The Scholarly Fingerprinting Industry

Jefferson Pooley

Muhlenberg College

pooley@muhlenberg.edu

jeffpooley.com

The Scholarly Fingerprinting Industry

Jefferson Pooley

Elsevier, Taylor & Francis, Springer Nature, Wiley, and SAGE: Many researchers know that the five giant firms publish most of the world's scholarship. Fifty years of acquisitions and journal launches have yielded a stunningly profitable oligopoly, built up from academics' unpaid writing-and-editing labor. Their business is a form of IP rentiership—collections of title-by-title prestige monopolies that, in the case of *Nature* or *The Lancet*, underwrite a stable of spinoff journals on the logic of the Hollywood franchise.

Less well-known is that Elsevier and its peers are layering a second business on top of their legacy publishing operations, fueled by data extraction. They are packaging researcher behavior, gleaned from their digital platforms, into prediction products, which they sell back to universities and other clients. Their raw material is scholars' citations, abstracts, downloads, and reading habits, repurposed into dashboard services that, for example, track researcher productivity. Elsevier and the other oligopolist firms are fast becoming, in other words, surveillance publishers (Pooley). And they are using the windfall profits from their existing APC-and-subscription business to finance their moves into predictive analytics.

Elsevier is the farthest along. In 2015, its parent company RELX Group announced its "transformation" from publisher to a "technology, content and analytics-driven business," adding that the firm is "systematically migrating all of our businesses towards electronic decision tools" (RELX Group, *Annual Report 2014* 5, 4). By then, Elsevier's decade-

26 Apart from contacting your own university library (the OA office or your subject librarian), please feel free to get in touch with us if you are looking for more information on questions raised in this article. The team of the Fachinformationsdienst Anglo-American Culture (FID AAC) is looking forward to entering into conversations from which both sides can benefit.

long acquisition binge, up and down the research lifecycle, was already underway. In the past decade, it acquired Pure (2012), Mendeley (2013), Newsflo (2015), SSRN (2016), bepress (2017), Parity Computing (2019), and, in spring 2022, Interfolio, the "Faculty Information System" provider. Together with ScienceDirect, the firm's web-based journal delivery platform, and Scopus, its citation index, Elsevier has assembled a portfolio of knowledge products that spans lab software to research assessment. These are, in a sense, services with benefits: reference management from Mendeley and journal access from ScienceDirect both furnish scholars' behavioral data back to Elsevier. The company then sells the processed data back to universities and other clients in the form of "research intelligence," i. e., prediction products like SciVal and Pure that score researcher impact and productivity.

Elsevier, to borrow a computing phrase, has become a full-stack publisher. Its thousands of journals might be seen as data-delivery vehicles—in themselves and by way of trackable engagement. Though some of these researcher-facing services are costly indeed, the core dynamic is not unlike the surveillance businesses built by Google and Facebook (Zuboff). The key difference is that Elsevier gets to charge its customers twice, first through sky-high subscription-and-APC rates and, secondly, for the "decision tools" generated by the legacy business's behavioral surplus (RELX Group, *Annual Report 2021* 5). As CUNY law professor Sarah Lamdan put it in a 2021 talk, "[y]our journals are spying on you" (*Your Journals*). Earlier this year, internet sleuths discovered that Elsevier had embedded a per-download tracker in its PDF metadata (Hansen). Psychologist Eiko Fried followed up with a GDPR data request, which yielded a spreadsheet haul of torrential size. The company, Fried revealed, is tracking article engagement at the granularity of specific image views. The precise ways that these and other data are mined, sorted, and processed into prediction products like SciVal is, of course, shrouded in proprietary secrecy. Elsevier touts what it calls its Fingerprint® Engine, which applies machine learning to its vast trove of researcher data ("signals") to assign, for example, a list of weighted concepts to a particular researcher (Picadio). As the RELX Group boasts in its latest annual report, the company's "research intelligence portfolio"—sold to university management, corporate R&D executives, funders, and policy-makers—now generates over a third of Elsevier's revenue (*Annual Report 2021* 21, 23). The company states that it expects to improve on its 2021 profit margin which, at 38 percent, places Elsevier among the world's most lucrative businesses.

