

Measuring the Results of Science Investments

Julia Lane¹ and Stefano Bertuzzi²

Historically, federally funded basic and applied scientific research has promoted scientific knowledge, innovation, economic growth, and social well-being. However, there is increasing pressure to document the results of these research investments in a scientific manner (1, 2) and to quantify how much of the work is linked to innovation (3).

Is it possible to create a system in which the effects of scientific research can be described? If so, what would be the inputs, outputs, and structure of the system? What scientific disciplines should inform the formulation of such a model? Creating a system in which the effects of scientific research can be described on an ongoing basis—without increasing the burden on research institutions and principal investigators—is difficult.

The current scientific data infrastructure is based on identifying, funding, and managing high-quality science, not on understanding its impact. The main sources of data on research and development in the United States—the Survey of Federal Funds for Research and Development (the federal funds survey) and the Survey of Federal Science and Engineering Support to Universities, Colleges, and Nonprofit Institutions—were designed to describe the types and levels of science investments, not their impact or effects (4). There are systems available to capture outcomes (for example, various health and economic information systems) but they do not link inputs with outputs and outcomes. Historically, there have been limited resources devoted to rigorous evaluations of science investments (5). Indeed, the roadmap published by the National Science and Technology Council (NSTC) Science of Science Policy Interagency group in 2008, found that “current science and technology investment decisions are based on analyses that lack a strong theoretical and empirical basis” (6).

¹Science of Science and Innovation Policy, National Science Foundation, Arlington, VA 22230, USA. E-mail: jlane@nsf.gov ²NIH Office of Science Policy Analysis, National Institutes of Health, Bethesda, MD 20892, USA. E-mail: stefano_bertuzzi@nih.gov

The challenge is not limited to the United States; other countries have been developing systematic ways of describing the results of science investments. Since 1986, the Higher Education Funding Councils in England has assessed research with its Research Assessment Exercises (now a Research Assessment Framework) intended to assess the quality, impact, and vitality of funded research. Their lessons are salutary: Although the exercises did help to improve research quality, the process of producing the data was burdensome

“The intent is to leverage revolutionary digital technology to capture the broad scientific, social, economic, and workforce impacts of science investments.”

and complex (7). In 2009, the European Union EUFORIA conference, which examined the impact of the Framework Programme (FP) 6, included, as a major recommendation, of building a database of project results for future FPs, noting that “getting robust data on the FPs in terms of participation and results is the foundation for any evaluation” (8). In 2011, the Japanese government is creating a program to advance the science of science and innovation.

A high-quality system should be based on describing the activities of scientists and clusters of scientists. Of course, the direct output of research is knowledge, which includes even research “failures,” and is difficult to measure. Despite this, the system should include proximal measures of scientific output (such as publications, citations, and patents) and go well beyond simple publication counts to the identification of emerging and interdisciplinary areas. It should also include broader outcomes, such as better health, clean energy and environment, the training of an analytically oriented workforce, and increased competitiveness. It should be structured to compare differences in outcomes and outputs of the recipients of science funding relative to a comparable control group that did not receive funding.

The development and analysis of such a system will not be easy—there are multiple feedback loops and long lags—and it is important to go beyond an accounting exer-

Science agencies and research institutions are building the infrastructure to evaluate results of federal funding of scientific research.

cise. However, there are useful precedents in other fields of policy in the United States. The Institute for Education Sciences has had a major impact on the quality of education policy. It has funded high-quality evaluations and brought together experts in economics, education, and other fields to provide evidence about the effects of education investments (9). The Center for Evidence-Based Policy has identified high-quality evaluations in a variety of policy areas, ranging from crime to health care to labor markets (10).

Developing such a system and the associated data infrastructure will require financial and intellectual resources. Other efforts to put together a data infrastructure describing the outcomes of research and development (R&D) investments, both by the private and the public sectors, no longer function for a variety of reasons (11). The new focus on accountability, combined with new technology and the broad-based commitment of key stakeholders, may result in a better outcome.

Currently, key data elements are dispersed across federal agencies and research institutions or are in third-party databases. For example, information about what science is being funded is often neither in structured format nor systematically shared across agencies; administrative information about the students supported by federal funding is housed at research institutions, but not by the agencies; and the universe of data on patents, publications, and citations is typically maintained by such third-party sources as the U.S. Patent and Trademark Office and the Web of Science. Similarly, research institutions, rather than federal agencies, typically have better access to data on subawards, vendors, and overhead expenditures, and these are not typically available in a way that can be mined and studied analytically. Reported outputs are only captured during the funding period (typically 3 to 5 years), often manually and in an unstructured format. The reporting burden is very high: The Federal Demonstration Part-

nership has estimated that some 42% of principal investigators' time is spent on administrative tasks (12).

