Usability and privacy in academic libraries: regaining a foothold through identity and access management

Benefits arising from initiatives to streamline the user experience for academic researchers and students must be balanced against GDPR and information security measures that institutions must take to protect their members' personal data. Using the example of Bath Spa University Library's role in single sign-on projects in collaboration with the IT department and a third-party software supplier, a way in which academic libraries can more robustly enter the conversation surrounding user privacy is suggested. Identity and access management is one area of collaboration in which the librarian's traditional commitment to patron or user privacy can be upheld.

Keywords
Identity and access management; usability; privacy; security; data; ethics

PETER REID
Digital Services Librarian
Bath Spa University

Introduction

The tensions between freedom and security online have come increasingly to the attention of society in recent years. The poles of conversation swing between values of usability, and security and privacy. A balance must be struck, since a greater emphasis on the one can mean trade-offs in the other.[1] It is possible and perhaps necessary to take up a position in this dynamic, as an institution, or as a company. For example, in the cases of Google and Apple, Google appears to be betting that users will continue to value the usability benefits of its free services, balancing this freedom against the knowledge (and perhaps unease) that this is afforded by the massive amount of personal data that they, the users, are choosing to give away in exchange. Meanwhile, Apple seems to be trusting that its users will continue to pay for its services which, amongst other benefits, assure a greater guarantee of privacy.[2]

Librarians have traditionally upheld the maintenance of a patron's right to privacy as a professional duty.[3] However, in the case of academic libraries in particular, the library's role in managing collections has largely switched from that of custodian of local content to broker of access to subscribed content hosted elsewhere. In addition, the infrastructure for managing access – in so far as it is managed on-campus – depends at least as much on the IT department as staff within the library. It also depends on other access brokers and IT staff working for myriad publishers and content aggregators. This work is fundamentally collaborative and relies on sharing information about the library, its users and its resources. Therefore, since much information about library users' behaviour is no longer under the library's direct control, the challenges of upholding the professional commitment to privacy have changed quite drastically.

'Librarians have traditionally upheld the maintenance of a patron's right to privacy as a professional duty'

One often-overlooked area where the tensions between usability and security become a professional concern for library services is the area of authentication and authorization: the matter of managing how users log into library-provided subscribed content. Using examples from Bath Spa University Library's involvement in this area, this article discusses how and why a library – and indeed the wider institution it serves – can go beyond a privacy statement on its website and begin to better embody its professional principles in the technology it uses to provide services.

Identity and access management

Libraries generally participate in the area of identity and access management in two ways. First, they will ensure the terms of the licences signed with publishers of scholarly content are met in practice (since the subscription model usually requires that only members of that institution have access). Second, when the user receives an error or refusal-of-access message to a resource that has been sourced and logged into through the institution or its library, their help will be sought.

A third way a library might participate in this area is when brokering arrangements with a publisher or aggregator on a user's legal and ethical entitlement to privacy and the institution's necessary legal commitment to information security. As well as legal awareness, in order to have an effective voice in this area, the library requires staff to have knowledge of its institution's identity provider (IdP) software and how this works as an intermediary between the university, the publisher and the user or reader (just like a traditional librarian). An IdP makes decisions on what information about a student to release to third parties – information which the student has trusted to the institution.

Within the IdP, details about users are categorized as 'attributes' – standardized data types which can contain information such as which institution they are from, the nature of the relationship between the institution and the individual (for example, whether they are a student, staff member, an affiliate, or a combination of the above), a unique personal identifier (commonly, an institutional username), right down to the individual's first name, last name and e-mail address.

Third-party services and applications, such as a publisher platform, require varying levels of this data to be released by the institution (IdP) in order to make a technical decision whether to allow a user through its website 'paywall' or not.

A basic indication of whether they are from the institution which pays for the publisher's services will often suffice for access, but in order for the third party to provide more optimized services (and arguably, for better usability), more detailed or granular (or more personal) data may be requested (see Figure 1).