The other publishing colossi are playing catch up. Taylor & Francis, a unit of the UK-based intelligence conglomerate Informa Group, has been expanding its "knowledge services" through acquisitions like the Faculty of 1000 platform last year (*Annual Report 2021* 51-55). The division's profit margin, at 37 percent, was just hairs off the Elsevier

pace (51). Wiley, meanwhile, recently rolled out its journal platform Literatum, built by the software firm it acquired in 2016, Atypon. "Know thy reader," reads the firm's pitch. "Literatum's analytics module tracks and combines publishing-specific content usage data with readers' site behavior" (Atypon). Wiley's margin last year was 35 percent (John Wiley & Sons 32). Springer Nature's parent company, Holtzbrinck, for its part, owns its own full-stack research lifecycle offerings, including the Scopus competitor Dimensions, Pure competitor Symplectic, impact tracker Altmetric, and data repository figshare (Holtzbrinck).

Elsevier's main competitor, tellingly, is Clarivate, a firm that began as the Institute for Scientific Information (ISI) in the late 1950s (Wouters). ISI's founder, Eugene Garfield, helped establish the field of bibliometrics through the company's Science Citation Index. In 2016, ISI was spun off as Clarivate in a $3.5 billion private equity deal, with Garfield's citation index—renamed Web of Science—the new company's crown jewel (Clarivate 5, 12–13). Sold to over 9,000 universities and other customers, Web of Science builds on what was, in Garfield's citation graph, the original academic prediction product. What Clarivate is selling, after all, is bets on future scholarly productivity and impact. A key growth strategy, the company states, is "moving up the value chain by providing our customers with predictive and prescriptive analytics" (Clarivate 10). Late last year Clarivate—which reported an astonishing 42 percent profit margin—acquired ProQuest, the sprawling library vendor, for over $5 billion (Clarivate 9, 13). The data generated from ProQuest's library products will almost certainly feed Clarivate's own "research intelligence" offerings, Converis and InCites. If anything, Elsevier's leg up on Clarivate has been its access to the rich behavioral surplus produced by its publishing business.

More acquisitions and inter-firm jockeying will proceed at the pace of Wall Street. What is fast emerging is a small band of vertically integrated knowledge brokers, most of them, in Björn Brembs's phrase, "corporations formerly known as publishers" ("Off to Paris"). Elsevier and its peers, indeed, have used their enormous publishing profits to finance their full-stack acquisitions. In that respect, surveillance publishing is an insult-to-injury story. Scholars justly complain about the insanely lucrative scholarly publishing industry, whose subscription and APC windfalls are made off their unpaid labor. Now Wiley and the others are extracting a second rent, without the consent or notice of scholars.

Most scholars, after all, have no idea that their behavioral cream is getting skimmed for profit. If widely exposed, these next-level predations could build momentum for a nonprofit, academy-led alternative to the oligopolists. As historian Aileen Fyfe has chronicled, the current joint-custody arrangement—nonprofit universities and for-profit publishers—is a recent and reversible development. A community-owned infrastructure is, with slow care, getting built out, with the aim to support new and established scholar-led publishing initiatives. Another

scholarly communication world really is possible. We need, however, researcher buy-in in light of predictable—if short-run—prestige penalties; funders and librarians, too, must be shaken from their APC-and-subscription slumbers. The emerging surveillance publishing economy, in that respect, is an opportunity of sorts. A range of scholar-critics, including Renke Siems, George Chen, Leslie Chan, Björn Brembs ("Algorithmic Employment"), and Sarah Lamdan (*Data Cartels*), have begun to sound the alarm. Our task is to amplify their accounts—to spread the word about surveillance profits—in support of the campaign to restore custody over scholarly publishing.