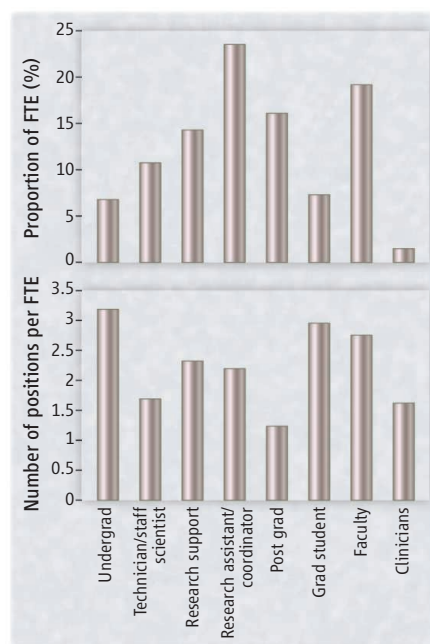
It is important to address these deficiencies; otherwise, impact estimates will be biased or unachievable. Numerous case studies estimate that the full outcomes are often felt more than a decade after the research is initiated. Capturing activities of students is similarly critical; they not only form the workforce of the future but generate scientific, social, and economic activity. Characterizing the funding and outcomes of interdisciplinary research within and across federal agencies will require being able to describe the structure of proposals, awards, and publications (4) and building information systems that link outputs to inputs or infrastructure investments. Estimating impact not only requires capturing data and comparing the outputs and outcomes of the activities of both funded and unfunded scientists but thinking carefully about appropriate counterfactuals. It is important to be clear about the policy question of interest and to develop a full cost-benefit analysis (9).

The STAR METRICS (Science and Technology for America's Reinvestment: Measuring the Effects of Research on Innovation, Competitiveness, and Science) is an attempt to focus both financial and intellectual resources to address some of these challenges in the United States. The program is being developed by a consortium consisting of the National Institutes of Health (NIH) and the National Science Foundation (NSF) under the auspices of the White House Office of Science Technology and Policy (OSTP). The Department of Energy and the Environmental Protection Agency are joining that consortium. The goal is to work collaboratively with research institutions to build a scientific data infrastructure that brings together inputs, outputs, and outcomes from a variety of sources in an open a fashion as possible. A major functional aim is to reduce, as much as possible, manual reporting by research institutions and principal investigators. The use of such automated tools as CiteSeerX, which facilitates the capture of outputs produced by principal investigators, offers great promise in fulfilling this aim. Such an approach should simultaneously reduce the reporting burden and increase the period over which outputs can be measured. Similarly, text-mining tools and topic-modeling approaches can be used to represent the information within proposals and scientific documents to describe the nature of scientific investments. The design is intended to permit scientists to provide input into the way in which knowledge is created

and transmitted in their disciplines, as well as to engage social and behavioral scientists for modeling the impact of interventions.

STAR METRICS began as a small pilot with seven institutions in July of 2009 in cooperation with the Federal Demonstration Partnership. By May of 2010, a Memorandum of Understanding had been signed with the participating agencies; Office of Management and Budget approval was received in July 2010 to expand the program. Since then, more than 60 institutions have signed participation agreements and at least 50 more have indicated interest in participating.

In practical terms, STAR METRICS is structured in two phases. The first phase ascertains the immediate effect of science spending on employment. It uses administrative records within participating institutions to document how many scientists (including graduate students, undergraduate students, and research staff) are supported by federal science funding, as well as to capture information on subawards and subcontracts. Only 14 data elements are required (13); STAR METRICS is now capturing that information electronically from institutional financial records (without personal identifiers) without burden for the scientists. This process, described in detail at <https://www.starmetrics.nih.gov>, has enabled generation of tables



Individuals in occupations supported by science funding. (Top) The distribution of FTEs in occupations directly supported by science funding. **(Bottom)** The number of distinct individuals per FTE directly supported by science funding per FTE. [Source: STAR METRICS data for 45 institutions, third quarter 2010].

