



From paywall builders to data tracking moguls or... How the big publishers have put on a new super villain costume

Didier Torny

► To cite this version:

Didier Torny. From paywall builders to data tracking moguls or... How the big publishers have put on a new super villain costume. Politics of technoscientific futures, EASST, Jul 2022, Madrid, Spain. hal-03885480

HAL Id: hal-03885480

<https://hal.science/hal-03885480>

Submitted on 5 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

From paywall builders to data tracking moguls or... How the big publishers have put on a new super villain costume.

Didier Torny

communication given at the [2022 EASST Conference](#) in Madrid, July 2022.

Elsevier is a felon, that is a given. This company epitomizes all the crimes, misdemeanors and petty theft that can be accomplished by a publisher. Its wikipedia page is so full of affairs, scandals and raunchy stories that it would be enough to read it to give a talk to an academic congress. And yet Elsevier still find new ways to extract value from academic communities, which both produces new profits and new critiques. This communication is the story of the publisher becoming a data company.

An endless list of academic misdemeanors

In 1880, It all begin with a “borrowing” as we say in academic life, which may also be named an homage, a plagiarism or a steal, depending on the point of view. When the company was founded, it took over a logo from an ancestral famous printing Dutch family, whose name was Elzevier (yes, with a Z).



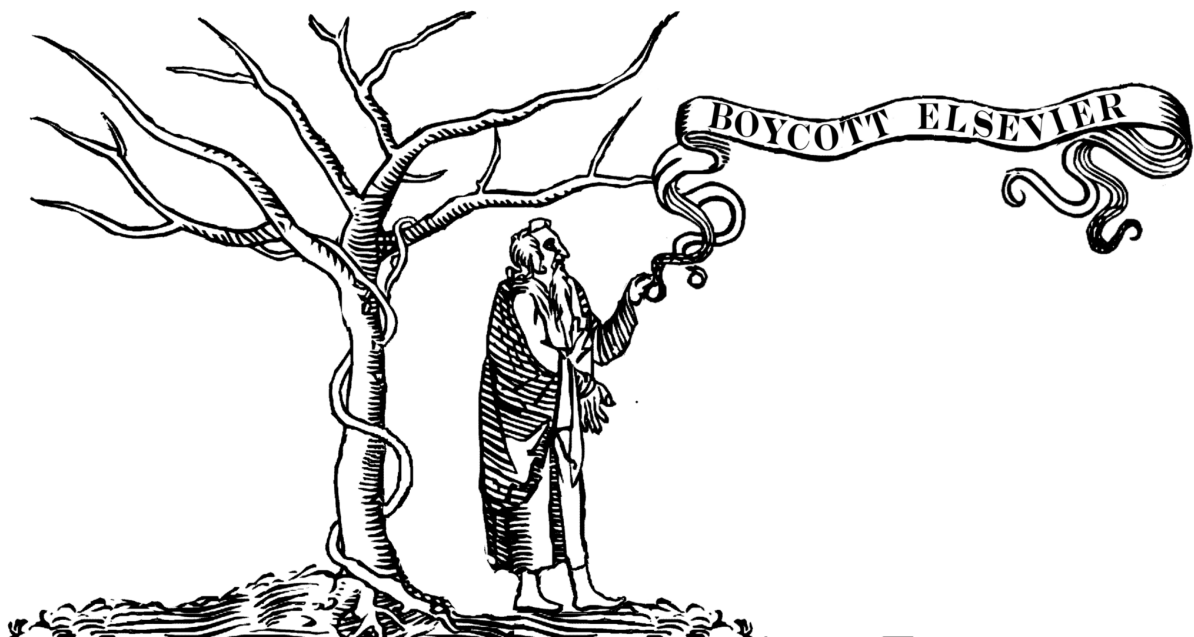
ELSEVIER

As a Dutch publisher, they first put out journals in the language of the country, but the need to go into exile in England in 1940, and no doubt a rather specific vision of scholarly communication, led the company to launch English-language journals in the post-war period. Along with Pergamon, Elsevier is certainly the inventor of the concept of the ‘international journal’ and created a global market for scientific writings, with customers all over the world and, what is more, a profitable market. This led to cycles of development and acquisitions, which continue to this day. But as we know in the world of superheroes, “with great power comes great responsibilities”.

And indeed they are responsible. The list of “problems” attributed to Elsevier can be categorised into three different groups: firstly, a propensity to act in a “sloppy and dirty” manner, for example in copyediting failures, by selling closed articles for which authors have already paid an APC, or by not acting in front of legitimate requests for retracting articles, as in the very recent following example.

Secondly, its constant pursuit of profit leads it to bend academic rules. Above and beyond offering researchers Amazon vouchers to write reviews on products, one of the most famous examples is the publication of journals in Australia that were *de facto* advocacy media for Merck pharmaceutical products, through a subsidiary that is cited by Sergio Sismondo as an example of ghost management².

