

# Smart Libraries Newsletter

News and Analysis in Library Technology Developments

 ALATechSource  
alatechsource.org

225 N. Michigan Ave., Suite 1300, Chicago, Illinois 60601-7616, USA



## Smarter Libraries through Technology Ethical Analytics

By Marshall Breeding

Libraries, like most other organizations, face increasing pressures to strengthen their impact and continually optimize how they allocate resources. Collecting data, detecting patterns, and creating visualizations that can be readily interpreted have become important tools for managers. Organizations increasingly value data-driven decision making and are investing in tools to provide those capabilities. The possibilities for collecting and exploiting abound and are used aggressively in the commercial and social networking arenas. The pervasive drive to exploit data presents challenges for libraries that strive to use it to benefit their organizations, yet stay within bounds of their professional values and privacy policies.

Libraries have a long tradition of counting events and actions and producing statistics describing their collections and operations. Advancing into the realm of analytics represents a natural step forward. Technologies available today provide powerful tools for guiding the management of a library to shape its collections and services to optimally serve its constituents. Analytical tools can be incredibly useful to measure performance and to anticipate future outcomes. Daily operational decisions and organizational strategies can be formed through the patterns and trends seen in relevant data, presented via well-designed visualizations.

Library requirements for data management and analytical tools differ from other organizations, especially from those in the commercial sector, which take a more aggressive approach to personal data. Libraries have a strong interest in presenting personalized services to their users, but also impose strict limits on the use of personally identifiable information. These limits do not apply to all categories of data. The ethical framework for data privacy of library patrons differs from that of library workers and other sources of data generated through operational applications.

### Patron Privacy: A Basic Assumption

Libraries hold a high standard of privacy for their users. This approach to privacy enables patrons to access library materials confidentially, without concern that their information will be shared with any other individual or organization. Not only do libraries protect the personal details of their patrons more strictly than would be the case in a commercial context, but they also retain data related to patron interactions with collection items and services within strict parameters.

Operations require collection of some data during an active period of use, though on the completion of that transaction, all data are removed or anonymized. Patterns of data retention consistent with library ethics were initially established with physical items and the manual or automated circulation systems that manage lending transactions. Library lending involves connecting a record associated with the patron with a record associated with a collection item. Until the patron returns the item, an active link enables communication channels for messages such as overdue, fine, or recall notices.

Once that transaction is complete, any link between the patron and item record is released. Any logs are then sanitized to eliminate record transaction details that could be used to reconstruct the use of an item by the patron. Rather than deleting transaction record entirely, libraries generally replace any data elements that directly represents the patron with placeholders

### IN THIS ISSUE

**EBSCO Information Systems Introduces Panorama for Analytics**  
PAGE 4

**Smart Libraries Q&A**  
PAGE 6

that anonymously preserve significant categorical information needed for statistics and analysis.

These privacy enhanced practices likewise apply to digital transactions, despite the complications related to the many different platforms and networks involved, each of which has the opportunity to collect detailed data. Commercial providers of digital content may not share library patron privacy values. Amazon, for example, gains access to some data when libraries lend ebooks to patrons using Kindle readers. Despite these complications, libraries strive to enforce the confidentiality of patron transactions for their digital services.

## Other Organizational Data

Data related to tasks performed by library workers do not necessarily fall within the same level of protection as applies to patron transactions. Library workers do not expect anonymity as they perform their work on systems provided by the library or its parent institution. Integrated library systems, for example, will capture productivity statistics that can be used for assessment of individual workers or the institution. Library workers often use the email and other communications systems of their parent organizations, which do not necessarily guarantee employee privacy and may be subject to freedom of information act requests. Analytics tools used by libraries should be able to differentiate the privacy and legal frameworks associated with each category of data.

## Performance and Assessment

Libraries expect their business systems to generate data that can be used to measure performance and impact of their resources, including personnel, collections, and financial. Collection of data related to tasks performed within a system informs decisions related to the number of personnel needed to meet demands and the allocation of personnel according to changing workloads. Libraries especially benefit from data related to collections. On the back end, pricing and vendor data enable libraries to evaluate the value, performance, and competitiveness of their suppliers. Given constrained budgets, libraries must use all available means to build collections tuned to meet the information needs of their constituents and to make selection, renewal, and cancellation decisions based on the impact and cost effectiveness.