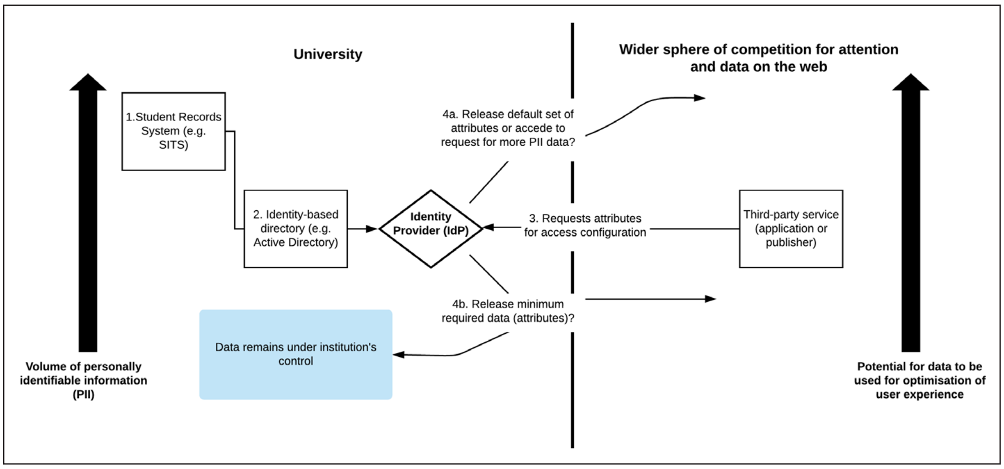
University
Wider sphere of competition for attention and data on the web
1.Student Records System (e.g. SITS)
4a. Release default set of attributes or accede to request for more PII data?
2. Identity-based directory (e.g. Active Directory)
Identity Provider (IdP)
3. Requests attributes for access configuration
Third-party service (application or publisher)
4b. Release minimum required data (attributes)?
Data remains under institution's control
Volume of personally identifiable information (PII)
Potential for data to be used for optimisation of user experience

Figure 1. Typical user data flow in configuring a third-party service for access through a university's identity provider

Keeping in mind that balancing usability and privacy is a strategic decision each web-based business must make, it follows that precisely which and how many attributes a provider asks for from the institution can say something about its underlying business model. For instance, at the time of writing, SAGE, Wiley-Blackwell and JSTOR each require a single attribute to indicate that the user is from the host institution in order to authorize access to its journal holdings.[4] In contrast, in order to grant institutional access to LinkedIn Learning (formerly Lynda.com), LinkedIn require the IdP to release a personal token by which to remember the user each time they visit the site, the automatic release of the user's institutional e-mail address, as well as strongly encouraging users to associate their use of the learning platform with a pre-existing LinkedIn account.[5] The difference is that the former three platforms, in return for a subscription fee, fulfil their contract with the institution by providing the student with access to a set of resources which are presented in the same way for each user from the university. LinkedIn Learning wishes to personalize the user experience and, by doing so, grow the user base of a data-driven social network, an identity which is embedded in the user's life beyond the confines of their university membership. While access to both requires an annual institutional subscription, LinkedIn Learning could make additional if not greater profit in addition to this thanks to the institution's willingness to 'pay' for services with its users' data.[6] This attribute release may or may not accord with the library's, the institution's or indeed the user's values or long-term interests. How the institution might respond to this challenge is discussed later in this article.

'balancing usability and privacy is a strategic decision each web-based business must make'

Change of identity, loss of personalization: Bath Spa's migration of single sign-on provider

There are numerous library-subscribed resources which require the use of a personal user token by which to remember returning users too – notably reference management services and e-book collections which rely on digital rights management. In those examples, this tie to a personal account on the platform is crucial to the service being provided. In the case of reference managers like EndNote and RefWorks, it enables users to create bespoke collections of references for particular assignments and referencing styles: for e-book collections, it asserts the publisher's and author's legal right to control copyrighted material.

To remember returning users in cases where a platform-based personal account is needed, the UK Access Management Federation (UKAMF) recommends that the IdP releases a particular attribute called eduPersonTargetedID. (The UKAMF is a Jisc-run service which facilitates the standardized exchange of data between institutions and service providers for the UK.) The institution and its IdP ensures that this attribute has a unique value both for each user and for each resource that receives it – so that, crucially, individual user behaviour across websites and services cannot be tracked or shared.[7]

The configuration as well as the maintenance of privacy-preserving access arrangements can benefit significantly from a front-facing, user-centred service such as the modern academic library. Bath Spa University Library deepened its involvement in this area in 2017, during a project to migrate the institutional identity provider to a hosted solution. The IdP software had before that point been managed by the Library, having been purchased to simplify the log-in experience for Library databases in 2010. Over the years, IT services had gradually begun to use the software to provide single sign-on to other university services, such as the virtual learning environment (VLE) and staff/student intranet. The 2017 project presented the opportunity to collaborate more fully.