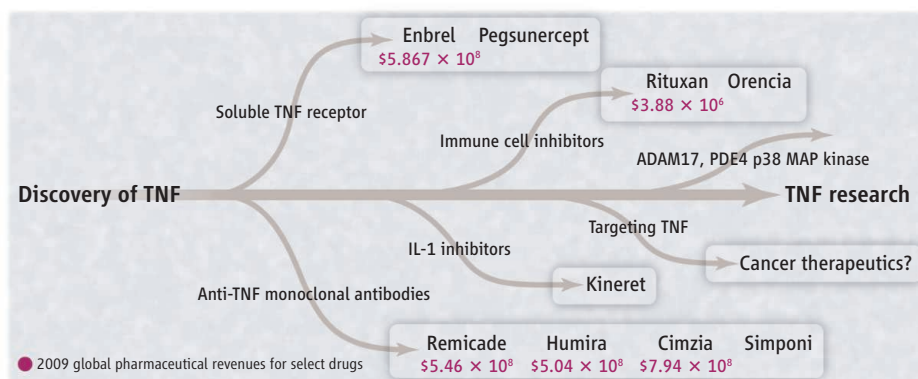
and maps of jobs and positions immediately traceable to science funding at each institution. Federal agencies use the same reports, aggregated from multiple institutions. Source data can be generated with minimal burden and cost—the typical institution requires less than 20 hours of staff time to generate the initial report. Subsequent reports are automated.

A graphic visualization of the type of report generated for each university is shown in the first figure. Science funding supports a wide range of occupations (top), and the nature of research means that science funding supports more individuals than are conveyed by simple counts of full-time equivalent (FTE) workers or students (bottom).

Phase I also provides estimates of how many additional jobs are created that are directly attributable to firms whose goods and services result from the spending of research institutions. These institutions, unlike federal agencies, have data that can be used to derive the industry and geographic location of their vendors and subcontractors. In combination with publicly available data from the Economics Directorate of the Census Bureau, we can estimate the payroll associated with payments and, hence, the number of jobs.

Phase II is designed to capture outputs and outcomes beyond the initial employment effects captured by phase I. The intent is to leverage revolutionary digital technology to capture the broad scientific, social, economic, and workforce results of science investments. Almost all scientific activity is eventually captured in electronic form. At least initially, this means we need to develop ways in which scientists' activities can be automatically, rather than manually, reported to science agencies. Phase II is likely to take at least 5 years to achieve the intermediate goals we have laid out here. Research institutions are developing structured information architectures to capture current and more accurate information about scientists' interests, activities, and accomplishments, including, for example, the VIVO Project (<http://vivoweb.org>), the Harvard Profiles System, and others. Brazilian science agencies have developed a system (Lattes Platform) for researchers and scientists to register and build curricula vitae and to capture scientific outcomes. The STAR METRICS team is beginning to consult with the scientific community to identify viable approaches.

An initial consultation meeting with the vice presidents for research of universities participating in phase I was attended by high-level representatives of more than 40 research institutions in October 2010. One suggestion from that meeting has been that



An example of research investments. Linking the discovery of TNF and its related properties to NIH investments in research grants.

the federal agencies could implement single progress reports and/or common biographical sketches, with a uniform electronic reporting template. The bureaucratic framework already exists, in the form of the uniform Research Performance Progress Report (14). Implementing the approach might involve providing tools that could streamline reporting, such as automated biographical sketches, profiles, and annual reports. In cases where data elements, such as publications and other ways of transmitting scientific knowledge, can be labeled with unique identifiers, scientists' reporting burden would be reduced. The consensus at a recent technical workshop on this topic was that if the federal agencies set up the core empirical infrastructure and data, the scientific community could create good software tools for building automated reports (15).

Another approach is to use existing administrative data, such as these from the U.S. Patent Office, to link patent data and the associated critical publications to their intellectual provenance in federally funded research. (16). That research has already generated insights into understanding collaboration networks and the way in which initial research investments ripple through science. For example, the second figure uses automated analysis of patent data and scientific connections to trace the path from the initial discovery of tumor necrosis factor (TNF) to successful biotech drugs. We also plan to expand the use of the existing patent database to provide automated visualizations of technologies supported by NIH- and NSF-funded research, as well as the firms using them.

We began by asking what scientific disciplines would inform the development of the system. There are many possibilities. For example, knowledge organization systems theory may inform the conceptual approach, which requires the maintenance of a set of relations between different areas of scientific knowledge and the maintenance of continu-

ity between past, current, and emerging ways of describing science (17). The fact that science is becoming increasingly team-oriented may necessitate drawing on the advances in network analysis and graph theory to describe the complex and changing nature of scientific collaboration. Even something as seemingly straightforward as describing what science is being done, which is beyond the current reporting capacity of many science agencies, may draw on recent advances in topic modeling (18).