## Collection Use and Impact

Libraries also depend on data related to the use of collection items, provided these data operate within the confines of

privacy concerns. For physical and digital items, libraries depend on data describing the volume of use and the categories of users. The circulation systems of ILS or LSP products provide definitive data regarding the use of physical items. Measuring use of digital items is more complex and comes from diverse sources, such as publisher-supplied COUNTER statistics, proxy servers, or link resolvers. The key concern for libraries is that any data collected about resource usage describe only the categories of users accessing materials and never specific users.

## Reports: Basic Counts and Statistics

All library automation products offer some level of reports and statistics. Integrated library systems, for example, produce basic reports of daily, monthly, and annual circulation totals by categories, and breakdowns of new acquisitions by disciplines. These products come delivered with standard reports generated automatically and may also offer customized reports that can be programmed to create lists or tables addressing any category of data available in the system. The custom reporting capabilities of systems may include procedures to count or extract data using SQL queries or API tools. Reports provide simple representations of data but do not necessarily provide additional perspective or insights.

## Analytics: Beyond Statistics and Reports

In recent years, expectations have advanced beyond basic reporting and statistics to more sophisticated analytical tools. Analytics provide a more sophisticated set of tools to enable organizations to more fully explore any available data to answer operational or strategic questions. These tools are scaled to work with massively large data sets and can combine data from many categories and sources. They include more advanced computational capabilities that can extract patterns and trends that may not be apparent using standard reporting tools. Some may incorporate machine learning and other aspects of artificial intelligence to analyze large and complex data sets.

Data tables can be difficult to interpret, so most analytics packages create visual representations that summarize data and trends. These visualizations make complex data scenarios or computational analysis accessible to managers or administrators. An analytics package will enable users also to explore the data underlying any graph or visualization. It is also common to bring together several visualizations into a single page to form a dashboard that gives a broad overview of a variety of related activities. Dashboards serve as landing pages that lead into more detailed data summaries or graphs.

While reports address data within a single system, analytics can incorporate data from many disparate sources. In the library context, an analytics package can bring together multiple categories of data from the ILS, usage data from content providers, selected data from student information systems, census or other demographic data sets, building entrance counters, or Wi-Fi and other network traffic data.

## Underlying Technologies

Most library-oriented products rely on commercial analytics engines. General purpose analytics engines can be licensed as infrastructure components in the same way that an integrated library system would make use of relational database. Major options include Oracle Business Intelligence, Tableau, MicroStrategy, IBM Cognos Analytics, and SAP Analytics Cloud. Each analytics engine offers different capabilities, integration options, and data models. In most cases, which analytics engine a vendor chooses for its analytics infrastructure will be transparent to the library. Library-specific analytics modules or platforms provide front-end interfaces and data connectors and import tools. Libraries will want to pay careful attention to any data sharing or privacy terms associated with the analytics service.

*Analytical tools wield substantial power and can be deployed well within the bounds of library ethics to produce significant benefits.*

## Ethical Issues and Professional Values

Data warehouses and analytical engines have powerful capabilities. In a commercial environment, these tools can combine data from many sources to compile detailed profiles of preferences and spending for an individual consumer. These data in turn power advertising and ecommerce platforms to enhance revenue through highly personalized marketing techniques.

These technologies must be approached cautiously by libraries. Even when libraries follow standard approaches to anonymize data sets, they should also be careful to prevent the possibility of re-identifying transactions to personal identities through triangulation with other data sets. Advertising networks routinely associate use data with specific individuals even when the sources may not have definitive identifiers by techniques such as clueing off of network traffic data.

Data analytics enable many scenarios that may benefit library decision making, but that raise ethical questions. Historical data on the items borrowed or accessed digitally

enables multiple services, and patrons may expect to view such lists in their profiles. Libraries can offer patrons an option to retain this data, but the specific terms may not be consistently understood. Is it within library ethical practices to collect this data by default, with an option to opt out? Are library systems able to use this data in recommendation engines? If such non-anonymized data exists, it may be challenging to ensure that it is not shared beyond the specific boundaries defined in the terms of privacy statements.

