SCIENCE

# How Scientific Publishers Monitor Researchers

In recent years, major publishers of scientific publications have invested heavily in data collection. Researchers fear that research is sinking into a race for performance.

By Aurelien Defer

Published yesterday at 6:00 p.m., updated at 5:10 a.m. • Reading 6 min.

Article reserved for subscribers

BORIS SEMENIAKO

Jonny Saunders, a neuroscientist from the University of Oregon (USA), made a surprising discovery this winter. From his workstation, he goes one evening in December 2021 thanks to the network of his faculty to the site of the renowned scientific publisher Elsevier. “I had heard of these scientific publishers who collected data on their users, he tells Le Monde. I just wondered by what means they do this and if these trackers show up in the code of their web pages. For some time now, the scientific community has been concerned about a new practice by major publishers: the collection of increasingly detailed data on

researchers and their work.

That evening, Jonny Saunders then opens a text in Elsevier's online software and, after a few quick manipulations, realizes that he is "observed" by three different tools which send information in real time to servers. third.

Read also | How the Covid-19 pandemic has disrupted scientific publishing

“ One of them communicated identifying data to Elsevier's servers, in particular the proxy of the university from which I was connected, which allows them to identify me precisely,” he recalls. Later, he notices between the lines of code that “events” are also described and collected by the editor and inform the latter very precisely about his reading behavior.
Different information is then coded if he places his cursor on the upper part of the text, if he reduces the browser window, if he stays on the page for a long time or not, etc.

## Elsevier prefers to dodge controversy

Surprised by the accuracy of the data collected by Elsevier, Jonny Saunders tweeted immediately to alert his peers. But a month later, apart from a small wave of indignation that remained confined to the English-speaking scientific community, the news did not make much noise. We asked Elsevier to respond.

The Dutch publisher refuses to "comment specifically" on what Jonny Saunders revealed but acknowledges the existence of a significant collection of data on its users. And this for several purposes, assures the company: “We use data tracking tools in order to provide and improve our services, (…) to help us authenticate users, secure our services, detect fraud and abuses, (…) to facilitate efficiency and productivity in research. Elsevier also mentions, without making direct reference to them, the very specific processes discovered by Jonny Saunders: users. »

Read also | Upheavals in the evaluation of scientists

Behind these few lines of defence, Elsevier seems to assume the turn taken in recent years by the leaders of the scientific publishing market, which are turning massively towards the collection and use of data.

In 2016, the British group RELX, parent company of Elsevier, had moreover announced, in an annual report, the “organic development of analytical tools based on increasingly sophisticated information (…)”, passing from a simple publishing group to a "technology, content and analytical tools company".

## “Datafication” of scientific publication

For Jefferson Pooley, professor of information and communication sciences at Muhlenberg University (United States), the recent affair involving Elsevier is irrefutable proof of this “datafication” of scientific publication. "To my knowledge, there hasn't been another discovery of this kind of behavior monitoring in the research world that is so precise, so granular," he explains.

In a research work entitled "Surveillance Publishing" and published in November 2021, Jefferson Pooley dates the beginnings of this publishers' appetite for data to the early 2000s. While the business models of Gafam, like that of Google, are built around “data” as a commodity, scientific publishers also see potential profits in it.

Those who dominate the market today have acquired in recent years companies capable of collecting all kinds of data, in particular to measure the impact and performance of a research work and its author. Elsevier thus acquired the company Pure in 2012, enabling it to

collect data on the performance of academics or Plum Analytics, in 2017, to quantify the impact of publications.

Read also | A new tool to track “bogus” scientific articles written by software

This information, as well as that on the behavior of researchers, is distinguished by its precision from that traditionally drawn by digital players from anyone browsing the Internet. Especially since scientists, because of the oligopolistic structure of the publishing market, very often have no other choice but to go through the leading publishers and obey their rules. Many of them are unaware that certain browsers or tools could prevent them from being tracked, or are unaware of the very existence of such collection tools.

Some of this data is sometimes monetized directly by publishers through the sale of services, such as Pure's, to universities. But it is difficult to determine the use that publishers have of it. "We don't know the full range of metrics that these publishers collect," says Jefferson Pooley, who points to a particularly opaque system. We know they collect the traditional metrics (citation metrics, number of downloads, reads…) and we now have good reason to believe that other data is being monitored, including how we as researchers read an article , the time we spend there, the highlights we use, etc. »

## Towards a science that works on performance

Always with the aim of better measuring research performance, scientific publishing houses are turning to increasingly precise tools.

Recently, the American start-up Scite, already a partner of publishing leaders such as Wiley and Sage, has offered a unique service for collecting data on scientific citations. It allows its customers to determine if, when an article is cited, it is because it is approved or on the contrary criticized by the author and, in this way, avoids the problems of precision encountered by traditional metrics. .

Read also | Diving into the questionable practices of scientific publications

Jaele Bechis, researcher at the Office of Theoretical and Applied Economics of the University of Lorraine, has studied at length, as part of her thesis and her research, the market for scientific publication in the digital age. “You have to keep in mind that these are commercial publishing houses whose objective is to make a profit,” she describes. They are now implementing strategies that large companies in other areas are also adopting. And if some of this data “can be used by the scientific community”, refractory academics argue that a commodification of science, accelerated by these “datafication” strategies, is to be feared.

**Reluctant academics argue that a commodification of science, accelerated by these “datafication” strategies, is to be feared**

According to them, sharpening these impact factors so precisely and going so far as to inspect the behavior of researchers would amount to motivating scientific production by performance and no longer by the search for knowledge. Thus, the most fashionable fields, the most spectacular works and the most widely read authors would have a better chance of being financed and published while others, less fashionable, would be neglected by this system, suggests the researcher. Jefferson Pooley.

“National research agencies, particularly in the English-speaking world, have already begun to emphasize the importance of impact factors,” he explains. This trend is likely to intensify as publishers invest in data. »

**Aurelien Defer**

## Services