Works Cited

Alfano, Veronica, and Andrew Stauffer, eds. *Virtual Victorians: Networks, Connections, Technologies*. New York: Palgrave, 2015. Print.

Aljoe, Nicole, and Elizabeth Maddock Dillon. "Decolonizing the Archive: Remix and Reassembly." *Early Caribbean Digital Archive*. Northeastern University, n. d. Web. 15 Dec. 2022. https://ecda.northeastern.edu/home/about/decolonizing-the-archive/.

Allington, Daniel, Sarah Brouillette, and David Golumbia. "Neoliberal Tools (and Archives): A Political History of Digital Humanities." *Los Angeles Review of Books* 1 May 2016. Web. 28 July 2022. https://lareviewofbooks.org/article/neoliberal-tools-archives-political-history-digital-humanities/.

Ancion, Zoé, et al. Action Plan for Diamond Open Access. *Zenodo*, 2022. Web. 26 Sept. 2022. https://zenodo.org/record/6282403#.Y1A2Gy8RppQ.

Atypon. "Analytics." *Atypon*, n. d. Web. 20 Aug. 2022. https://www.atypon.com/products/literatum/analytics/.

AuROA: *Publizieren und Open Access in den Geisteswissenschaften: Erkenntnisse aus dem Projekt AuROA zu den Stakeholdern im Publikationsprozess*. Essen, 2022. Web. 16 Sept. 2022. https://projekt-auroa.de/wp-content/uploads/2022/03/AuROA-Publizieren-und-Open-Access-in-den-Geisteswissenschaften.pdf.

Bargheer, Margo, and Kizer Walker. "Library Publishing and the University Press in the United States and Germany: Lessons from Two Academic Contexts for

Sustaining the Scholarly Book." *Bibliothek: Forschung und Praxis* 41.3 (2017): 291-307. Web. 15 Dec. 2022. https://doi.org/10.1515/bfp-2017-0037.

Barnett, Fiona, Zach Blas, Micha Cárdenas, Jacob Gaboury, Jessica Marie Johnson, and Margaret Rhee. "QueerOS: A User's Manual." *Debates in Digital Humanities*. Ed. Matthew K. Gold and Lauren Klein. Minneapolis: U of Minnesota P, 2016. 50-59. Print.

Bertram, Lillian-Yvonne. *Travesty Generator*. Blacksburg, VA: Noemi, 2019. Print.

Bode, Katherine. "The Equivalence of 'Close' and 'Distant' Reading; or, Toward a New Object for Data-Rich Literary History." *Modern Language Quarterly* 78.1 (2017): 77-106. Web. 4 Sept. 2022. https://doi.org/10.1215/00267929-3699787.

Brembs, Björn. "Algorithmic Employment Decisions in Academia?" *björn.brembs. blog*. Björn Brembs, 23 Sept. 2021. Web. 12 Sept. 2022. http://bjoern.brembs. net/2021/09/algorithmic-employment-decisions-in-academia/.

---. "Off to Paris for #FENS2022 with Two Posters." *björn.brembs.blog*. Björn Brembs, 8 July 2022. Web. 12 Sept. 2022. http://bjoern.brembs.net/2022/07/ off-to-paris-for-fens2022-with-two-posters/.

Bruns, Andre, et al. *Diamond Open Access Journals Germany (DOAG)*. Bielefeld University, 2022. Web. 26 Sept. 2022. https://pub.uni-bielefeld.de/ record/2963331.

Carroll, Stephanie Russo, et al. "The CARE Principles for Indigenous Data Governance." *Data Science Journal* 19.1 (2020): 1-12. Web. 15 Dec. 2022. http:// dx.doi.org/https://doi.org/10.5334/dsj-2020-043.