Libraries should also be vigilant to ensure that analytics or data practices do not introduce bias or inequities into their information ecosystems. Strategies or intentions that libraries conceive to provide materials and services to vulnerable or underserved segments of their communities may not necessarily be addressed algorithmically. Recommendation services can also introduce unintended biases. Both in the collection and presentation of content items, libraries must oversee any analytic-driven processes to avoid bias. Principles of diversity and inclusion should be applied to the data, algorithms,

and analytics that shape the performance of library systems and platforms.

## Indispensable Technologies

These concerns do not mitigate the value of data and analytical tools for libraries. It is important for libraries to invest in the technologies able to help them manage their resources and carry out their missions. This genre of technology products wields substantial power and can be deployed well within the bounds of library ethics to produce significant benefits. Libraries that implement these tools need to be aware of any implications related to privacy and bias and configure their use accordingly.

## An Expanding Genre

Many of the major library technology vendors offer enhanced analytics packages integrated into their platforms or as optional product offerings.

- **Alma Analytics:** included as a built-in component of the Alma library services platform. <https://exlibrisgroup.com/products/alma-library-services-platform/alma-analytics>
- **BLUEcloud Analytics:** an optional component of the BLUEcloud suite that can work with either the Symphony

or Horizon ILS products. <https://www.sirsidynix.com/bluecloud-analytics>

- **CollectionHQ:** an analytics package oriented to library collection management, offered by a division of Baker & Taylor. <https://www.collectionhq.com>
- **Gale Analytics:** combines local ILS data with public demographic data sets. <https://www.gale.com/databases/gale-analytics>

- **OCLC WorldShare Report Designer:** provided with WorldShare Management Services or the Tipasa ILL workflow manager. <https://www.oclc.org/en/worldshare-report-designer.html>

Analytics offerings for libraries continues to expand. This issue of *Smart Libraries Newsletter* features the latest new offering, Panorama from EBSCO Information Services.

## EBSCO Information Systems Introduces Panorama for Analytics

Following a period of design and development, EBSCO Information Services has released Panorama. The new platform provides a diverse set of capabilities to help libraries manage each aspect of their operations using empirical evidence.

Panorama was created primarily for academic libraries. It provides insights on the performance and value of collections and services and the library's impact on their campus communities. EBSCO designed Panorama as a vendor-neutral analytics service.

Panorama extends EBSCO's portfolio of SaaS products, which also includes EBSCO Discovery Service, Stacks, EBSCO Knowledge Services, and Full Text Finder, as well its services for the FOLIO library services platform.

### Data Sources and Ingestion

EBSCO positions Panorama as a self-service analytics platform. One of the key concepts behind Panorama involves streamlining the ingestion of data, enabling library personnel to spend less time with the details of acquiring data and focus more on analysis. EBSCO has developed automated pipelines to feed data into the platform, including internal library sources, institutional data, and usage reports from suppliers.

The platform is delivered with a set of data ingestion tools programmed for common data sources of interest to the library. These connectors are programmed to transform and load data into the platform without the need for detailed technical programming by the library. In addition to these prebuilt connectors, libraries can also create customized data sources, following templates defined for the platform. The data connectors perform the initial load of data sets with automated refresh intervals.

Some of the connectors currently available include:

- Multiple categories of data from the ILS or LSP, including

bibliographic and holdings records, patron demographics, circulation data, and acquisitions data. Products currently supported include FOLIO, Ex Libris Alma, and Innovative Sierra. Connectors for other ILS products, including Sirsi-Dynix Symphony, are planned.

- Data representing other areas of library service, such as instruction or outreach.
- External sources of data related to library collections include COUNTER statistics from content providers. Upcoming versions of the product will include collection and use data from additional EBSCO products, such as GOBI, EBSCO Discovery Service, EBSCONET, and EBSCOhost.
- Institutional data describing user interactions with library resources and outcomes can be acquired through student information systems, authentication services, and campus research management systems.

The data from the multiple sources associated with a given institution are loaded into a single repository. This aggregation of data enables analysis of library data with external sources. Each institution using Panorama has a dedicated data repository, isolated from other institutions using the product.

### Underlying Technologies

EBSCO delivers Panorama as part of its growing software-as-a-service offerings. The product is entirely web based and does not require any local software.