This highlighted differences in the professional cultures of the groups involved. For example, IT, according to its remit, prioritizes a systemic and rule-based approach to technical challenges, and has justified pride in the expertise of its staff to provide elegant solutions that work for the great majority of use cases. Academic librarians on the other hand have a more singular purpose – to connect communities of readers with relevant reading material as smoothly as possible (to quote S R Ranganathan, 'save the time of the reader'[8]), and even operational and technical-focused librarians work relatively closely to the front-line touch-points of this endeavour. In a university, librarians also provide a long-established tradition of liaison with the academic and teaching staff. While this focuses attention on the end-user experience of the individual as well as the core business of the university, with all its possible nuance, this focus can risk paying too much attention to the smaller details. There is also a danger that professional services staff may self-disqualify from problem-solving in IT or web-based projects due to the perception that it is too complicated, or that they themselves are 'not technical'. In this way, librarians risk excluding themselves from a crucial arena of decision-making.

Therefore, when a specific challenge emerged on the project at Bath Spa, the means to transcend these cultural differences was needed. Midway through, it transpired that the particular way that the unique eduPersonTargetedID attribute was generated for the Library's services in the old identity provider could not be replicated in the new hosted service. IT were surprised to learn from the Library that this risked the loss of personal data to thousands of single sign-on users of third-party applications, particularly related to notes made against e-books and references used in the writing of assignments. This would have been seriously disruptive.

'librarians risk excluding themselves from a crucial arena of decision-making'

The necessary bridge was provided by a service-neutral project management structure. This enabled regular meetings of the project board with representatives from IT teams, the Library and the new IdP software supplier, Overt. It also focused attention on a user-centred quality assurance test plan that flowed from written user stories, enabling issues to be identified early and resolved.

From left to right, Bath Spa Library and Learning Service's Deputy Director (Digital and Research), Senior Network Engineer, IT's Information Security Manager, and Network Manager, pictured in front of Bath Spa University's 'Media Wall'. All were involved in the IdP project in various ways, and their roles all have a remit to ensure creative work and teaching at the University stays within various technical and legal boundaries (e.g. copyright, GDPR, network security).

IT commissioned a script from Overt which replicated the unique user value, which the Library was able to test to ensure the personalized data which users expected to return to was maintained seamlessly. Indeed, it took some agonizing and detailed project work over three months to achieve the desirable (if unexciting) outcome: that on the migration date, the students and staff of the University did not notice anything.

In order to maintain the personal data across future updates to the Shibboleth single sign-on standard, the Library was further required to do very detailed work to help its suppliers update the unique eduPersonTargetedID values when the workaround script was decommissioned in the summer of 2019. This took a lot of staff time, and while it is true that compromising through the use of a simpler cross-service user attribute such as e-mail or eduPersonPrincipalName (which is in effect their University username) would have precluded the need for the work, the eventual solution provided stronger protection for users' privacy.

Overall, a balance between the need for a smooth user experience providing institution-wide access to subscribed resources and the preservation of the institutional member's privacy had been struck.

Discussion: Institutional ID over Social ID

'a balance between the need for a smooth user experience … and … institutional member's privacy'

The offer of individual membership to a service or site is of course a key source of the massive data businesses that have emerged on the web, along with all its controversies. Naturally, every player in the online marketplace – including universities and their libraries – has to respond to the tendency or bias towards monitoring and maximizing usage of their web-based service via the tracking and aggregation of their users' unique IDs.

It is sometimes assumed unquestioningly that tracking student activity across all university-provided services, including library databases, is already in effect. It is possible to argue that user data should be used to improve services and alleviate the pressure we are under to improve student experience and justify tuition fees.[9] And it can be the case that other university-provided services, including externally hosted VLEs, are configured to receive batch imports of new users, extending as far as an entire copy of the institution's user base. Like many web businesses, third-party services may depend on the number of registered users and corresponding statistics like time spent on the application to generate revenue, so mass sign-ups of institutional members will rarely be refused; indeed, the practice may be encouraged in services' instructions to technical staff who are asked to configure resources on behalf of academics or librarians.[10]

'privacy-preserving means of providing and sustaining personalized memberships of third-party sites are possible'

However, in counterbalance to what has been a prevailing trend to make liberal use of students' implied consent, our sector is likely to become increasingly mindful of the General Data Protection Regulation (GDPR) requirement for 'privacy by default and design'.[11] As in the example of Bath Spa's work with Overt, privacy-preserving means of providing and sustaining personalized memberships of third-party sites are possible. The institutional liaison with the third-party service – in our example, the librarian – is the party in control of that discussion. This is one small example of how an institution can decide, as Google and Apple are doing, how much and what way it wishes to engage with the pressures to maximise usage through data-driven business models.

