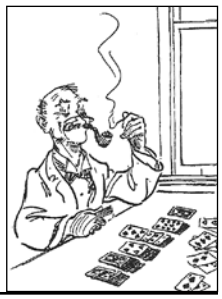
Library Analytics: Shaping the Future — Data, Privacy and the User Experience



by **Neil Scully** (IT Director, OpenAthens) <openathens@eduserv.org.uk>

Column Editors: **John McDonald** (EBSCO Information Services) <johnmcdonald@ebsco.com>

and **Kathleen McEvoy** (EBSCO Information Services) <kmcevoy@ebsco.com>

Librarians are being given more insight into online user activity, but the question of how best and how ethically to use the available data has never been more relevant.

With single sign-on access, users identify who they are by verifying which learning institution they are from. But with the **Facebook/Cambridge Analytica** data breach a hot topic, the arguments over privacy, ownership, transparency and exactly how data is stored and used has never been more important.

Publishers know that making the online user experience as positive as possible is key, but simple access must balance security and transparency, permissions and privacy.

Librarians and publishers are looking for detailed analytics so they can see who is using online services and where they are being used. These analytics do not need to be at an individual level because aggregated groups or trends also provide a great deal of value. Librarians and publishers want to ensure that end users can access as many library resources as possible. Core to this is the digital identity, which builds trust between the library and the user. The digital identity is not an email address or first name and surname but an opaque ID tied to the user data in the institutional user directory.

The digital identity authenticates the user and allows publishers to know where that user is from and what they are using the resource for. The data that can be collected carries a huge amount of value and enables strategic analysis and planning. For example, professors can see how many students have accessed course material and amend teaching literature to make sure more specialist content is utilized. It can allow faculty to plan classes on how to get the most out of library resources and to sit down with students to encourage them to make full use of online library services. Research has shown that students who access more course materials online are more likely to do well in their studies. Librarians can also see if students aren't using available resources and can re-invest their money where it will have the greatest impact.

However, there is a flip side to data like this being collected. That data is valuable, but who owns it and who manages the rights to analyze and interrogate that data? The concerns over privacy are centered around who owns the permissions — the user, organization or publisher?

We have all heard about **Cambridge Analytica** and the fall-out over Facebook data. It has come to light at the UK Parliament's digital, culture, media and sport select committee that a lot more than 87 million people might have had their data processed and analyzed without their permission and certainly without their knowledge. Former employee **Brittany Kaiser** said the consulting firm had a number of personality quizzes designed to extract personal data from the social network, including **Aleksandr Kogan's** *This Is Your Digital Life* app.

This leveraging of data without permissions has caused a major trust issue. But with General Data Protection Regulation (GDPR) which aims to give people more control over how organizations use their data and heavily penalizes organizations that don't comply with the rules — a regulatory framework for maintaining that trust is being created. GDPR provides a common standard that people controlling and processing data must follow and the protection of the individual is at the heart of the regulation.

But provided data collection is ethical and legal, institutions could be doing more to realize the value and insight contained within the data they hold. This also creates an opportunity for scholarly publishers — who can be helped to realize these benefits but always within the framework of permission — both user and institutional permissions.

Examples

With single sign-on access platforms, users can verify they are who they say they are and can be taken to a particular publisher, such as **ScienceDirect**. Librarians can then be given insight into who the user is and where they are coming from. For example, a UK University could have up to 80 partnerships or affiliated colleges across the UK and overseas.

They can now see how people from their different partnerships are engaging with their resources and use that data to optimize their collection. A lot of further education institutions struggle to get students to use the library. Now staff can look in six months' time and monitor how many people are using which platforms and for what, and base training around it.

There are of course dangers, in the general sense, of using data. Focusing on Brexit, an extreme example might be that names and data could be collected on all academics with a specific political leaning whose research talks down the UK's withdrawal from the European Union. There are many within academia who believe this kind of research should be protected.

A recent **OpenAthens** conference, called, "Championing the User," focused on current and future online users. **OpenAthens** is a gateway to secure online services through single sign-on access.

Commercial Director **Jon Bentley**, in his talk, "The Authentication Landscapes of Tomorrow," tackled the importance of trust to the user and discussed his Facebook data. He downloaded his own Facebook data ahead of the event, describing it as "shocking." Until the data held on him was delivered in its totality it had not been possible to comprehend the breadth and depth of the data that had been collected — nor how far back it went. It was easy to argue that the data was not an authentic reflection of his own identity and he is encouraging other Facebook users to download their own data to understand how much information is held on them and how it could be used to create an inaccurate profile of who they really are.

Bentley also cited the *Financial Times* as an example of a publication that is "phenomenal" at using data in a legal and compliant manner in order to create the best product and service possible for its customers. With a legitimate, user-centered approach, academic institutions can do more to make use of their data and create services that are shaped around their users and ultimately improve outcomes for all involved.

OpenAthens' Head of Sales **Rob Scaysbrook** says many institutions struggle to get students to use their library. "Analysis of data could help reverse this if it is used in the right way. **OpenAthens** is giving libraries insight into who the users are, where they are coming from and what journals and databases they are reading."

Growth Areas

Future considerations need to focus on data relationships that libraries, individual users and publishers are comfortable with, then on how that data can be managed, analyzed and best utilized. One Australian healthcare library has 30 different user types from pharmacists to medical students. If libraries require funding, these analytics can show which groups are taking advantage of their online resources and how often.

Heat maps are now available showing where users are coming from, which are proving to be a big hit with librarians because they lift a veil on the value different user communities are placing on the digital resources that are available. The way reports are being made is changing and they are becoming more flexible and as a result offering more value. It is now much easier for library staff to access and understand analytics and take advantage of the reports that are available. Many North American academic libraries are using these data and resources more, as are those in the

continued on page 63

Library Analytics ...
from page 62

U.S. healthcare sector, healthcare libraries in Australia and other countries around the world.

Many people can be very nervous about sharing data with a third party and want confidence in the technology and security surrounding that. And this is a global concern. People want an assurance that they don't have to share their data and that data won't be shared without their permission. But when it is collected and processed legally, it creates opportunities for all parties to gain rich analytics that can support decision making and improve services and ultimately deliver better outcomes.

Conclusion

Making the online user experience as positive as possible is vital and publishers know this. But privacy must not be lost as a result of easy access. Publishers need to be sympathetic to user concerns when it comes to taking and analyzing data. GDPR will help in providing a regulatory framework while allowing more people to recognize the value within data. But **Cambridge Analytica** is just one example of a situation that has highlighted dangers of data exploitation.

We know that librarians and publishers are looking for detailed analytics so they can see who is using online services and where and how much value this can bring to their future strategy. They want to ensure that end users can access as many library resources as possible and target those reports and articles that are doing well, as well as those that aren't.

Central to this digital identity governance — establishing trust between the library and the user — is using tools and technology which set a pseudonymous ID as a default. This identity authenticates the user and allows publishers to know who they are (e.g., where they are coming from without their names associated) and why they are using the resource.

With technological improvements, it is now much easier for users to access analytics and understand them. New features include the ability to open, save and favorite reports meaning they can make more comparisons and collate the data more effectively.

Some users can be very nervous about sharing lots of data with a third party and the security and policy issues surrounding this need to be addressed. They will need assurances that they don't have to share their data and that data won't be shared without their permission. However, one of the key messages is that without it, services will not evolve to be the very best they can be for all users.