

Who Uses Open Access Research? Evidence from the use of US National Academies Reports



A fundamental principle of open access is that publication technology enables the widest possible audience for research findings. However, the extent to which open research is used outside of academia is often underexplored. Drawing on a dataset covering over a million user comments about their use of US National Academies consensus study reports, [Ameet](#)

[Doshi](#), [Diana Hicks](#), [Matteo Zullo](#) and [Omar I. Asensio](#) find widespread use of open research in the public sphere.

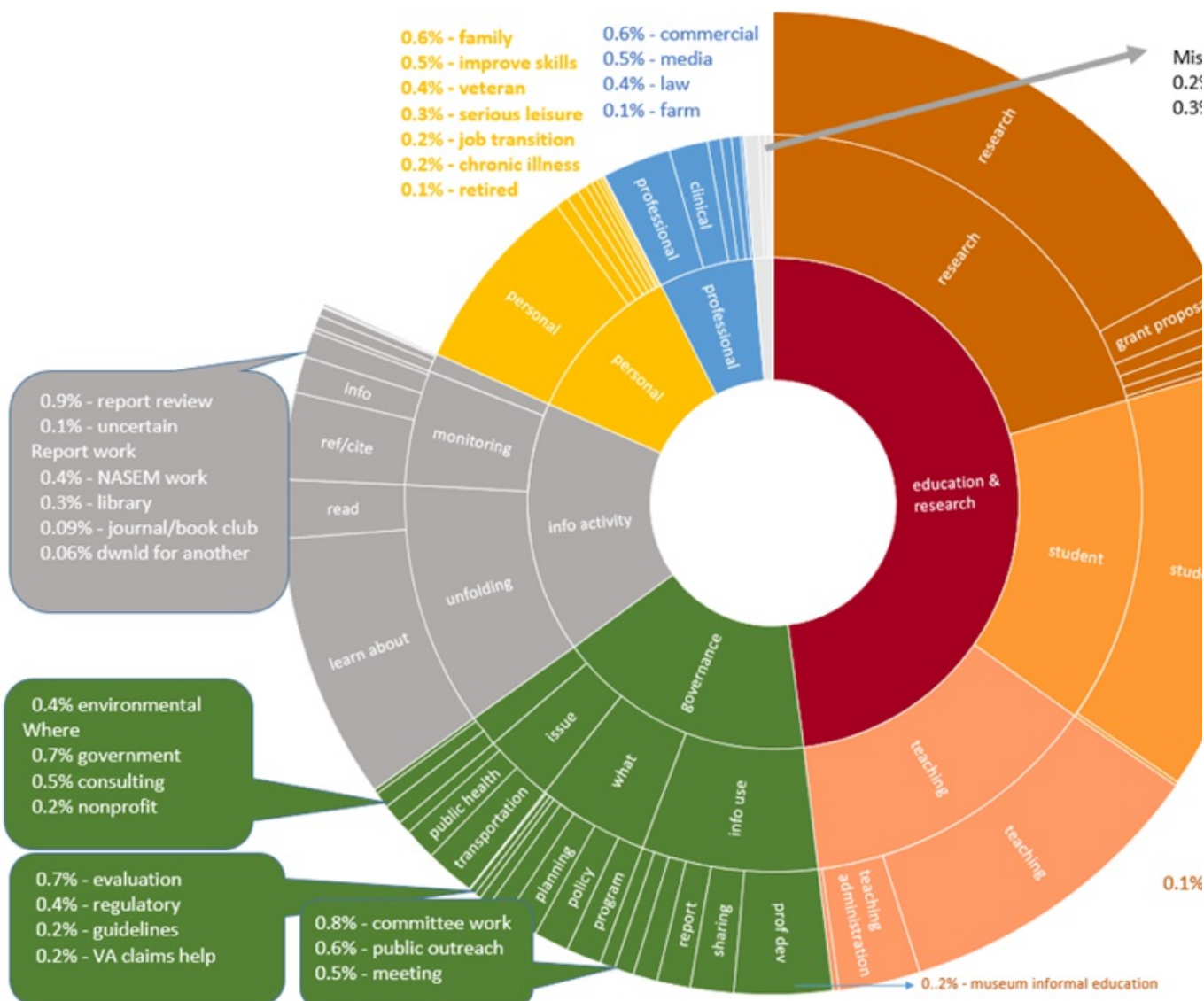
A central argument in favour of open access is the claim that the public benefits from having direct access to research. Beginning with the earliest open access manifestos, the [Budapest Open Access Initiative](#) (2002), the [Berlin Declaration on Open Access to Science and the Humanities](#) (2003) and the [Bethesda Statement on Open Access Publishing](#) (2003), OA adherents advanced their argument based on first principles: that the public has an *inherent right* to publicly-funded research. Most of these manifestos explicitly include non-researchers and the lay public as potential intended audiences for open access literature.