Third, its concern for protecting its intellectual property leads it to numerous actions opposing open access, unlimited text and data mining, or even metadata sharing. Elsevier thus funds numerous lobbying actions, and one aiming at the US Congress led to the “Cost of Knowledge” petition in 2010. This petition called for a boycott to write, review or make editorial work for the company. It was signed by tens of thousands of academics and led to some mocking of Elsevier logo.



Michael Eisen, CC BY 3.0 <https://creativecommons.org/licenses/by/3.0>, via Wikimedia Commons

To sum it up, If the kids of Bruno Latour had been STS PhD students in 2010, they would probably have authored a paper entitled “Portrait of a publisher as a wild capitalist”. But that wouldn’t have predicted what happened next.

From academic publisher to data company: a very public transition

In fact, Elsevier continued to thrive as a publisher despite the tens of thousands of petitioners. But the company has changed its core business and has significantly expanded its range of services until it no longer appears as a publisher. Take two exemplary acquisitions: in 2013,

Elsevier purchased Mendeley, a library management service, and for some, it was as if the Empire had bought the rebels.

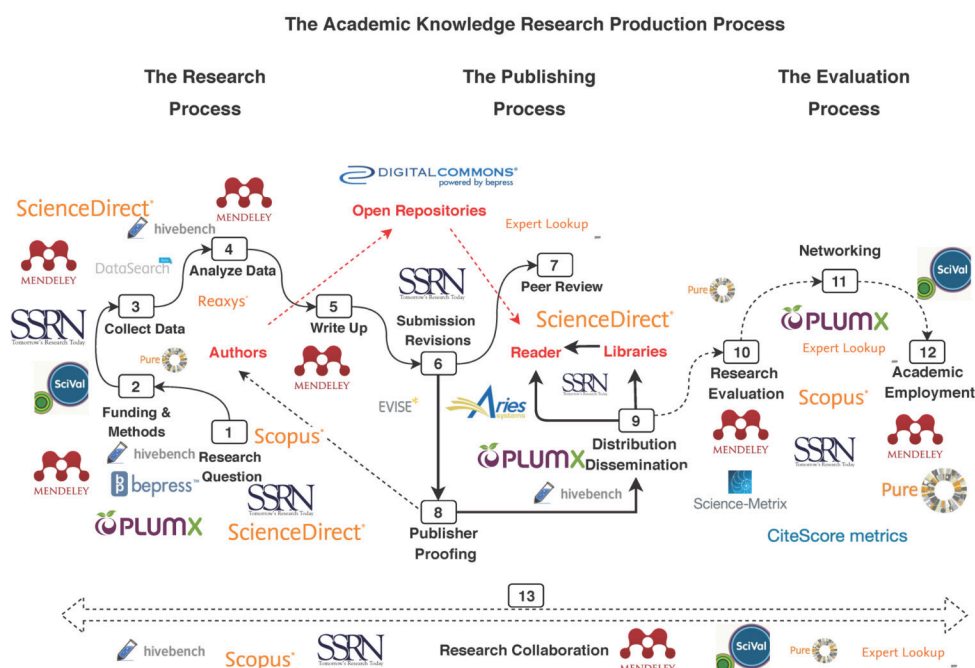
This was me long ago when Mendeley sold off to Elsevier. They used to have the best desktop reference manager, but now it's clear that [@zotero](https://t.co/8YvTVs1Y8A) is miles ahead. <https://t.co/8YvTVs1Y8A> pic.twitter.com/TmmYMs3hgb

— Timothée Poisot, Ph.D. (@tpoi) [July 4, 2022](#)

Elsevier had two objectives: on the one hand, to extend its information retrieval ecosystem, and on the other, to collect data on Mendeley users, potentially authors and reviewers. These same objectives were reflected in the acquisition two years later of SSRN, a preprint platform then specialising in the social sciences.

“Elsevier is now getting closer and closer to researchers with business models that don’t involve libraries,” says Joe Esposito, a publishing consultant in New York City. “The positioning is well thought out: lock up revenues to the legacy publishing business, move into areas where piracy is not much of an issue, create deeper relationships with researchers and become more and more essential to researchers even as librarians become less so.”³

The series of purchases aim to control the bricks directly used by researchers whose research projects, results, research data, texts read, cited, reviewed, tweeted, etc were linked and. Elsevier being able to identify them. But as the 2019 comprehensive diagram below shows, they are not the only target of the “new Elsevier”.



Chen, G., Posada, A., & Chan, L. 2019. Vertical Integration in Academic Publishing : Implications for Knowledge Inequality. In Chan, L., & Mounier, P. (Eds.), *Connecting the Knowledge Commons — From Projects to Sustainable Infrastructure : The 22nd International Conference on Electronic Publishing – Revised Selected Papers*. Marseille : OpenEdition Press.

