Follow the electronic footprints: how to track impact without asking scientists to lift a pen.

by Blog Admin   November 2, 2011

*Citation counts and other traditional methods of reporting the results of research are proving inadequate, but now that almost all scientific and economic activity leaves electronic footprints, we have the tools to do better writes **Julia Lane,** Program Director of the Science of Science & Innovation Policy program at the **National Science Foundation**. A U.S. approach that links grants to the scientists that receive them, the students they train, the intellectual products they produce and the resulting social and economic outcomes of their research offers some ideas for U.K. practice.*

The goal of science funders and research institutions is to support the creation, transmission, and adoption of knowledge. That goal has, by and large, been achieved in the United States, where universities typically rank the highest in the world in terms of research results – both in terms of scientific achievement such as Nobel Prizes and economic impact such as productivity growth. Yet the tide of accountability, which has already affected the practice of medicine and school teaching, is now lapping at the feet of scientists. Although reporting on the results of research has always been a requirement, the passage of the American Recovery and Reinvestment Act (ARRA) in 2009 required the scientific community to document on a much broader scale the employment impact of their activities. And it is not just the government: the rise of such ranking systems as the Shanghai ratings have forced the scientific community to consider what constitutes good science.
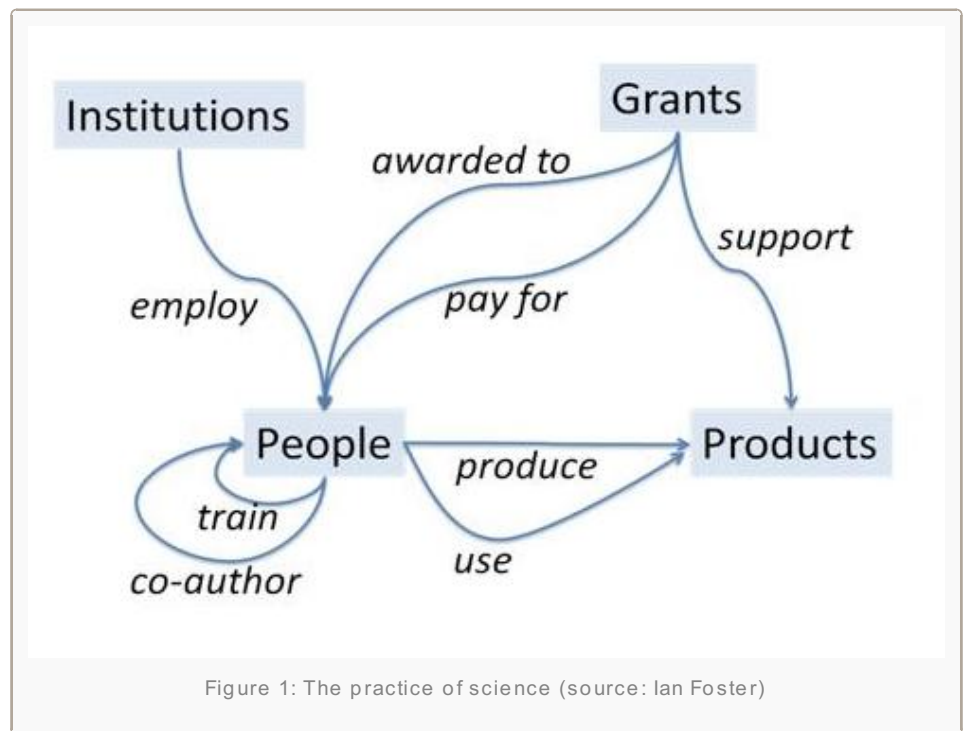
Yet we should think hard about developing reporting systems that not only serve to describe the results in ways that make sense to the public, but also use scarce resources to enable scientists to do science, not reporting. Current reporting systems, which are relics developed decades ago, cannot be used to answer that question well. Counts of documents, citations and patents, and their derivatives, are proving both burdensome and inadequate to either guide science management or to convey the results to a skeptical public.

