Chairman McCormick, Ranking Member Sykes, Distinguished Committee Members, thank you for inviting me here today. I am Jason Owen-Smith, Executive Director of the Institute for Research on Innovation and Science at the University of Michigan, a research data scientist specializing in science and innovation, the principal investigator on federal and privately funded scientific research grants, and a founder of national data infrastructure that clarifies the public benefits of university R&D.

Scholarly Publishing is a National Strategic Asset.

The US government invested $190 per person in academic R&D in 2024[1] because, as Vannevar Bush put it in 1945, "without scientific progress, no amount of achievement in other directions can secure our health, wealth, and security."[2] In other words, the benefits that result from the $4,831 per person that our country spent on health care[3] or the $2,499 per person America spend on national defense[4] that year depend on the vast stores of knowledge, technology, and expertise that public research investments produce.

Scholarly publication is one way to realize the value of taxpayer-funded research. Data about publications and their use are a treasure trove of information about research in important fields – such as Quantum Computing, AI, Hypersonics, or Synthetic Biology – about research work – such as collaboration, problem definition, and experimental design and about researchers themselves. Data about research and researchers could be national strategic assets used to identify promising areas for investment, amplify the work of America's rising scientific stars, bolster the distinctive strengths of our world class universities and national labs, help science agencies build evidence-based portfolios to pursue national priorities, or accelerate ambitious programs such as the Genesis Mission.[5] Instead, they are almost exclusively controlled, monetized, and frequently exported by private, often foreign-owned, entities whose policies and practices are beyond the scope of

---

[1] Total federally funded academic research expenditures in FY 2024 from the NSF Higher Education Research and Development HERD survey were $64.717 billion (NCSES, 2025). The US Census Bureau's annual population estimate for that year was 340,110,998 people (US Census Bureau, NST-EST2024-ALLDATA).

[2] Bush, Vannevar. 1945. *Science the Endless Frontier (75th Anniversary Edition). Washington, DC. National Academies Press*. P. 8.

[3] Center For Medicare and Medicaid Services. 2024. National Health Expenditures Fact Sheet. https://www.cms.gov/data-research/statistics-trends-and-reports/national-health-expenditure-data/nhe-fact-sheet. Accessed 04/10/2026.

[4] Congressional Budget Office, 2024. Discretionary Spending in FY 2024, An Infographic. https://www.cbo.gov/publication/61184. Accessed 04/10/2026

[5] Department of Energy. N.D. "The Genesis Mission: A National Initiative to Build the World's Most Powerful Scientific Platform." https://www.energy.gov/undersecretaryforscience/genesis-mission/genesis-mission. Accessed 04/12/2026.

public oversight.[6] This lack of transparency creates substantial risks for the talented people and frontier discoveries that are so central to our health, prosperity and security.

Major scholarly publishers do far more than publish research. They have become for-profit data brokers,[7] sophisticated research analytics companies whose industry position allows them to link and mine researchers' data without their informed consent, without federal oversight, and without apparent regard for potential risks to US research and national security. The resources that make this business model possible are largely paid for by American taxpayers. The work required to create them is largely done by US researchers. Yet almost all the resulting value is captured by a few, mostly foreign-owned companies.

We need to rethink public access to and use of these national strategic assets.

My recommendations aim to reduce unnecessary risks to the security of American research and the privacy of American researchers while spurring development of the accountable public infrastructure needed to make data built largely with US resources work for the American people.

This Committee could make concrete progress in both directions by taking two steps. First, draw on bipartisan policy statements – such as the 2022 OSTP Public Access memo[8], 2025 OMB OPEN Government implementation guidance[9], and President Trump's 2025 Restoring Gold Standard Science Executive Order[10] – to provide a regulatory framework for responsible, transparent public access to data about federal research and researchers. Second, support the use of proven public platforms – such as Vivo,[11] IRIS,[12] the National

---

[6] Winter, Caroline. 03/17/2023. "Market Consolidation and Scholarly Communication." Open Scholarship Policy Observatory. https://doi.org/10.25547/CZ92-S769.

[7] Lamdan, Sarah. 2022. *Data Cartels: The Companies that Control and Monopolize Our Information*. Stanford, CA. Stanford University Press.