In fact, Elsevier's other target market is higher education and research institutions, and even governmental institutions. The enclosure of the Elsevier ecosystem has, for example, guaranteed the company a position as a subcontractor in the construction of the first European open access monitor, which has been deemed scandalous by open access activists⁴. When a consortium of Dutch universities signed a transformative agreement with the publisher in 2019, this included the joint development of projects involving all kinds of data, being a Faustian pact with Lucifer where open science becomes sustaining Elsevier data infrastructure in exchange for open access papers⁵.

In a decade, Elsevier had become a data company, selling them to numerous clients, both academic and non-academic and defining itself in corporate documents as "a global leader in information and analytics, (which) helps researchers and healthcare professionals advance science and improve health outcomes for the benefit of society.", while making videos on the perfect world of information it designs with a product called PURE.

Naming a new supervillain : surveillance publishing

As of 2019, this transformation of Elsevier and, to a lesser degree, other big publishers, was a wake-up call for various institutions and authors. They started to formalise the list of new dangers that the construction of data tracking and information aggregation systems constituted, some of them specific to the academic world and others similar to those created by GAFAM-like companies. For example, a commission of the DFG published a briefing paper where they acted as alarm raisers for the following concerns⁶:

1. *entail a violation of academic freedom and the freedom of research and teaching;*
2. *constitute a violation of the right to the protection of personal data;*
3. *pose a potential threat to scientists, as the data could also become accessible to foreign governments and authoritarian regimes;*
4. *constitute an encroachment of competition law, as new participants barely have a chance to enter the market;*
5. *favour a reduction in the value of public research investment, since data on research activity can be collected by commercial research competitors or made available to them in return for payment in connection with industrial espionage.*

These fears may seem hypothetical, but the fact, for example, that Elsevier's parent company, RELX, has signed a huge contract to supply personal data to the US Immigration and Customs Enforcement agency has given some weight to their warnings. But what data are we talking about? Two facetious colleagues used the provisions of the GDPR to ask Elsevier for their data and documented their findings on the traces of their stay in the Elsevier Hotel. It contains a number of directly personal data (phone numbers, bank details, addresses), but above all a great deal of usage data on the opening of e-mails sent by the company, the most basic operations on Mendeley and Science Direct or, more amusingly or worryingly, the trace of persons consents and non-consents:

CONSENT	NOCONSENT	UNKNOWNCONSENT
0	1	0
0	0	1
0	0	1
0	0	0
0	0	0
0	0	0
0	0	0
0	0	1
0	0	1
0	0	0
0	0	1
0	0	1
0	0	1
0	0	0
0	1	0
0	0	1
0	0	0
0	0	0

Eiko I. Fried, Robin Niels Kok, [Welcome to Hotel Elsevier: you can check-out any time you like ... not](#), 2022

A new petition, twelve years after The Cost of knowledge, calls to “Stop Tracking Science“, which actually means stop tracking academics. In the new configuration, libraries are still a passage point between the big publishers and the researchers though not exclusive anymore. But the data that goes for these exchanges is now considered in a different manner: “they are even attempting to persuade libraries to install trackers inside university networks: the research behavior of all of us is being recorded in real time.” While identification has for long been presented as a necessary security to provide access to closed texts, it is now a source of concern, in a very similar manner as cell phone or internet tracking. To describe this phenomenon, several labels have been proposed, such as ”platformization of science”⁷ or “surveillance publishing”⁸ The big publishers are trying through various legal actions to present Sci-hub and Libgen not only as IP offenders, but also as dangerous hackers for the security of research institutions., while it is the same accusation that is directed towards them. So, in the end, for you, which ones are the supervilains threatening academic communities?

Endnotes

1. Sismondo S (2007) Ghost Management: How Much of the Medical Literature Is Shaped Behind the Scenes by the Pharmaceutical Industry? PLoS Med 4(9): e286
2. Van Noorden, R. Social-sciences preprint server snapped up by publishing giant Elsevier. *Nature* (2016). [
3. see for example Jonathan Tennant, & Björn Brembs. (2018, October 26). RELX referral to EU competition authority. Zenodo
4. This deal has been the object of a very long post. Also see this SPARC analysis
5. Data tracking in research: aggregation and use or sale of usage data by academic publishers. A briefing paper of the Committee on Scientific Library Services and Information Systems of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) 28 October 2021
6. Kunz, Raffaella: Threats to Academic Freedom under the Guise of Open Access: The Power of Publishers, Data Tracking in Science, and the Responsibilities of Public Actors, VerfBlog, 2022/3/18
7. Pooley, J., (2022) Surveillance Publishing, *The Journal of Electronic Publishing* 25(1)