Chen, George, and Leslie Chan. "University Rankings and Governance by Metrics and Algorithms." *Research Handbook on University Rankings*. Ed. Ellen Hazelkorn and Georgiana Mihut. Cheltenham: Edward Elgar, 2021. 425-43. Print.

Clarivate. "Form 10-K." *1-153*. Web. 12 Sept. 2021. https://s25.q4cdn. com/843006813/files/doc_downloads/2022/05/2021_12-Clarivate-Plc-FSs-DOC-10K-(32).pdf.

Clement, Tanya E., et al. "Collaborators' Bill of Rights." *Humanities Commons*. Humanities Commons, 2021. Web. 28 July 2022. https://hcommons.org/ deposits/item/hc:31187/.

"Colored Conventions Project." *Colored Conventions Project*. Web. 28 July 2022. https://coloredconventions.org.

Deutsche Forschungsgemeinschaft. *Academic Publishing as a Foundation and Area of Leverage for Research Assessment: Challenges and Fields of Action*. Bonn, 2022. Web. 26 Sept. 2022. https://www.dfg.de/download/pdf/ foerderung/grundlagen_dfg_foerderung/publikationswesen/positions papier_publikationswesen_en.pdf.

Drucker, Johanna. "Humanities Approaches to Graphical Display." Digital Humanities Quarterly 5.1 (2011): n. pag. Web. 15 Dec. 2022. http://www.digital humanities.org/dhq/vol/5/1/000091/000091.html.

Elsevier. "Elsevier Fingerprint Engine." *Elsevier*. Elsevier, n. d. Web. 12 Sept. 2021. https://www.elsevier.com/solutions/elsevier-fingerprint-engine.

Fried, Eiko. "Welcome to Hotel Elsevier: You Can Check-Out Any Time You Like ... Not." *Eiko-fried.com*. Eiko Fried, 9 May 2022. Web. 12 Sept. 2022. https://eiko-fried.com/welcome-to-hotel-elsevier-you-can-check-out-any-time-you-like-not/.

Foreman, P. Gabrielle, et al. *Colored Conventions Project*. Web. 2 Dec. 2022. https://coloredconventions.org/.

Fyfe, Aileen. "Self-Help for Learned Journals: Scientific Societies and the Commerce of Publishing in the 1950s." *History of Science* 60.2 (2022): 255-79. Web. 15 Dec. 2022. https://doi.org/10.1177/0073275321999901.

Galey, Alan, and Stan Ruecker. "How a Prototype Argues." *Literary and Linguistic Computing* 25.4 (2010): 405-24. Web. 15 Dec. 2022. https://doi.org/10.1093/ llc/fqq021.

Gallon, Kim. "Making a Case for the Black Digital Humanities." *Debates in Digital Humanities 2016*. Ed. Matthew K. Gold and Lauren Klein. Minneapolis: U of Minnesota P, 2016. 42-49. Print.

Greenspan, Brian. "The Scandal of Digital Humanities." *Hyperlab 23*. Carleton University, Jan. 2018. Web. 6 Sept. 2022. https://carleton.ca/hyperlab/2018/the-scandal-of-digital-humanities/.

Grusin, Richard. "The Dark Side of Digital Humanities: Dispatches from Two Recent MLA Conventions." *Differences* 25.1 (2014): 79-92. Web. 15 Dec. 2022. https://doi.org/10.1215/10407391-2420009.

Hansen, Morten. "Building Education Assets, One Crumb at a Time." *The Post-Pandemic University* 20 Mar. 2022. Web. 12 Sept. 2022. https://postpandemicuniversity.net/2022/03/20/building-education-assets-one-crumb-at-a-time/.

Hayles, N. Katherine. *Writing Machines*. Cambridge, MA: MIT, 2002. Print.

Herrmann, Sebastian M. *Data Imaginary: Literature and Data in Nineteenth-Century US Culture*. Heidelberg: Winter, 2022. Print.

