Broader scope is key to the future of 'science of science'

For years, researchers have interrogated scientists' own research practices. A computational research stream, often termed 'science of science', studies the signatures these practices leave in big data. As the field matures, it is looking for ways to use its data-driven insights to make a tangible mark in science policy.

From project funding proposals to article publication, scientists' actions produce data. Armed with advanced computational and machine learning tools, researchers working in the field of 'science of science' scour these big data for insights into the inner workings of the scientific enterprise. They hope to make discovery, research funding, evaluation and dissemination more efficient and equitable.

This June, we were at the first annual meeting of the 'science of science' community at the National Academies of Science (NAS) in Washington, DC. Researchers asked what big data analyses can tell us about hiring, productive collaborations, impact, disruption and the stubborn inequities in science. Listening from the sidelines, funders and policy makers considered how the science of science will help them to optimize the distribution of funds and how research can better meet the needs of the public.

'Science of science' is no longer a niche of computational specialist interest, but a stream of research with clearly defined use cases. But so far, big data analyses have led to relatively few concrete actions to improve science and science policy. As the field finds its place among other branches of research on research, its practitioners are looking for ways to use its data-driven insights to make a tangible mark in science policy.

The future includes interventions

Science of science should not just tell us how the world of science currently is — it should also push us towards the science that could be.

By analysing big data, researchers have confirmed deeply rooted intersectional inequalities in science production and publishing. And they have a fairly good understanding of how various dimensions of diversity, from gender and racial or ethnic composition to interdisciplinarity and geography, relate to outcomes[1]. "Correlational analyses based on large-scale datasets have been successful in uncovering a large set of highly reproducible facts and patterns that are highly generalizable across domains," says Dashun Wang, director

Credit: studiostockart/DigitalVision Vectors/ Getty

of the Center for Science of Science and Innovation at Northwestern University.

But correlations only reflect what is in the data; they do not directly imply an action plan.

Correlations are also influenced by multitudes of biases rampant in the world of science. These inequalities are reflected in both the available data and traditional impact metrics. Similar to biases in machine learning, decisions based on correlations in data could amplify, rather than ameliorate, existing inequalities. For example, if well-funded teams accrue more citations than less-funded ones, does that mean that we should increase their share of funding even more?[2] If multidisciplinary teams accumulate fewer citations, should we discourage such collaborations?[3]

We need data-driven insights from the field to address some of the biggest issues in the scientific enterprise. Science of science researchers are becoming increasingly aware that they need to complement correlational work with causal inference and interventional techniques.

Citation diversity statements in academic articles are one example of a tangible step to improve science, aiming to correct the gender and racial/ethnic disparities in citation lists[4]. "We are aware of numerous gender biases in science, yet we still cite like it is 1995. The gender composition of the

scientific world is not reflected in citation patterns," says Dani Bassett, a professor at the University of Pennsylvania and one of the researchers behind the Citation Diversity Statements initiative. "Crucially, just knowing that these biases exist will not by itself change the field — we also need to test specific interventions and nudges."

Science-informed practices and science policy will probably require interventional research, not just correlational. Still, this does not mean that purely data-driven work is not valuable. The massive scale of data is an advantage in itself. And the breadth of computational analyses can help researchers to design the right interventions.

"Narrow interventional studies that are common in economics may sometimes miss some of the broader patterns that we see in big data," says Wang. "The science of science of the future will benefit from a flourishing ecology of both observational and experimental studies."

This future, it seems, is destined to blur the boundary between science of science, a field with origins in the computational sciences, and other branches of research on research, such as psychology and economics of science.

Closer to the public

Science evaluation, funding and its relationship to society are also in need of actionable insights from science of science. At the meeting at NAS, Rush Holt Jr, a former US House Representative, presented the community with a challenge — addressing "the deep chasm between science and the public." Scientific success by itself means little if it does not serve society's needs. And scientists' internal measures of impact, widely used in science of science analyses, do not capture this dimension of success.

The deluge of COVID-19 research has been a triumphant success judging by traditional quantitative measures — citations and journal impact factors. Top COVID-19 papers racked up tens of thousands of citations[5] in 2021, and many journals publishing this work doubled their impact factors[6]. But not so much if we measure

success by the rates of vaccine refusal and overall fatalities. And while this is not exactly a failure of the scientific enterprise itself, it does signal that something is broken in the relationship between traditional science metrics and public impact. Can science of science help?

"The real challenge for science of science, as a discipline, is how we can evaluate how well the scientists are providing the public with the understanding, the ownership and the expectation to apply science in conducting their own affairs," said Holt.

Becoming more actionable also means broadening the reach of research and diversifying data sources. Open publication and patent data are now widely available. They include more and more journals and previously neglected institutions outside Europe, the US and Oceania[7]. But this is less true for some of the behind-the-scenes aspects of the scientific enterprise, such as peer review, editorial decisions and evaluations of funding proposals.

Rikke Nørding Christensen, a senior impact partner at Novo Nordisk Foundation (a major science funder), says that behind-the-scenes data carry tremendous potential to help design concrete interventions, but currently there are major barriers in the way of access. Evaluation committees, for instance, may not be the best way to select the most promising research proposals, but without the fine-grained evaluation data from foundations, the community cannot know. "Science of science could do much more if funders and organisations were more open with their data," says Christensen.

Companies, funders and publishers are reluctant to share their behind-the-scenes data and do so only subject to restrictive data sharing agreements. Although there are important legal and privacy considerations, a culture of more open data sharing would empower the science of science community to expand their analyses and go beyond the traditional impact metrics. Furthermore, only through openness and collaboration can funders design large-scale randomized clinical trials to effectively test causal theories.

The list of issues that the scientists of science aim to solve is long and familiar: inequalities in science making and dissemination, suboptimal funding decisions and modes of collaboration, subjective evaluation, questionable research practices and more. For many years, researchers working in the fields of metascience have used the tools of their own disciplines to interrogate science making. Science of science, as a computational field, should not be narrowly limited to analyses of existing big data. To have a greater impact in all corners of the scientific enterprise, it must keep expanding the scope of both its methods and its data sources, integrating more closely with other branches of research on research and learning from these closely related partner fields. ❐

References
1. Wang, D. & Barabási, A. L. *The Science of Science* (Cambridge Univ. Press, 2021).
2. Wuchty, S., Jones, B. F. & Uzzi, B. *Science* **316**, 1036–1039 (2007).
3. Sun, Y., Livan, G., Ma, A. & Latora, V. *Commun. Phys.* **4**, 263 (2021).
4. Zurn, P., Bassett, D. S. & Rust, N. C. *Trends Cogn. Sci.* **24**, 669–672 (2020).
5. Google Scholar reveals its most influential papers for 2021. *Nature Index* https://go.nature.com/3amP25O (2021).
6. Quaderi, N. Journal Citation Reports 2022: COVID-19 research continues to drive increased citation impact. *Clarivate* https://go.nature.com/3OPu8LD (2022).
7. Singh Chawla, D. *Nature* https://doi.org/10.1038/d41586-022-00138-y (2022).