[8] White House Office of Science & Technology Policy. 08/25/2022. "Ensuring Free, Immediate, and Equitable Access to Federally Funded Research." https://bidenwhitehouse.archives.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf. Accessed 04/10/2026.

[9] Executive Office of the President, Office of Management and Budget. 01/15/2025. "Phase 2 Implementation of the Foundations for Evidence Based Policymaking Act of 2018: Open Government Data Access and Management Guidance." https://bidenwhitehouse.archives.gov/wp-content/uploads/2025/01/M-25-05-Phase-2-Implementation-of-the-Foundations-for-Evidence-Based-Policymaking-Act-of-2018-Open-Government-Data-Access-and-Management-Guidance.pdf. Accessed 04/10/2026.

[10] President Donald J. Trump. 05/23/2025. Executive Order: Restoring Gold Standard Science. https://www.whitehouse.gov/presidential-actions/2025/05/restoring-gold-standard-science/. Accessed 04/10/2026.

[11] VIVO: Connect, Share, Discover. 2024. https://vivoweb.org/. Accessed 04/10/2026.

[12] Institute for Research on Innovation and Science. 2026. Regents of the University of Michigan. https://iris.isr.umich.edu/, Accessed 04/10/2026.

Data Platform,[13] the California Digital Library,[14] and the Statewide Longitudinal Data Systems[15] – to create the core of a US-based public data system that will ensure data about research and researchers becomes a first-class strategic asset for the nation.

Major Publishers Have Become Sophisticated Data Analytics Companies

Large, for-profit publishers were once oriented toward review, editing, and dissemination of finished pieces of research such as scholarly articles. Today, their high-tech research analytics enterprises often depend on sophisticated, AI-enabled products and pay-for-access data models.

Let me share an example with you. According to their parent company's financial reports, Elsevier, the world's largest scientific publisher, ". . . helps advance science and healthcare by combining high-quality, trusted scientific and medical information and data sets with innovative technologies to deliver critical insights that support better outcomes."[16] The image in Figure 1 is one of the company's visual summaries of its AI-enabled analytic pipeline. Massive amounts of data flow in from many sources, cleaned, integrated information suitable for further processing or consumption by clients flow out.

At Elsevier, such data may include publications and citations, researchers' clicks, downloads, manuscript submissions, and reviews,[17] literature searches on company platforms like SciVal, or digital traces of work done using the collaborative bibliographic tool, Mendeley.[18] Preprint services, like Elsevier's SSRN, can indicate which literatures, authors and ideas researchers follow most closely.[19] Researcher trace data like these can assemble a detailed picture of professional and scientific work. When combined for many

[13] National Data Platform. N.D. https://nationaldataplatform.org/documentation/. Accessed 04/10/2026.

[14] California Digital Library. 2026. Regents of the University of California. https://cdlib.org/. Accessed 04/10/2026.

[15] National Center for Education Statistics. "Statewide Longitudinal Data Systems Program." Institute of Educational Sciences. https://nces.ed.gov/Programs/SLDS/. Accessed 04/10/2026.

[16] RELX 2025 Annual Report. https://www.relx.com/~/media/Files/R/RELX-Group/documents/reports/annual-reports/relx-2025-annual-report.pdf p. 8. Accessed 04/10/2026.

[17] Hinchliffe, Lisa Janicke. 02/06/2017. "Making a Few Elsevier Predictions." https://lisahinchliffe.com/2017/02/06/elsevier-predictions/. Accessed 04/06/2026.

[18] Esposito, Joseph. 01/272015. "When is a Feature a Product, and a Product a Business?" The Scholarly Kitchen. https://scholarlykitchen.sspnet.org/2015/01/27/when-is-a-feature-a-product-and-a-product-a-business/. Accessed 04/10/2026.

[19] Schonfeld, Roger C. 02/09/2017. "When is a Publisher Not a Publisher? Cobbling Together the Pieces to Build a Workflow Business." The Scholarly Kitchen. https://scholarlykitchen.sspnet.org/2017/02/09/cobbling-together-workflow-businesses/. Accessed 04/10/2026.

researchers such data become a powerful basis for many types of analyses. That power, Figure 1 suggests, is precisely what Elsevier sells.

