

‘All your data are belong to us’: the weaponisation of library usage data and what we can do about it

UKSG eNews 527

11 November 2022



By **Caroline Ball** – Academic Librarian, **University of Derby**, #ebookSOS campaigner
Twitter: [@heroicendeavour](#), [Mastodon: @heroicendeavour@mas.to](#)

and **Anthony Sinnott**, Access and Procurement Development Manager, **University of York**
Twitter: [@librarianth](#)

What do 850 football players and their performance data have in common with academic libraries and online resources? More than you’d think! The connecting factor is data, how it is collected, used and for what purposes.

‘[Project Red Card](#)’ is demanding compensation for the use of footballers’ performance data by betting companies, video game manufacturers, scouts and others, arguing that players should have more control over how their personal data is collected and particularly how it is monetized and commercialised.

Similarly, libraries’ online resources, whether a single ebook or vast databases, are producing enormous amounts of data, utilised by librarians to assist us in our vital functions: assessing usage and value, determining demand and relevance.

But are we the only ones using this data generated by our users? What other uses is this data being put to? We know for certain that vendors have access to more data than they provide to us via **COUNTER** statistics etc, but we have no way of knowing how much, what types, or what is done with it.

Witness the recent controversy generated by [Wiley’s removal of 1,379 e-books from Academic Complete](#). Publishers like Wiley determine high use by accessing statistics generated by our end-users via the various e-book platforms through which they access the content. This in itself is indicative of our end-user/library data being provided to third parties without our knowledge or consent, particularly concerning given our licences are with vendors and not publishers. We are also not privy to what data-sharing agreements exist between vendors and publishers. Should we allow library usage data to be weaponized against us in this fashion? What recourse do we

have to push back against this practice of ‘data extractivism’, to either withhold this data from publishers and vendors or prohibit them from using it for their own commercial gain?

It should concern us that most of the user agreement licences that govern library e-resources make scant reference to this data. The [JISC model e-book licence](#), for example, contains one clause (6.2) referring to vendor use of data generated by our users (compared to multiple clauses governing what those users can do with publisher-owned material) and that specifically relates to the use of personal data and GDPR and entirely fails to address the aggregate data being collected from libraries that may be anonymised and therefore falls outside the scope of the 2018 UK Data Protection Act.

Almost all of us would balk at the idea of libraries providing detailed behavioural data relating to print books to vendors and publishers: *the most borrowed books, when, how long were they borrowed for, how far did readers get, which pages did readers spend longest on, which books did they read next, how long did they read in total, what notes did they make, what bookmarks, what terms did they look up in the index...* Almost certainly, were we to stand at the shoulders of our users, silently observing their reading and note-taking, they would vigorously object to such an invasion of privacy. Why then do we so thoughtlessly permit publishers to do this to our digital users?

As a sector, we seem to have sleepwalked into a position where mass observation, under the guise of engagement monitoring, has been allowed to creep into a huge swathe of our digital resources, unseen, unregulated and ill understood. Whilst individual and personal data has been protected, anonymised mass data is being extracted, aggregated, and supplied by vendors to publishers. This is almost guaranteed to create opportunities, not for libraries, universities, public libraries, and end users, but for these commercial entities.

Vendors use this surveillance and data oversight as a selling point for their products, turning themselves into providers of ‘learning analytics’ every bit as much as e-books. In other words, marketing the benefits of enhanced student engagement through data-driven metrics. Yet again, why is this a benefit for digital resources and not an invasion of privacy? Why should we accept this in the digital world, and not in the analogue world? What are the pedagogic implications of turning education and learning into Big Data to be crunched and analysed, users morphed into statistics to be monetised?

[As more than one critic has](#) noted, more than anything, learning analytics track compliance. We don’t know how much a student has learned, how much information they have retained, how it relates to the rest of their knowledge - merely how much they have conformed to this new data-driven template of instruction. And increasingly there is no opt-out from that - if students do not consent to their learning, reading, research (their very curiosity), being monitored and tracked, there are no alternatives available, no other options. Not for our students, and not for us either.

Another concern is the potential impact of misuse of this data by vendors and publishers. We are already aware of one recent example, demonstrated by Wiley and **ProQuest** - removal of high-use titles from packages in order to re-sell them back to libraries via more profitable e-textbook subscription models. Indeed, can we call it misuse when there is nothing illegal about this practice, and as far as we know no violation of GDPR is taking place, and no breach of licence terms or contract clauses occurring? We may think this approach immoral, but it is only taking place because the sector has turned a blind eye for so long.

The #ebookSOS campaign has been pushing back against ebook prices and unfair licence terms, but unfortunately as a sector we are in a situation of our own making. Publishers pitch their titles at prices they believe the market will bear, justified in these decisions by the data we have so blindly provided to them. They know which titles our students use the most, down to the edition, chapter, and page. We have permitted this, even enabled it.

Unless libraries respond to this type of data extractivism in an effective and collective way, this will become the reality of how digital resources exist within the ecosystem of learning environments. We need to establish a widespread opt-out position and knowledge about global data usage while we still can.

Would the same #ebookSOS campaigning approach work for this data issue? Unlikely. How can we campaign and agitate and rally our troops to begin pushing for satisfactory resolutions without proper information? This isn't even an inability to envisage what victory might look like; in our current situation, we don't even know what the battlefield looks like! We are so ill-informed about the data and user metric landscape emanating from libraries and our users that it is almost impossible to properly articulate what an ideal result would be like.

Therefore there is work that has to be done. We need to rapidly increase the scope of our knowledge, and it is imperative that this happens soon as it is already late in the game from the perspective of publishers and vendors. We need to push sector bodies to begin questioning suppliers, and we also need consortia bodies to consider user data in framework agreements to the benefit of public and university libraries. We need to question suppliers ourselves and share that information across the sector. We must urgently discover what is happening with our data and make determinations about what value is being derived from it. We must have licences that cover the issue of all data generated from digital resources, not just identifiable personal data. We need to know about all the data collected by platforms and vendors, not just those statistics included in COUNTER. We need transparency about which entities have access to this data (i.e., publishers) and what they do with it.

The longer-term purpose of this would be to develop a sector wide choice as to whether to allow or disallow this form of data monitoring. From this initial stage it could be a mechanism that allowed for libraries to monitor usage of their own collections but precluded vendors and publishers from seeing and/or acting upon it. Obviously there would be technical and logistical hurdles to overcome and it clearly requires a level of institutional commitment and maturity to negotiate. But one thing is certain, it won't ever be something we can use to our own benefit if we aren't prepared to fight for it.

If you would like to know more about this subject, the webinar [‘All your data are belong to us’: the weaponisation of library usage data and what we can do about it](#) will take place on the 5th of December.

These views are the authors' own and do not necessarily reflect the views of UKSG.



This UKSG Editorial is taken from the industry newsletter UKSG eNews, published every two weeks exclusively for UKSG members. The newsletter provides up-to-the-minute news of current issues and developments within the global knowledge community.

To enjoy *UKSG eNews* and other member benefits [become a UKSG member](#).