Holtzbrinck Publishing Group. "About Us." *Holtzbrinck Publishing Group*. Georg von Holtzbrinck GmbH & Co., n. d. Web. 12 Sept. 2022. https://www.holtzbrinck.com.

HumanitiesCommons. n. d. Web. 29 Sept. 2022. https://hcommons.org.

"HuMetricsHSS." *HuMetricsHSS*. HuMetricsHSS: Humane Metrics Initiative. HuMetricsHSS, 2022. Web. 28 July 2022. https://humetricshss.org.

Informa Group. *Annual Report 2021: Digital & Data Acceleration. London: Informa Group*, 2022. Web. 12 Sept. 2012. https://www.informa.com/globalassets/documents/investor-relations/2022/informa-annual-report-2021.pdf.

John Wiley & Sons. "Form 10-K." *2022, 1-111*. Web. 12 Sept. 2022. https://s27.q4cdn.com/812717746/files/doc_financials/2022/q4/Wiley-10K-Annual-Report.pdf.

Johnson, Jessica Marie. "Markup Bodies: Black [Life] Studies and Slavery [Death] Studies at the Digital Crossroads." *Social Text* 137 36.4 (2018), 57-79. Print.

Kartheus, Wiebke, and Dorothea Schuller. "Open Access in English and American Studies." *Open Access Network*. open-access.network, 18 July 2022. Web. 26. Sept. 2022. https://open-access.network/en/information/subject-specific-open-access/english-and-american-studies.

Kirschenbaum, Matthew. "What Is Digital Humanities and What's It Doing in English Departments?" *Debates in Digital Humanities*. Ed. Matthew K. Gold. Minneapolis: U of Minnesota P, 2012: 3-11. Print.

Kreutzer, Till. *Open Content: Ein Praxisleitfaden zur Nutzung von Creative-Commons-Lizenzen*. 2nd ed. Bonn: Deutsche UNESCO-Kommission e. V., 2016. Web. 24 Sept. 2022. https://irights.info/wp-content/uploads/2015/10/Open_Content_-_Ein_Praxisleitfaden_zur_Nutzung_von_Creative-Commons-Lizenzen.pdf.

Lamdan, Sarah. *Data Cartels: The Companies That Control and Monopolize Our Information*. Stanford, CA: Stanford UP, 2022. Print.

---. "Your Journals Are Spying on You: Research Surveillance in Library Products." Videotaped Presentation, Indiana University Bloomington Libraries, 22 Oct. 2021. Web. 12 Sept. 2022. https://media.dlib.indiana.edu/media_objects/76537m18z.

Lee, Maurice. "Evidence, Coincidence, and Superabundant Information." *Victorian Studies* 54.1 (2011): 87-94. Web. 25 June 2016. https://doi.org/10.2979/victorianstudies.54.1.87.

McPherson, Tara. "Why Are the Digital Humanities So White? Or Thinking the Histories of Race and Computation." *Debates in Digital Humanities*. Ed. Matthew K. Gold. Minneapolis: U of Minnesota P, 2012: 139-60. Print.

Monks-Leeson, Emily. "Archives on the Internet: Representing Contexts and Provenance from Repository to Website." *The American Archivist* 74.1 (2011): 38-57. JSTOR Web. 15 Dec. 2022. https://www.jstor.org/stable/23079000.

Moretti, Franco. "The Roads to Rome: Literary Studies, Hermeneutics, Quantification" *New Left Review* 124 (2020): 125-36. Web. 16 Dec. 2022. https://newleftreview.org/issues/ii124/articles/franco-moretti-the-roads-to-rome.

Morrison, Toni. *Beloved*. 1987. New York: Vintage Classics, 2007. Print.