Where the institution wishes to engage in a usage statistics or engagement-driven learning analytics project, for example, if it can align its business needs with the IT team and, in particular, the student records team, then with a closely managed IdP a university can release a meaningful indicator of the individual user's affiliation with the university alongside the attributes required for access. This way, via the use of an IdP's usage logs or dashboard, usage metrics of expensive resources can be triangulated against any number of variables. These need not go out to the provider, but can still be present for university stakeholders to gain analytical insight (Figure 1). Whether this is the student's course, or other category based on market research, it empowers the institution, and also localizes the risk of data security breaches.

On the other hand, universities and libraries might look to a different approach, one more in line with the tradition of protecting patron privacy, and embody this in its online connections to third-party services. In the case of the eduPersonTargetedID attribute, if it holds third parties to accepting *only* this as its means of identifying returning users, then it could vouchsafe a privacy policy that takes a counter stance against prevailing business practices that are inspiring a growing degree of public mistrust. In this vein, the GÉANT Data Protection Code of Conduct for publishers, providers and home organizations (e.g. IdPs) encourages the release of the minimum amount of user attributes for the provision of the service in question.[12]

Technologist Jaron Lanier called the provision of research services in print libraries 'the very last bastion where you aren't being spied upon'.[13] While, in the author's opinion, much ground has already been lost in the rush to avoid being disrupted by the giants of Silicon Valley over the past 20 years (not least, through the widespread adoption of free Google apps for education), being seen to reclaim some authority in this space may be of significant appeal to an idealistic student market which, if not rebelling from social media, is like many in the population, increasingly ambivalent.[14]

'the GÉANT Data Protection Code of Conduct … encourages the release of the minimum amount of user attributes for the provision of the service'

Conclusion

There is a specialist role for libraries in the systematic review and restriction of data released to third-party providers – both as agents of technical understanding and collaboration, and when negotiating subscriptions and renewals. The example of Bath Spa University's collaboration with Overt is useful in that it was with a company that was contracted at the right stage of its development, of a size whereby resources could be allocated flexibly, working relationships with the relevant technical and operational teams easily established, and in a project structure where the right people were identified to deliver the work. As new features such as analytics tools are developed, this arrangement enables the Library to test new features that are relevant to its business, while being fully informed on the privacy and security implications of the use of that data.

More generally, there is an important role for libraries to play in developing new, privacy-orientated standards and access management systems. For example, SeamlessAccess.org (that builds on RA-21,an initiative which promises a 'continue with X University' button on sites like ScienceDirect in the 'social login' mode similar to 'continue with Facebook/Google/LinkedIn' options on many commercial sites), and the GÉANTCode of Conduct mentioned above.[15]

'there is an important role for libraries to play in developing new, privacy-orientated standards and access management systems'

This active role will put libraries in a position to meaningfully factor adherence into licence negotiations, giving us an empowered position on not only business and operational matters, but ethical ones, too. If Apple and Google are betting their futures in part on which philosophy wins out in the usability vs privacy debate, surely universities as a whole, and their libraries, should be able to take up a leading position.

Abbreviations and Acronyms
A list of the abbreviations and acronyms used in this and other *Insights* articles can be accessed here – click on the URL below and then select the 'full list of industry A&As' link: **http://www.uksg.org/publications#aa**.

Competing Interests
The author has declared no competing interests.

References

1. Gurpreet Dhillon, Tiago Oliveira, Santa Susarapu, and Mario Caldeira, "Deciding between information security and usability: Developing value based objectives," *Computers in Human Behavior* 61, Issue C (August, 2016), DOI: https://doi.org/10.1016/j.chb.2016.03.068 (accessed 22 October 2019).

2. "Apple's quiet AI acquisition," *AngelList Weekly*, November 21, 2018, https://angel.co/newsletters/apple-s-quiet-ai-acquisition-112118 (accessed 22 October 2019); Daniel Greene and Katie Shilton, "Platform privacies: Governance, collaboration, and the different meanings of 'privacy' in iOS and Android development," *New Media & Society* 20, no. 4 (April, 2018), DOI: https://doi.org/10.1177/1461444817702397 (accessed 22 October 2019).

3. "Library Bill of Rights," American Library Association, http://www.ala.org/advocacy/intfreedom/librarybill (accessed 22 October 2019); "IFLA Statement on Privacy in the Library Environment," IFLA, https://www.ifla.org/publications/node/10056 (accessed 22 October 2019); "The Ethical Principles", CILIP, https://www.cilip.org.uk/general/custom.asp?page=ethics (accessed 22 October 2019); Michael Zimmer, "Assessing the Treatment of Patron Privacy in Library 2.0 Literature," *Information Technology & Libraries* 32, no. 2 (June, 2013), DOI: https://doi.org/10.6017/ital.v32i2.3420 (accessed 22 October 2019).