Elsevier's 2022 acquisition of the faculty information tool *Interfolio* went even further by

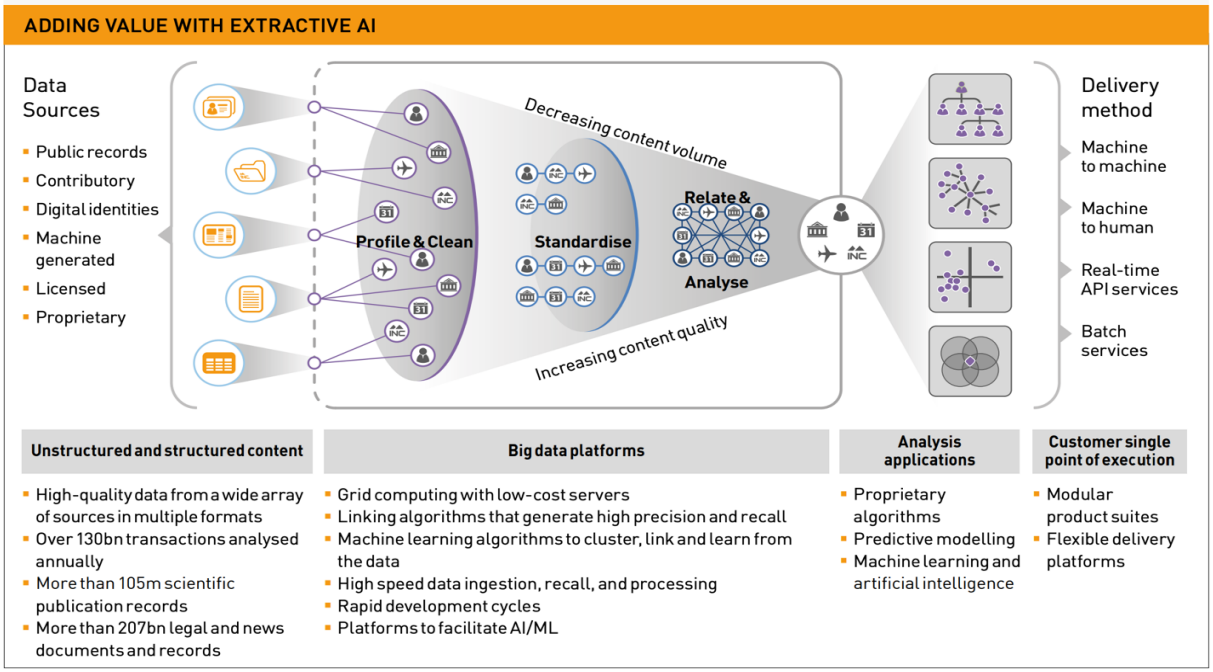

Figure 1. Data Pipeline Summary from RELX Annual Financial Report, 2025. Page 7. https://www.relx.com/~/media/Files/R/RELX-Group/documents/reports/annual-reports/relx-2025-annual-report.pdf. Accessed 04/10/2026

adding a product used for academic hiring, annual performance reviews, promotion and tenure decisions. *Interfolio* combines university HR and sponsored projects systems data with documents produced by faculty, such as personal statements, research plans, copies of prepublication works like grant proposals, and reference letters that typically include frank, detailed evaluations of in progress research.[20]

Potentially affected researchers have noticed the risks adding linked evaluation data to a system like the one outlined in Figure 1 could pose. In a letter protesting *Interfolio*, 25 NYU faculty emphasized ". . . massive violations of faculty and student intellectual property rights . . . profiling intellectual currents and individuals, and ultimately potentially threatening national security."[21]

---

[20] Hart, Nick, Suzette Kent, Julia Lane & Nancy Potok. 07/25/2025. "Knowledge Nation and Security at Risk." *NextGov FCW.* https://www.nextgov.com/ideas/2025/07/knowledge-nation-and-security-risk/406958/

[21] New York University Faculty of Arts & Sciences. Minutes of the Tenured and Tenure Track Faculty Senators Council Meeting of February 5, 2026. https://www.nyu.edu/content/dam/nyu/facultyGovernance/documents/T-FSC%20Meeting%20Minutes-2-5-26.pdf. P. 6. Accessed 04/09/2026.

