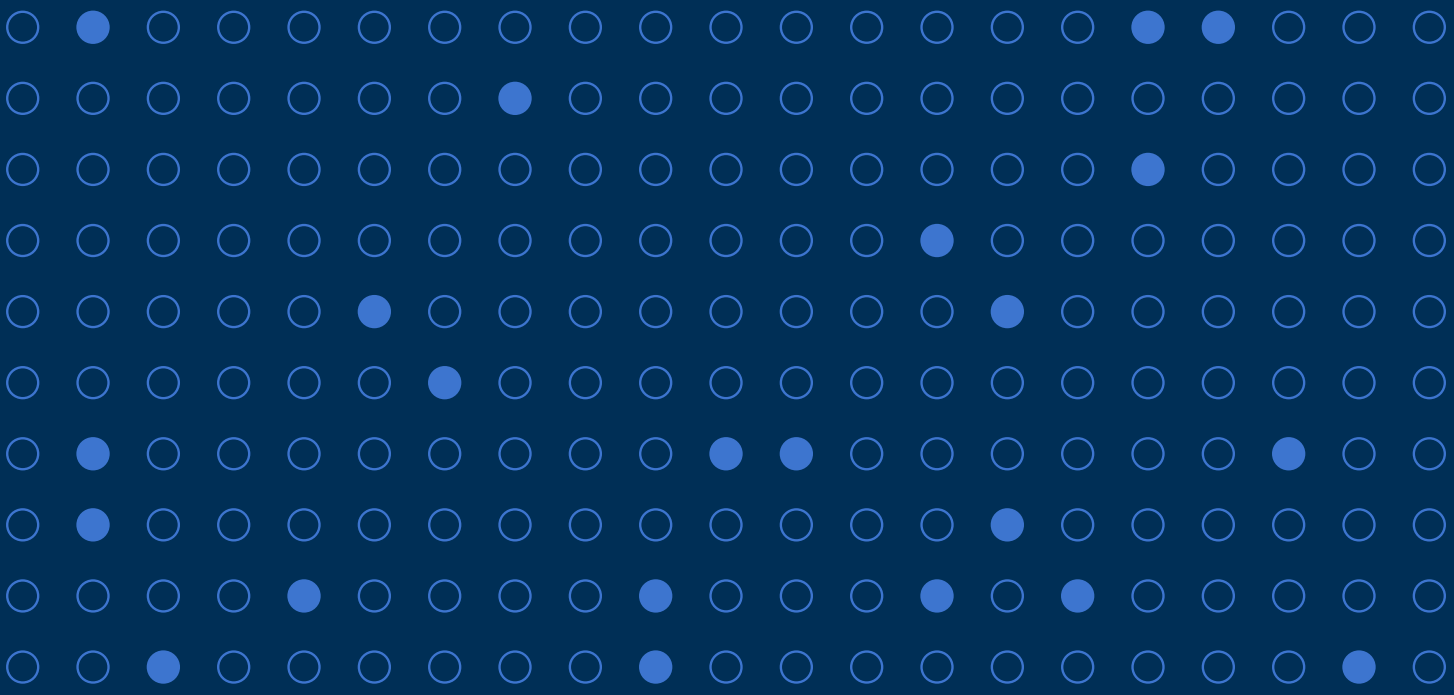
Library Analytics — Proving Value While Maintaining Privacy

With shrinking budgets, staff cuts, and the need for academic libraries to show their contribution to the overall academic mission, libraries are in a crucial position to invest in sound analytics and data. This type of data, however, needs to extend past traditional library metrics to demonstrate the synergy between the library and the entire academic campus. Of course, this means gaining an understanding of users on campus, how they leverage the library and what criteria create an academically successful cohort vs. "at-risk". Thus, an opposing force has been created between leveraging library data to make evidence-based decisions and maintaining library user privacy.

This white paper aims to explore the topic of privacy in library analytics while presenting best practices and solutions which can help a library improve its library analytics approach and maintain proper data security standards to uphold the privacy of the end-user.



The History of Privacy in the 20th Century — Highlights of Important Dates

Understanding the issue of data privacy within the library world means taking a glimpse at the history of data privacy across the globe. Legislation, literature, and other articles have shaped the vision of data privacy across all sectors.

While the issue of privacy can be traced back to the US Constitution, below is a timeline of global highlights relating to privacy in the 20th century. [1]

- **1914: Establishment of the FTC**

The leading federal agency in the US often involved with privacy issues.

- **1948: United Nations Declarations of Human Rights, Article 12**

States that "No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks."

- **1974: FERPA**

FERPA, the Family Educational Rights and Privacy Act protects the privacy of student education records and applies to any educational institution that receives funds from the U.S. Department of Education.