Panorama is built on top of components from Tableau Software, which was acquired by Salesforce in 2019. Tableau provides a robust set of tools to analyze data and create visualizations and has grown to be one of the leading products in this sector. Many organizations, including libraries and



universities, use Tableau products directly for data analytics.

Tableau provides a solid foundation for Panorama. While libraries could develop their own environment based on Tableau or other analytics engines, EBSCO has created a product specifically for academic libraries that eliminate the need for technical development as well as the staff time of data librarians or analysts in normalizing and loading data sets and crafting reports and visualizations.

For product developers such as EBSCO, licensing an analytics engine such as Tableau allows them to focus their development on workflows and features, rather than on building complex infrastructure components.

## Platform Interfaces

Panorama follows a visual approach to presenting the results of its analysis, providing a variety of dashboards populated with multiple graphs or visualizations pertaining to the area of interest. The Panorama interface features a collection of dashboards, each comprising a set of visualizations that describe some aspect of library or collection performance. These

dashboards give an immediate impression of the status of its area of concern. Most of the visualizations include links that present a more detailed view of the underlying data.

According to John McDonald, director of product management for EBSCO Information Services, the dashboards and visualizations of Panorama give libraries real-time insights into impact of their collections and services. Since the data sources are continuously updated, the visualizations represent current collection or use patterns, enabling libraries to proactively address access problems or to make adjustments in the corresponding area of operations.

The initial dashboard presents an overview of the library, including visualizations for multiple high-level areas of activity such as print loans and collections, electronic resources and overall use levels, patron breakdowns, and collection expenditures. Any of these visualizations can be selected to invoke more detailed charts and analytics.

One of the dashboards aims to provide insights between library use and student outcomes. This dashboard relies on data from the student information system to illustrate inferences between library use and grades. It addresses the