O'Donnell, Daniel. "'There's No Next about It': Stanley Fish, William Pannapacker, and the Digital Humanities as Paradiscipline." *Dpod Blog*. Internet Archive, 22 June 2016. Web. 13 Dec. 2022. https://web.archive.org/web/20160310141643/http:/dpod.kakelbont.ca/2012/06/22/theres-no-next-about-it-stanley-fish-william-pannapacker-and-the-digital-humanities-as-paradiscipline/.

Pannapacker, William. "The MLA and the Digital Humanities." *The Brainstorm Blog: The Chronicle of Higher Education Online*. Archive.org, 28 Dec. 2009. Web. 28 Sept. 2022. https://web.archive.org/web/20121004134824/http:/chronicle.com/blogPost/The-MLAthe-Digital/19468/.

Picadio, Doug. "Fingerprinting: What Is It, and How Can I Use It." *Presentation. Pure International Conference, Barcelona, 10 Oct. 2017*. Web. 12 Sept. 2022. https://www.elsevier.com/__data/assets/pdf_file/0004/525613/Day1_Sala3_11_50_D_Picadio.pdf.

Pooley, Jefferson. "Surveillance Publishing." *The Journal of Electronic Publishing* 25.1 (2022): 39-49. Web. 15 Dec. 2022. https://doi.org/10.3998/jep.1874.

Ramsay, Stephen. "On Building." *Defining Digital Humanities: A Reader*. Ed. Melissa Terras et al. Farnham: Ashgate, 2013. 243-45. Print.

RELX Group. "Annual Report and Financial Statements 2014." *London: RELX Group, 2015*. Web. 3 Oct. 2022. https://www.relx.com/~/media/Files/R/RELX-Group/documents/reports/annual-reports/2014-annual-report.pdf.

---. "Annual Report and Financial Statements 2021." *London: RELX Group, 2022*. Web. 12 Sept. 2022. https://www.relx.com/~/media/Files/R/RELX-Group/documents/reports/annual-reports/relx-2021-annual-report.pdf.

Roth, Camille. "Digital, Digitized, and Numerical Humanities." *Digital Scholarship in the Humanities* 34.3 (2019): 616-632. Web. 13 Dec. 2022. https://academic.oup.com/dsh/article-abstract/34/3/616/5161109.

Schöch, Christof. "Big? Smart? Clean? Messy? Data in the Humanities." *Journal of Digital Humanities* 2.3 (2013): 2-13. Web. 13 Dec. 2022. http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/.

Siems, Renke. "When Your Journal Reads You: User Tracking on Science Publisher Platforms." *Elephant in the Lab*. Zenodo, 14 Apr. 2021. Web. 12 Sept. 2022. https://zenodo.org/record/4683778#.Y1Aoxi8RpQI.

Smith, Henry Nash. "Can 'American Studies' Develop a Method?" *American Quarterly* 9.2 (1957): 197-208. *JSTOR* Web. 15 Dec. 2022. https://doi.org/10.2307/2710743.

University of Oxford. "OA Publication Routes." *Open Access Oxford*. The University of Oxford, 2013. Web. 26 Sept. 2022. https://openaccess.ox.ac.uk/open-publication-routes/.

Wilkinson, Mark D., et al. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3.1 (2016). Web. 15 Dec. 2022. https://www.nature.com/articles/sdata201618?ref=https://githubhelp.com.

Wise, Gene. "'Paradigm Dramas' in American Studies: A Cultural and Institutional History of the Movement." *American Quarterly* 31.3 (1979): 293-337. *JSTOR* Web. 15 Dec. 2022. https://doi.org/10.2307/2712378.

Wissenschaftsrat. "Empfehlungen zur Transformation des wissenschaftlichen Publizierens zu Open Access." Köln, 2022. Web. 26. Sept. 2022. https://www.wissenschaftsrat.de/download/2022/9477-22.html.

Wouters, Paul. "Eugene Garfield (1925–2017)." *Nature* 543 (2017): 492. Web. 12 Sept. 2022. https://www.nature.com/articles/543492a.

Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: Public Affairs, 2019. Print.