- **1995: EU Protection Directive**

Regulates the processing of personal data within the EU. Would later be superseded by the EU Right to be Forgotten in 2012.

- **1996: HIPPA**

Health Insurance Portability and Accountability Act was created to protect Personally Identifiable Information maintained by the healthcare and health insurance industries from theft and fraud in addition to other regulatory issues.

- **1999: Gramm Leach Bliley Act**

Requires financial institutions to explain how customer data is shared.

- **2003: State Data Breach Notification Laws**

California was the first state to implement data breach notification laws.

- **2018: GDPR**

The General Data Protection Regulation deals with data protection and privacy for the European Union (EU) and the European Economic Area EEA). It also applies to the transfer of personal data outside of the EU and EEA.

What the timeline snapshot demonstrates is that a library must be committed to protecting user privacy and data. What's more, libraries need to implement proper policies, procedures, and infrastructure to handle big data security and continually evolve compliance efforts with data.

Big Data and Privacy. Where do We Start?

The term "big data" has become popular with all sectors of business and education. Big Data is defined as "extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human behavior and interactions".[2] If optimized correctly, big data can provide actionable insights and evidence-based decision-making.

When it comes to big data, libraries have arrived at a crossroads. Either continue the traditional role, focusing on common library metrics such as circulation, usage, etc., or embrace the big data mindset to evolve their position. Providing more data on interactions with the entire academic community can transform the library, not only into a major influencer on the trajectory of the academic institution but potentially as the central hub of academic metrics. So the million-dollar question is, will academic libraries "play a central role in the meta-fourth industrial revolution as central-information providers" or will they "keep a profile as organizations that continue to provide "traditional" services to patrons?"[3]

Extending the Library's Value: Leveraging Learning Analytics

Learning analytics combines data mining and processing along with integrated technologies and reporting capabilities to identify trends, holistically understand its impact on student success or to identify the academic "at-risk" students, an academic institution must gather and analyze a wealth of learning analytics data. Tracking student demographics and creating predictive modeling allows for deep insight into how all parts of the institution work together to influence academic success and demonstrates activities and interests that correlate with learning. [4]

Learning analytics can be used to improve several core educational components of the academic institution including identifying target courses, improving curriculum, improving instructor performance, and post-educational employment (to name a few). But the shared interest for the library and the institution is student learning outcome, behavior, and process. [4]

Students: Learning Behavior and Process

Improving how students learn and conduct research is in part dependent on how skilled they are in information literacy. Information literacy can be defined as the competencies an individual summons to locate, retrieve, evaluate, select, and use information resources. [5] It is quite common for libraries to spearhead information literacy instruction and for librarians to become the facilitators of learning. Therefore, having impactful data that correlates library usage with student success: not only demonstrates the importance of the library in academic life, but solidifies the librarian(s) as the stewards and builders of a very important foundation for learning.

For libraries, this evolutionary data approach has a two-fold challenge. One, of course, is this apparent issue around privacy. For librarians there is an ethical responsibility to protect patron privacy and to reconcile that with a need for detailed user engagement metrics. The other is the challenge of automating and streamlining secure access to many data sets. Often, librarians are faced with accessing and gathering data from disparate platforms and manually combining data for analysis. While this approach was somewhat sustainable with traditional library metric reporting, it cannot scale with additional data sets that are needed to provide a holistic view of the library's influence. Not only is the current manual process of data analysis painstaking, but it also increases the risk of data loss and more importantly the possibility of a data breach.

Upholding Privacy

A 2020 article in Security Boulevard underscores the growing need to protect Personally Identifiable Information within higher education:

> *COVID-19 has been something of a perfect storm for colleges and universities. As we uncovered in our recent Education Cybersecurity Threat Index, the pandemic has made addressing cybersecurity weaknesses an urgent operational necessity in higher education. Between the changing nature of how education is delivered, and emboldened threat actors, higher ed now faces an unprecedented threat level.* [6]

PII has obvious attributes that can easily be identified and agreed upon by your library and technology teams. Surprisingly, though, some attributes of PII are broad and often left to be interpreted. For example, some of the major institutions that care about PII have varying definitions of what these attributes are:

Put simply, Personally Identifiable Information (PII) is any information that could identify a person such as an email, address telephone number, etc. In a more complex way, PIII can be a combination of data points such as race, geographic location, and birthday that when used together can identify an individual. [7]

Any information relating to an identified or identifiable natural person ("Data Subject"); an identifiable person is one who can be identified, directly or indirectly, by reference to an identification number or one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.

The Department of International Law of the Organization of American States defines PII as… "all personal information of any kind […] to any kind of data, whether or not private that may affect rights if used by data processors.