We now have the tools to do better. Almost all scientific and economic activity now leaves electronic footprints. We can use 21st century technologies to reduce the burden on researchers, and describe the real conduct of science. We can describe what science investments have been made in a more intuitive way, so that elected officials and program managers alike can characterize where and what investments have been made. We can describe how many students have been trained, and in what areas, so that the public understands the immediate as well as future results, and program managers can identify future skill shortfalls or surpluses in scientific areas. And we can provide much richer information about the technological advances tied to science investments, and the firms who make use of them, so that the evidence of economic impact has a broad empirical basis.

We can do all this without asking scientists to lift a pen.

The approach in the U.S. is based on the development of a more scientific approach to describing the conduct of science. The STAR METRICS data infrastructure is a voluntary collaboration between research institutions and funding agencies – both of which have the goal of fostering and describing good science. The approach is to use existing data to link grants to the scientists that receive them, link them to the students they train and the intellectual products they produce and then (eventually) to the resulting social and economic outcomes (Figure 1). This is done by combining new tools and data in a flexible data platform that can be used by the scientific community to develop a variety of different metrics depending on both the questions that are posed and the theoretical framework.

This nascent system is being used in a number of ways; three are described here. One is to use new tools to describe what science is being done. Another is to use new data sources to link the grants to the people supported and to describe which types of students and postdoctoral researchers are being trained, and in which areas. A third exploits a recently developed dataset to link grants to people and in turn to patents to describe one type of economic impact.

Figure 1: The practice of science (source: Ian Foster)

What science is being done?

Scientists should not have to manually develop keywords to describe what science is being done. New scientific tools have been developed that can be used is to classify funding based on scientists' written descriptions of their work. Scientists – both computer scientists and computational linguists – have developed topic modeling tools that permit the machine reading of large text based data collections that enable this automatic classification on a large scale. The STAR METRICS team applied topic models in excess of 100,000 NSF award abstracts from 2000-2010 to provide one representation of the contents of this large text based data collection.

How many students are being trained as a result of science investments? Current reporting systems do not systematically document who is being supported by science funding. Yet it is critical to capture this information not only because innovation is increasingly driven by scientific teams – but also because the future workforce can be trained in the scientific method by working on research awards. New STAR METRICS data sources can be used to link each project, together with the project topics, with the staff who work on them. The STAR METRICS system, in a voluntary collaboration with the universities, draws information about the staffing of each project directly from institutional payroll records, thus automatically capturing information on how many people are working on each project, the proportion of time that is devoted to each grant, as well as their occupations . No personally identifiable information is used, and no principal investigator had to lift a pen to do this reporting.

What are some of the results of science funding? Current reporting systems typically require that the principal investigator manually report their patents during the period of the award. The STAR METRICS data infrastructure makes use of a more scientific source of data: the new disambiguated patent database developed by NSF funded researchers that links U.S. Patent office data to identify unique inventors over time. This is a non-trivial task because the United States Patent Office (USPTO) does not require consistent and unique identifiers for inventors. The advantage to using this automated approach is that patents that directly cite federal grant funding can be automatically captured, resulting in higher quality and more complete reporting than is likely during an award period. Since much scientific research is cumulative, the data can be used to automatically identify patents that have been filed by researchers previously funded by NSF awards. And, since much scientific research builds on previous work, the data can also be used to automatically identify patents that cite patents funded by NSF awards.

In sum, the wave of accountability should not – and must not – result in less or lesser science. The scientific community should proactively use scientific tools to be more accountable. There is a saying in Washington that if you're not at the table, you're on the menu. We should be at the table.

Julia Lane is Program Director of the Science of Science & Innovation Policy program at the National Science Foundation.

Related posts:

1. Impact from beyond the grave: how to ensure impact grows greater with the demise of the author
2. Continual publishing across journals, blogs and social media maximises impact by increasing the size of the 'academic footprint'.
3. 'Maximizing The Impacts Of Your Research: A Handbook For Social Scientists' now available to download as a PDF
4. From blogging to print: My journey to creating impact
5. Making impact with history: how policy makers have much to learn from historians and social scientists and why academic writing must strive for clarity