More personally, I have been unable to ascertain either the general contents of Elsevier's data stores, or the extent of the data the company holds about me. I use Mendeley, maintain SciVal and SSRN profiles, search for literature on Scopus, track others' use of my research with ScienceDirect, publish in, review for, and have served on editorial boards of Elsevier journals. I often participate in faculty hiring, graduate student and post-doctoral recruitment via *Interfolio*. These types of tools help me and thousands of other researchers do our work. They also collect data about us and our work to feed the Elsevier product funnel depicted in Figure 1.

Evaluative, pre-publication, and workflow data are valuable precisely because they suggest the directions that research is heading. At scale, such data can help clarify the public value of taxpayer-funded research, inform researchers' selection of promising projects, assess the potential trajectories of fields, and help funders construct portfolios to achieve specific goals. Elsevier makes and sells products designed with these uses in mind.[22]

Given how many Elsevier products I use, I must assume the company holds data about me and about my work that is extensive and highly detailed. Even though I am an expert in these types of data, I cannot determine just how extensive and detailed they might be, let alone how they are being used or who can access them.

I find the prospect that detailed data about me and my work might be included in products designed to give others a leg up in races where I may also be competing disturbing. The idea that taxpayer-funded data about thousands of other researchers who may be working on much more sensitive or dangerous topics than I do could find its way into products that could give our national rivals a leg up in key fields is, frankly, frightening.

Academic and Business Incentives Are Not the Same.

A 2020 study found that five for-profit publishers[23] owned 56% of journals and published 57% of articles across all research fields.[24] Only two of those companies are based in the US. Elsevier owns more than 3000 scientific journals, handled more than 4.2 million article submissions and published better than 795,000 articles in 2025.[25] Nevertheless, their revenue growth "continues to be driven by the evolution of the business mix toward higher growth, higher value analytics and tools."[26] In other words, publications themselves are not

[22] https://www.elsevier.com/products/analytical-services. Accessed 04/12/2026.

[23] Elsevier, Springer, Wiley, Taylor & Francis, and Sage.

[24] Kim, S.-J. and K. S. Park (2020). Influence of the top 10 journal publishers listed in journal citation reports based on six indicators. *Science Editing*. 7(2), 142–148.

[25] Hart et. Al 07/25/2025. Op. cit. 17.

[26] Hart et. Al 07/25/2025, Op. cit. 19.

the most valuable part of Elsevier's portfolio. The data their use creates and the products that can be built from them are.

Citations provide a useful example of the ways that usage data, market power, academic and business incentives collide. Over time citations stratify journals into reputational bands. The status hierarchies that result make some more attractive publication venues than others. Higher status journals draw more readers, more attention, more citations, and more submissions. More submissions generally mean lower acceptance rates, which bolster prestige and, sometimes research quality.[27] Sought after journals become "must have" subscriptions that carry premium prices and are bundled with less desirable titles, much as early cable companies used ESPN to leverage purchase of less attractive channels.

Reputational hierarchies help keep subscriber prices high because career success often depends on making a discovery *and* reporting it in the "right" venue. An article in one of Elsevier's marquee journals, *Cell* (Impact Factor 42.5) or *The Lancet* (Impact Factor 88.5), can help make a career.[28] The "publish or perish" incentive structure of academic careers pushes researchers toward high-status venues.

For younger researchers, that pressure can force a choice between playing the status game and career success. In fields I know well, entry-level faculty positions commonly receive 300-400 applications. Departments typically interview 3-5 finalists for a single position. For good or ill, marquee publications often mean the difference between making and missing the interview cut. Journal status hierarchies loom very large for early career researchers. As a result, even very established researchers, at least those who advise students and post-docs, are also likely to feel reputational pressures and to convey the need for access to those journals to their librarians and subscription managers.

At the same time, publishers' terms commonly include non-disclosure provisions that prevent potential customers from knowing what others pay for the same services. This fact undermines the negotiating positions of subscribers like universities or government agencies. A recent analysis of 158 contracts between two publishers and university libraries concluded "each library negotiates a highly customized contract . . ." because "concentration of ownership tilts the battlefield toward large publishers."[29]

Publisher's costs stay low thanks to public research subsidies. Publishers typically own the rights to articles, but taxpayers typically support both the research articles report and the

---

[27] Jones, Jamaica. 2026. "A Precarious Task: Science Indicators, Research Assessment, and Science Policy in the United States." Dissertation. University of Pittsburg. P. 180.