European Union, Data Protection Directive defines PII as:
How PII is defined is also varied by country. [8] With data security and PII as a major concern within the academic institution, how can a library leverage analytics while simultaneously protecting PII / student information?

Proving Value While Maintaining Privacy

Without a clear definition of PII attributes, the first step for many libraries is to align their privacy attributes to their institution's definition of PII and to ensure their privacy policies around data also comply with global, country, and sometimes even regional (or state) privacy laws.

For example, when one academic library partnered with EBSCO Information Services (EBSCO) as an alpha partner for the library analytics platform, **Panorama™**, library staff worked with EBSCO's director of technology to understand what PII characteristics may not be loaded into the platform. An example of a student's PII is their IP address. However, according to most laws, including CCPA regulations on IP addresses, the IP address is not PII if it cannot be linked directly to a household.

Of course, this example demonstrates that it is beneficial for the library to partner with the IT or data security department within the academic institution to establish a framework and define attributes PII. Additionally, the library must leverage not only the institution's privacy policy but overall privacy laws as well as **guidelines** established by library associations.

Conclusion

Protecting privacy with big data must go beyond traditional protocols. With the amount of data that can be leveraged, and regulations to uphold, libraries need to consider how to scale today and in the future. One way to scale is finding the right library analytics platform. Below are some key elements to consider when building out the right library analytics strategy and when leveraging the right analytics platform.

Automation: Who owns data and how it is stored varies within each platform, content type, or institution. For example, there may be data owned by the institution but not shared with the library (and vice versa). Or library data may be contained within third-party vendor platforms, stored in various formats, and shared with various employees. [10] The lack of standardization can certainly jeopardize data security. But it also leaves the library unable to meet the demands of big data analysis and lose at the game of scale. The automation of library data creates a more sustainable practice, one that meets security requirements, but also saves time and effort for staff and allows for scalability.

Cloud Technology: Leveraging cloud technology within an analytics platform allows for several benefits. Of course, the obvious is centralizing and streamlining data. But cloud technology also reduces human error, improves security and privacy of data, and ensures the backup of data is available no matter your system failures. Simply put, cloud technology acts as an extension of your local IT capabilities.

Visualization Tools: The inherent problem with big data is of course its size. Leveraging visualization tools that can present data in a digestible format provides faster identification of trends and easier reporting capabilities. A simplified representation of data can be used to showcase the library's value to institutional stakeholders.

Going With the Right Vendor: With all this sound advice, it leaves a library to look for the most effective path to balancing big data analytics and user privacy. That is why it is important to look to vendors that not only have analytics technology to support the above points but take a holistic approach to safeguarding security and user data protection. One such application is the (International Standards Organization) ISO 27001 certification, which is attained through a rigorous audit and assessment by an independent third party of company processes and the security of its products and services. When considering a library analytics platform, consider working with a vendor who has received this type of certification. Today and in the future, librarians need a clear picture of the library's landscape across collections, services, operations, and academic influence. Implementing the right library analytics platform can provide more advanced metrics with greater context in a secure environment.

Panorama™

Insights Through Analytics

Learn more about **Panorama**, EBSCO's next-generation library analytics platform.

Learn More

Works Cited

1. **https://safecomputing.umich.edu/privacy/history-of-privacy-timeline**
2. Prindle, Sarah, and Amber Loos. "Information Ethics and Academic Libraries: Data Privacy in the Era of Big Data." Journal of Information Ethics, vol. 26, no. 2, Fall 2017, pp. 22–33.
3. Garoufallou, Emmanouel, and Panorea Gaitanou. "Big Data: Opportunities and Challenges in Libraries, a Systematic Literature Review." College & Research Libraries, vol. 82, no. 3, May 2021, pp. 410–435
4. Avella, John T., et al. "Learning Analytics Methods, Benefits, and Challenges in Higher Education: A Systematic Literature Review." Online Learning Journal (OLJ), vol. 20, no. 2, June 2016, p. 13.
5. David Bawden. "Information and Digital Literacies: a Review of Concepts." Journal of Documentation, vol. 57, no. 2, Apr. 2001, pp. 218–259
6. **https://securityboulevard.com/2020/12/why-higher-education-is-a-prime-target-for-cybercriminals/**
7. **https://www.law.cornell.edu/cfr/text/2/200.79**
8. Cherie L. Givens. Information Privacy Fundamentals for Librarians and Information Professionals. Rowman & Littlefield Publishers, 2015.
9. **https://iapp.org/news/a/are-ip-addresses-personal-information-under-ccpa/**
10. **https://er.educause.edu/articles/2017/8/the-academic-library-and-the-promise-of-ngdle**