Yet, beyond invocations of *noblesse oblige* to “wider society” and [utopian hopes](#) to feed “curious minds,” the focus of OA conventions and manifestos [largely ignore the nature of use by the general public](#). Instead, these declarations functioned as statements-of-intent prompting action to expand no paywall access to research. Even when detail is provided, the imagined uses of open access materials often remain within the research realm: under-resourced scholars operating in the Global South, for instance, or to speed the pace of discovery and innovation within the triple-helix of university, industry and government sectors. While these are undoubtedly valid justifications for expanding access to research, left out of these potential user communities are the *dark universe* of people who are not research scientists or academic scholars.

The open access community has heretofore largely focused on overcoming the economic, legalistic and technological hurdles to create sustainable pathways to research. However, understanding and using scholarly research is non-trivial. Reading scholarly work more often than not requires specialised grounding in disciplinary concepts in order to parse the language of the domain. Is someone, who may not have strong grounding in the language and theory of a subdiscipline, willing to take the time and effort to overcome those barriers? Furthermore, the shift towards open access comes with significant costs to institutions and authors, as well as risks for smaller non-profit publishers. Is the global movement towards OA worth the risk to the established edifice of scholarly publishing and, by extension, to the advance of science itself? Specifically, what are the returns that accrue to society for moving publications to the open access model? It has now been twenty years since the Budapest declaration. What can we actually say about the public benefits of a more open scientific publishing ecosystem?

To help answer these questions [we analysed data from the US National Academies of Sciences, Engineering and Medicine \(NASEM\)](#) by classifying 1.6 million US-based comments about how NASEM’s consensus study reports are used by the public. [NASEM’s reports](#) consist of authoritative, independently-researched, consensus-based analysis on policy issues across domains. Since Abraham Lincoln first chartered the National Academy of Sciences in 1863, NASEM’s consensus study reports have served as influential scientific evidence for policymakers. The most downloaded reports are built on social science expertise in education and policy, in addition to medical knowledge. All consensus reports were made open access in 2011, and downloaders are prompted with a request to “please take a moment and tell us how you will be using this PDF.” The paper applies deep learning and natural language processing to label over a million comments, a task which would have otherwise required an inordinate amount of time and resources to accurately annotate the data. The deep learning neural network classifier implemented is [Google’s BERT](#), a transformer-based classifier, which uses bidirectional training based on the well-known *attention* mechanism to overcome limitations of one-directional approaches commonly used for text classification.

Our classification project reveals that the impact of these reports [extend far beyond the research community](#) (see *Results*, Fig 1). We find that half of all report downloads are used for non-academic purposes, including to improve the provision of services by medical professionals, local and regional planners, public health workers, and veterans' advocates, to name just a few of the 64 total categories of report use. Heavy use is made of Academies reports on STEM education and how people learn by teachers, school administrators and teachers' coaches. Other notable reports with their prominent users included *Dying in America* (chaplains), *Nutrient Requirements for Beef Cattle* (farmers), and *Best Care at Lower Costs* (clinicians and hospital administrators).



This picture suggests that taxpayer investments in open access to high-quality science do indeed pay dividends to society, broadly and at the local service level. The results also indicate a public motivated to improve their engagement with patients, students, clients, and fellow citizens, and seek out (and share) the best available evidence to solve problems at the coalface. This motivation by non-researchers to use and apply consensus-based research appears to overcome the challenge of parsing specialist jargon in technical writing. This finding also contrasts with the contemporary notion of a public completely misinformed by social media, though we do not dispute the very real issues surrounding social media manipulation. Additionally, we detect signals of “serious leisure” in the NASEM data, comprising about 4,300 comments. Serious leisure is a [sociological concept](#) introduced by Robert Stebbins to describe unpaid activities by individuals who engage in a systematic, self-directed pursuit of knowledge. The serious leisure devotee aims to continually expand understanding of their respective domains. These people downloaded reports relevant to wild edible plants ([Lost Crops of the Incas: Little-Known Plants of the Andes with Promise for Worldwide Cultivation](#)), astronomy ([New Worlds, New Horizons in Astronomy and Astrophysics](#)), and ham radio ([Handbook of Frequency Allocations and Spectrum Protection for Scientific Uses](#)).