[28] The impact Factor itself is a product made by Clarivate, one of Elsevier's competitors.

[29] An, et. Al 2024. Op. cit. P. 20.

work that goes into writing them. Universities usually pay for the time reviewers and editors spend to evaluate manuscripts. All those bills ultimately fall to taxpayers, yet the profits go to the publishing companies. Reputational pressures, market concentration, and opaque pricing models combine to leave US research organizations in a bind while Elsevier " . . . has a profit margin . . . higher than . . . Microsoft, Coca Cola, and Google."[30]

A recent commentary by Nick Hart, Suzette Kent, Julia Lane, and Nancy Potok, leaders in open data, privacy, data infrastructure and evidence-based policy, emphasizes why expanded researcher data in this landscape is a significant cause for concern.[31]

Interfolio is "used by over 400 American academic institutions and over 700,000 researchers to collect and process detailed confidential data about scientific hiring and promotion processes." It connects to "university payroll and human resource records" and "reflect[s] the often unpublished intellectual property and scientific collaborations of tens of thousands of American researchers and often their students."[32] Elsevier's privacy policies explain they use data to "enhance and improve our products, events and services and develop new ones."[33]

Likewise, the company's policy discloses that Elsevier may share data with "affiliates, trading names and divisions within the Elsevier group of companies and certain RELX companies." That portion of the policy includes a link to a list of office locations around the globe where data might be shared, including information technology offices in Beijing and Shanghai.[34]

What I anticipate would be most concerning to this committee is the following set of facts. Elsevier is a foreign company whose business practices are outside the scope of much US public oversight. It generates more than 1/3 of its revenue from research analytic products sold outside the US and Europe.[35] Those products are virtually certain to include data about

---

[30]Hagve, Martin. 08/17/2020. "The Money Behind Academic Publishing." *Tiddskriftet.* doi: 10.4045/tidsskr.20.0118. https://tidsskriftet.no/en/2020/08/kronikk/money-behind-academic-publishing . Accessed 04/10/2026; Marcum, Christopher Steven & Corinna Turbes. 02/04/2026. "The Cost of Confusion: Dispelling Myths About the NIH Public Access Policy and Article Processing Fees." *Upstream.* https://doi.org/10.54900/6m4xn-y3m49. https://upstream.force11.org/cost-of-confusion-dispelling-myths-about-the-nih-public-access-policy/ . Accessed 04/10/2026.

[31] Hart et. Al 07/25/2025. Op. cit.

[32] Hart et. Al 07/25/2025. Op. cit. P 1.

[33] Elsevier. "Privacy Policy" Last Updated: 30 October 2025. https://www.elsevier.com/legal/privacy-policy. Accessed 04/10/2025.

[34]Elsevier. "Elsevier Office Locations" https://www.elsevier.com/about/global-locations. Accessed 04/10/2026. The two offices are Reed Elsevier Information Technology (Beijing) Co. Ltd. and Reed Elsevier Information Technology (Beijing) Co Ltd – Shanghai Branch.

[35] Public financial documents are not detailed enough to determine how much business the company conducts with entities in specified countries of concern.

US researchers.[36] Since 2022, the company has owned a widely used tool, *Interfolio,* that integrates prepublication information about scientific and professional activities with university personnel data. Elsevier's privacy policy says such data could be used to develop new products and shared with certain partners. Among the partners detailed in a link included in the policy are two IT offices located in the People's Republic of China,[37] a *CHIPS and Science Act* country of concern.[38]

Hart and colleagues give voice to the risks and outline possible solutions. Rather than relinquishing this important terrain to foreign companies, they suggest, national leaders could "invest in a US based national research infrastructure centered on open-source research reporting tools and research data management tools with high standards for research security . . . under the control of US law."[39] Doing so would take important steps toward public access and toward the creation of a national data asset with many potentially valuable applications.