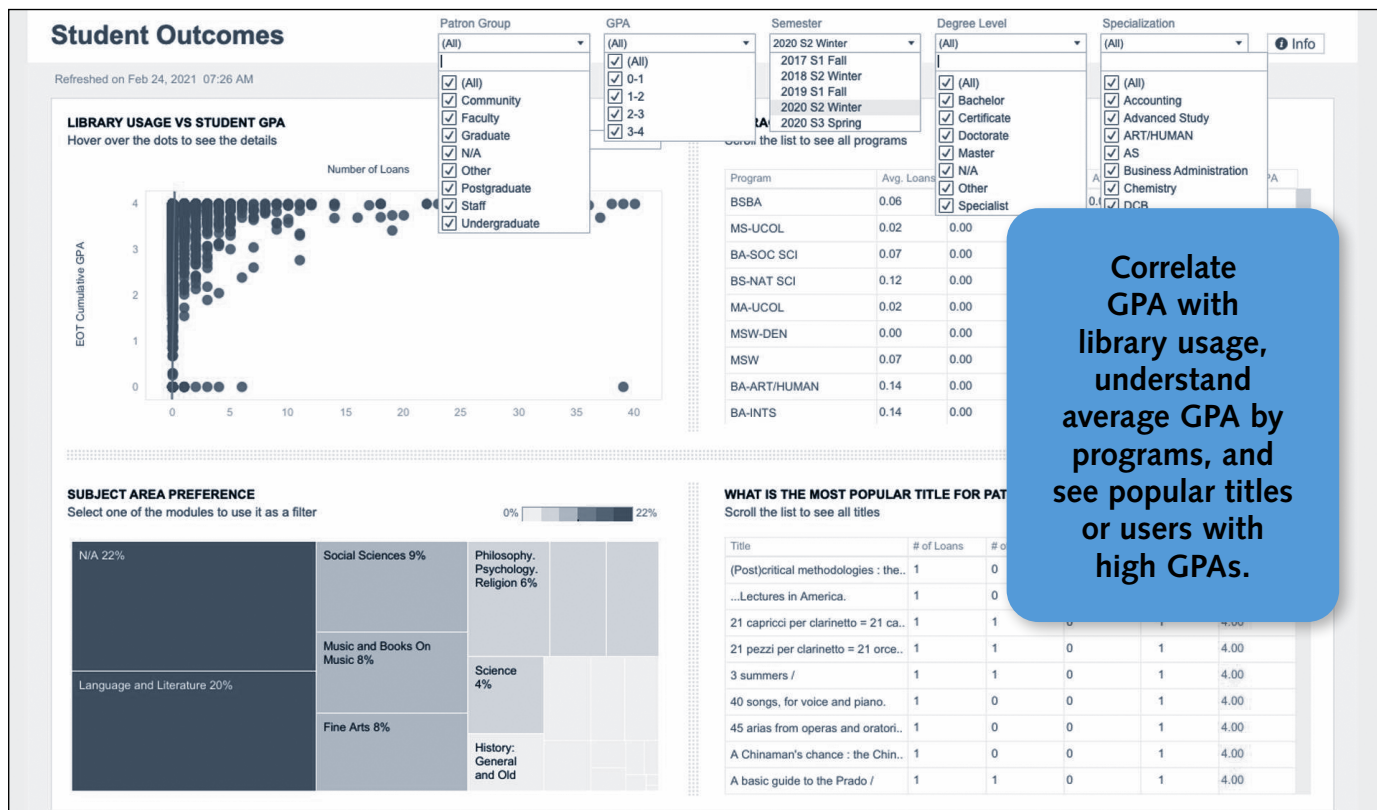


Figure 1. Panorama dashboard on Student Outcomes

question of whether higher levels of library use, such as book loans, correlate to higher grade point averages.

## Availability and Product Roadmap

EBSCO partnered with three libraries to inform its concept for a new analytics platform. University of Denver, Rensselaer Polytechnic Institute, and Chalmers University of Technology participated as alpha sites to refine the concepts and test the product as it developed. An additional set of institutions served as beta partners to test the data pipelines and expand the number of data sources. These included:

- University of South Carolina
- Seton Hall University
- University of Melbourne
- Missouri State University

The initial version of Panorama was released in March 2021. EBSCO will continue to expand the capabilities of the product to support additional data sources and corresponding dashboards and visualizations. EBSCO anticipates facilitating a community among institutions implementing Panorama to share customized tools, such as report parameters or analytical models.

## Smart Libraries Q&A

Each issue Marshall Breeding responds to questions submitted by readers. Email questions to Patrick Hogan, Managing Editor, at [phogan@ala.org](mailto:phogan@ala.org).

*What steps can the library take to improve the discoverability of items in the institutional repository?*

Libraries operating digital repositories can take several steps to ensure that the content can be easily found through the various search engines. Some of these tasks are related to general search engine optimization that would apply to any website, while others are related to the broader ecosystem of institutional repositories.

Search engine optimization techniques include tasks that ensure that each unique resource on a website can be identified and indexed by Google, Google Scholar, Microsoft Bing, Microsoft Academic, and others likely to be used by students and researchers. Even in the academic arena most literature searches begin with the general web. Search engine optimization can increase access to items in an institutional repository. Some of these techniques include:

- Verify that the robots.txt file does not impede access to the automated bots used by the search engines, as identified by user agent signatures. Although some user agents may need to be excluded to block bots that harvest content aggressively or that may download and republish content to unauthorized sites, the major search engines follow standard practices and should not burden web servers or degrade performance. The robots.txt file can optionally specify the location of a sitemap.
- Generate and update a comprehensive sitemap of all unique resources held in the repository. Sitemaps must follow the XML syntax as defined for the protocol.<sup>1</sup> These sitemaps improve the efficiency for search engine indexing, though most search engines will also crawl through all links presented in addition to those systematically listed in the sitemap. Including the link to a resource does not guarantee that it will be harvested and indexed.
- The repository should offer a unique landing page for each abstract or full text document.
- Each page delivered by the repository should embed machine-readable metadata to enable detection and indexing of resources. This structured metadata represents the citation details and unique identifiers for the resource, improving the ingestion by Google Scholar, Microsoft Academic, or other scholarly search tools. Google Scholar suggests presenting the embedded metadata following the style developed jointly with HighWire Press.<sup>2</sup> The metadata should include all authors, appropriate keywords or subject tags, and original date of publication. The date the item was uploaded to the repository should not be confused with the official publication date.
- The repository should present pages through straightforward coding and navigation. Complex Javascript or intermediary pop-up pages may inhibit site indexing.
- Any embedded metadata must be consistent with the visual presentation of the page. Even though structured metadata improves the quality and efficiency of automated indexing, Google fundamentally indexes what human visitors to the website see. Search engines may penalize or disable indexing for sites with inconsistent or misleading metadata.