The implications of this work are far-reaching. On the methodological side, the paper demonstrates the ability of machine learning techniques to enhance social science research and generate insights at scale. The techniques continue to improve, enhancing their precision and promising to exceed human ability to consistently make the subtle distinctions necessary to classify very large amounts of text for research purposes., members of the research team have been expanding the application of transformer-based algorithms into [other social science areas](#), including understanding consumer behaviour at scale with electric vehicle charging and smart meters.

Open access repositories require significant resources, both technological and human, to sustain and innovate. The National Academies Press, for example, has developed an engaging [user interface](#) to incentivize browsing and ease of access to NASEM publications. The PubMed Central server, developed and managed by the US National Institutes of Health (NIH), [requires millions of dollars per year to operate](#). Our research indicates there is an identifiable payoff to society for these taxpayer investments into people, technology and design to support OA publishing.

As we note in the paper, “[o]ur results establish the existence of demand for high-quality information by the public and that such knowledge is widely deployed to improve provision of services. Knowing the importance of such information, policy makers can be encouraged to protect it.” Librarians and open access advocates have long presupposed that open access to high-quality scientific knowledge could and should be viewed as a public good. Our empirical research suggests that the initial utopian aspirations regarding the public use and societal impact of OA may indeed rest on sound footing.

- This blog post is draws on the authors’ article, [Widespread use of National Academies Consensus Reports by the American Public](#), published in *Proceedings of the National Academies* first appeared at the [LSE Impact blog](#).
- Image Credit: Reproduced with permission of the authors.

[Please read our comments policy before commenting](#)

Note: The post gives the views of its authors, not the position USAPP– American Politics and Policy, nor of the London School of Economics.

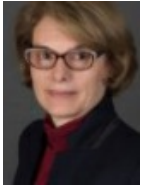
Shortened URL for this post: <https://bit.ly/388ci6s>

About the author



Ameet Doshi – Princeton University

Ameet Doshi is Head of the Donald E. Stokes Library at Princeton University and a doctoral student in the School of Public Policy at the Georgia Institute of Technology. Doshi’s research focuses on how non-scientists use open access research. Previously, he held librarian positions at the University of North Carolina, Wilmington and the Georgia Institute of Technology, and has served on the American Library Association’s Center for the Future of Libraries advisory board.

**Diana Hicks** – *Georgia Institute of Technology*

Diana Hicks is Professor in the School of Public Policy, Georgia Institute of Technology specializing in metrics for science and technology policy. She was the first author on the Leiden Manifesto for research metrics published in *Nature*, translated into 2 languages and awarded the 2016 Ziman award of the European Association for the Study of Science and Technology (EASST) for collaborative promotion of public interaction with science and technology. She co-chairs the biennial international Atlanta Conference on Science and Innovation Policy. Prof. Hicks has also taught at the Haas School of Business at the University of California, Berkeley; SPRU, University of Sussex, and worked at NISTEP in Tokyo. In 2018 she was elected fellow of the American Association for the Advancement of Science (AAAS).

**Matteo Zullo** – *Georgia Institute of Technology*

Matteo Zullo is a PhD candidate at the Georgia Institute of Technology focusing on educational analytics, science & technology policy, standardized testing, and AI. He holds a MS in Data Analytics from the same university and has worked in technology consulting after receiving his education in business economics in his native country.

**Omar I. Asensio** – *Georgia Institute of Technology*

Omar I. Asensio is an Assistant Professor in the School of Public Policy at the Georgia Institute of Technology and Director of the Data Science & Policy lab. His research focuses on the intersection of big data and public policy, with applications to energy systems and consumer behavior, digital innovation, smart cities, resource conservation and machine learning in transportation and electric mobility.