Researcher Data Infrastructure as a Strategic Asset

The opportunities and the risks apparent in the scholarly publishing landscape suggest that our nation needs an open, accountable public infrastructure suited to responsibly protect and use complex, sensitive data about research in sophisticated ways. Such an infrastructure could make more effective use of resources US taxpayers already pay for, increase ROI on public research investments, and address unnecessary risks to individual privacy and national security in the current landscape.

That ecosystem could also support AI-enabled science through projects like the Genesis Mission, which President Trump launched to ". . . harness federal scientific datasets . . . developed over decades of federal investments – to train scientific foundation models and create AI agents to test new hypotheses, automate research workflows, and accelerate scientific breakthroughs."[40] Achieving Genesis Mission goals[41] will require scientific data along with the kinds of research activity and researcher data I have been discussing. Public data infrastructures like those that I am describing to you now are designed for such a task.

---

[36] RELX Financials, Op. cit. p. 20

[37] Elsevier Global Locations. Op. cit.

[38] Elsevier Privacy Policy. Op. cit.

[39] Hart et. Al, Op. cit. p. 2.

[40] President Donald J. Trump. 11/24/2025. "Executive Order: Launching the Genesis Mission." https://www.whitehouse.gov/presidential-actions/2025/11/launching-the-genesis-mission/. Accessed 04/10/2026.

[41] US Department of Energy.03/17/2026. DE-FOA-0003612. "The Genesis Mission: Transforming Science and Energy With AI" https://science.osti.gov/-/media/grants/pdf/foas/2026/DE-FOA-0003612.pdf. Accessed 04/10/2026.

There are proven[42] models[43] for the key components of the system we need.[44] I direct one. The Institute for Research on Innovation & Science (IRIS) exists to solve problems that improve the public value of research. We have operated as an approved data repository for 11 years, anchor a secure, nationwide university data sharing consortium, and support secure research access through both a NIST 800-171 aligned virtual data enclave and the federal statistical research data centers.[45]

Today, with partners around the country, we are leading the *Industries of Ideas* project.[46] NSF's Technology, Innovation and Partnerships Directorate supported this effort to develop new ways to measure how AI is changing jobs, affecting employers and shifting demand for skills in regional labor markets. Industries of Ideas pursues its work with community stakeholders to put useful tools into the hands of real local decision makers.[47]

That is challenging, highly collaborative work. But it is doable. Many of the necessary[48] components[49] exist[50] and are proven[51] through decades of use. What has been missing is the will and the motivation to knit them together. Today's hearing seems to indicate that leaders like those serving on this committee are also motivated by the importance of these issues and the desire to protect the American research ecosystem, the people and work that power it, the interests and needs of the taxpayers who fund it, and the outputs it produces for our nation.

Improving the ROI on taxpayer supported research investments while protecting the security of American science and the privacy of American scientists requires us to treat data about research as a first-class strategic asset without relinquishing responsible public access and accountability. This Committee has an opportunity to advance gold standard science by instructing federal agencies to treat detailed data about publicly funded

---

[42] National Data Platform, Op. cit.

[43] Vivo, Op. cit.

[44] California Digital Library, Op. cit.

[45] Institute for Research on Innovation & Science (IRIS), Op. cit.

[46] Industries of Ideas. 2026. "Tracing public investment in AI research effects on workforce, education, nd skills." https://industriesofideas.ai/. Accessed 04/10/2026.

[47] Lane, Julia, Jason Owen-Smith & Bruce A. Weinberg. 2024. "How to Track the Economic Impact of Public Investments in AI." *Nature*. 630: 302-304; Owen-Smith, Jason. 2024. "AI and Work: How to Build New Data Foundations." *American Enterprise Institute*. https://www.aei.org/wp-content/uploads/2024/10/Owen-Smith-AI-and-work.pdf.

[48] ORCID. N.D. "ORCID: Connecting Research and Researchers." https://orcid.org/signin. Accessed 04/10/2026.

[49] United States Patent & Trademark Office. N.D. "PatentsView" https://www.patentsview.org/. Accessed 04/10/2026.

[50] Crossref. 01/27/2026. https://www.crossref.org/ Accessed 04/10/2026.

[51] National Center for Education Statistics. Op. cit.

research and researchers as a strategic national asset through support for proven, open, US-based, publicly accountable data infrastructure platforms and tools.

Thank you for your attention.