4. "UK Federation Information Centre|Documents/AvailableServices Browse.", UK Access Management Federation, https://www.ukfederation.org.uk/content/Documents/AvailableServices (Accessed 26 October, 2019).

5. "Privacy and Security Whitepaper: Account Center Employee Database Integration (EDI) and Single Sign-On (SSO)," LinkedIn, December, 2016: 7, https://business.linkedin.com/content/dam/me/business/en-us/sales-solutions/resources/pdfs/privacy-and-security-whitepaper.pdf (accessed 22 October 2019).

6. Roshan Sumbaly, Jay Kreps, and Sam Shah, "The Big Data Ecosystem at LinkedIn," *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data* (June 2013), DOI: https://doi.org/10.1145/2463676.2463707 (accessed 22 October 2019).

7. "Recommendations for use of Personal Data Version 4.1," UK Access Management Federation for Education and Research, July 2018, https://www.ukfederation.org.uk/library/uploads/Documents/recommendations-for-use-of-personal-data.pdf (accessed 22 October 2019). It is the creation of a multi-service identifier that is cited in the recent and well-publicized complaint by the Irish Data Commission against Google for suspected infringement of GDPR – since a multi-service ID is a means by which Google can sell a user's identity to firms which are competing to learn as much as they can about their online customers' behaviour. EdupersonTargetedID is a pseudonymous identifier that is specifically designed to prevent this: Jane Wakefield, "Google's 'secret web tracking rages' explained," *BBC News*, September 5, 2019, https://www.bbc.co.uk/news/technology-49593830 (accessed 22 October 2019).

8. Shiyali Ramamrita Ranganathan, "The five laws of library science," (Madras: The Madras Library Association, 1931), DOI: https://doi.org/10.1145/2463676.2463707 (accessed 22 October 2019).

9. Ken Chad and Helen Anderson, "The new role of the library in teaching and learning outcomes," *Higher Education Library Technology* (HELibTech) briefing paper, no. 3 (June 2017), DOI: https://doi.org/10.13140/RG.2.2.14688.89606/1 (accessed 22 October 2019).

10. "Privacy and Security Whitepaper," 3.

11. "Data protection by design and default," Information Commissioner's Office, https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-by-design-and-default/ (accessed 22 October 2019).

12. "Data Protection Code of Conduct, Version 1.0," GÉANT (June, 2013), http://www.geant.net/uri/dataprotection-code-of-conduct/v1 (accessed 11 September 2019).

13. Speaking in 2013, when MOOCs were a popular venture, Lanier also observed that 'at the moment, public institutions, particularly the universities, are racing to join into this digital system that will lead them to oblivion … you'll have your Harvards and your Stanfords, who will survive as a brand, then you kill all the intermediate schools.' "LIVE from the NYPL: Jaron Lanier," New York Public Library (October, 2013), https://www.nypl.org/sites/default/files/livelanier_10.10transcript_0.doc (accessed 22 October 2019).

14. Gina Pingitore, Vikram Rao, Kristin Cavallaro, and Kruttika Dwivedi, "To share or not to share: What consumers really think about sharing their Personal Information," *Deloitte Insights* (September, 2017), https://www2.deloitte.com/us/en/insights/industry/retail-distribution/sharing-personal-information-consumer-privacy-concerns.html (accessed 22 October 2019).

15. "Recommended Practices for Improved Access to Institutionally-Provided Information Resources: Results from the Resource Access in the 21st Century (RA21) Project," National Information Standards Organization (2019), https://groups.niso.org/apps/group_public/download.php/21376/NISO_RP-27-2019_RA21_Identity_Discovery_and_Persistence-public_comment.pdf (accessed 22 October 2019).

Peter Reid
Digital Services Librarian
Bath Spa University, GB
E-mail: p.reid@bathspa.ac.uk
ORCID ID: https://orcid.org/0000-0002-2264-7871

To cite this article:
Reid P, "Usability and privacy in academic libraries: regaining a foothold through identity and access management", *Insights*, 2019, 32: 33, 1–8; DOI: https://doi.org/10.1629/uksg.487

Submitted on 12 September 2019 Accepted on 21 October 2019 Published on 13 November 2019

Published by UKSG in association with Ubiquity Press.