Most repository platforms should offer configuration options consistent with search engine optimization. DSpace, for example, includes tools to automatically generate robots.txt and sitemaps.<sup>3</sup> Institutional repositories based on Fedora, TIND IR, CONTENTdm, or Digital Commons will each have their own procedures for search engine optimization and content syndication.

In addition to search engine optimization, libraries may also take additional measures to ensure that the server is well positioned in the ecosystem of open access repositories. The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), enables a repository to systematically contribute metadata to relevant search services. OAIster, originally developed and deployed by the University of Michigan, enables repository items to be included in a broad index of open access content, available at [oaister.worldcat.org](http://oaister.worldcat.org).<sup>4</sup> This service was acquired by OCLC in 2009.

Libraries can also add their open access repository content to the Unsub database maintained by Unpaywall, which offers a popular browser plug-in that helps researchers gain access to open access copies of scholarly articles. (See <https://unpaywall.org/sources> to register a repository). Additional services that facilitate access to open access content include:

- OpenDOAR (<https://v2.sherpa.ac.uk/opensoar>),
- Directory of Open Access Journals (<https://doaj.org>),
- CORE (<https://core.ac.uk>).

Items in an institutional repository can also see increased

use when they are included within the library's discovery service or online catalog. The major discovery interfaces, such as VuFind, Blacklight, Primo, Summon, and EBSCO Discovery Service, support the use of OAI-PMH to extract and maintain metadata from a local repository for inclusion into the library's main search index. Items found through the discovery interface would link to the full text of the item in the repository. Inclusion in the default library search tool may be the most effective way to increase use, especially when the content of the repository includes items of local or institutional interest. Most users may not think of visiting the search interface of the repository directly.

Search engine optimization and syndication of data with relevant search services can improve the mechanics of discoverability of the content held in an institutional repository. More fundamental issues include the nature of the content and the quality of the metadata. The impact of efficiently propagating metadata will be greatly inhibited if the items are not thoroughly described. Sparse or inconsistent metadata will impede discoverability when using the repository's search interface and on the external services that rely on that metadata. Most of all, the interest of the content will drive increased use of a repository. If the repository holds secondary copies of items held in more prominent destinations, placement in search results may be weakened. As with any other content resource, the key challenge concerns aligning content with the targeted audience. Libraries face obstacles in attracting high-impact content and producing rich metadata. The technical tasks in improving discoverability can be solved more easily.

## Notes

1. "Sitemaps XML format," [sitemaps.org](https://www.sitemaps.org/protocol.html). <https://www.sitemaps.org/protocol.html>
2. "Inclusion Guidelines for Webmasters," Google Scholar. <https://scholar.google.com/intl/en-us/scholar/inclusion.html#indexing>
3. "Search Engine Optimization," Confluence. <https://wiki.lyrasis.org/display/DSDOC5x/Search+Engine+Optimization>
4. "The OAIster database," OCLC. <https://www.oclc.org/en/oaister.html>



Smart Libraries Newsletter  
American Library Association  
225 N. Michigan Ave., Suite 1300  
Chicago, IL 60601-7616 USA  
Address Service Requested

---

---

NON PROFIT  
US POSTAGE  
PAID  
PERMIT 4  
HANOVER, PA

---

---

## May 2021 Smarter Libraries through Technology

*Smart Libraries Newsletter*

Marshall Breeding's expert coverage of the library automation industry.

---

### Editor

Marshall Breeding  
marshall.breeding@librarytechnology.org  
Twitter: @mbreeding

### Managing Editor

Patrick Hogan  
312-280-3240  
phogan@ala.org

### Digital Access for Subscribers

[journals.ala.org/sln](http://journals.ala.org/sln)

## TO SUBSCRIBE

To reserve your subscription, contact the Customer Service Center at 800-545-2433, press 5 for assistance, or visit [alatechsource.org](http://alatechsource.org).

ALA TechSource purchases fund advocacy, awareness, and accreditation programs for library professionals worldwide.

Production and design by the American Library Association  
Production Services Unit.

*Smart Libraries Newsletter* is published monthly by ALA TechSource, a publishing imprint of the American Library Association.

[alatechsource.org](http://alatechsource.org)

Copyright © American Library Association 2021. All rights reserved.