There are interesting questions to be answered with the restructured data. For example, what types of funding are most successful? Preliminary evidence suggests that the structure and type of multiuniversity and multidisciplinary collaborations matter (19). How important are institutions, like biological resource centers, in stimulating research? What evidence supports the notion that it is better (or worse) to fund junior versus senior researchers? What are the employment and earnings outcomes for students trained in science? An open and transparent approach, as well as full scientific engagement, is necessary. Federal agencies typically do not have resources to build complex models and develop analytical techniques necessary to tease out the marginal and average impact of interventions in different areas.

In addition to the financial resources that have been made available, we will also need to attract the intellectual resources of the research community. We believe the scientific challenge is compelling: The way in which scientists create, disseminate, and adopt knowledge in cyberspace is changing in new and exciting ways, and scientists should be fully engaged in describing and studying these changes. Collaborations between computer scientists and social scientists can capture these activities by means of new digital technologies and statistical techniques. We believe that the data being generated will attract new researchers

and students to the field. Finally, we hope that the active engagement of the federal science policy community through STAR METRICS will help ensure that the scientific advances in science measurement move the data available for science policy to the same analytical level as the data available for the study of education, labor, and health-care policy.

References and Notes

1. C. Macilwain, *Nature* **465**, 682 (2010).
2. P. Orszag, J. Holdren, Science and Technology Priorities for the FY 2012 Budget [memorandum], 12 July 2010; www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-30.pdf.
3. D. E. Stokes, *Pasteur's Quadrant: Basic Science and Technological Innovation* (Brookings Institution Press, Washington, DC, 1997).
4. Committee on National Statistics, *Data on Federal Research and Development Investments: A Pathway to Modernization* (National Research Council, National Academies Press, Washington, DC, 2010).
5. A. Arora, A. Gambardella, in *Contributions in Memory of Zvi Griliches*, J. Mairesse, M. Trajtenberg, Eds. (National Bureau of Economic Research, Cambridge, MA, 2010), pp. 91–115.
6. Interagency Task Group on Science Policy, *The Science of Science Policy: A Federal Research Roadmap* (Office of Science and Technology Policy, Washington DC, 2008).
7. C. Macilwain, *Nature* **463**, 291 (2010).
8. EUFORIA conclusions and recommendations, Czech Presidency of the Council of the EU (Prague, Czech Republic, 2009); www.eufordia2009.eu/dokums_raw/eufordiaconclusionsfinal2doc_1216100183.pdf.
9. J. Smith, A. Sweetman, in *Strengthening Evidence-Based Policy in the Australian Federation, Volume 1: Proceedings* (Productivity Commission, Canberra, Australia, 2010), pp. 59–101.
10. Social programs that work, <http://evidencebasedprograms.org/wordpress/>.
11. D. Hicks, paper presented at the workshop Science of Science Measurement, Washington, DC, 2 to 3 December 2010; <http://www.nsf.gov/sbe/sosp>.
12. S. Rockwell, *Res. Manage. Rev.* **16**, 28 (2009).
13. The federal agency and institutional award numbers, the overhead charged, the de-identified employee ID number and occupation, the proportion of earnings allocated to award, the FTE status, the subaward recipient DUNS number (a unique number to identify a business entity) and payment amount, the vendor DUNS number and payment amount, and the proportion of overhead associated with salaries.
14. Research Performance Progress Report, www.nsf.gov/bfa/dias/policy/rppr/.
15. Workshop on Changing the Conduct of Science in the Information Age, 26 April 2010 (Office of International Science and Engineering, NSF, Arlington, VA, 2010); <http://scienceofsciencepolicy.net/media/p/714.aspx>.
16. L. Fleming, V. Torvik, *From Grant to Commercialization: An Integrated Demonstration Database Which Permits Tracing, Assessing and Measuring the Impact of Science Funding* (NSF grants 0965279 and 0965341, NSF, Arlington, VA, 2009).
17. P. Lambe, in (15), pp. 40–47; <http://scienceofsciencepolicy.net/media/p/714.aspx>.
18. B. W. Herr II et al., *The NIH Visual Browser: An Interactive Visualization of Biomedical Research*, paper presented at Information Visualisation, 2009, 13th International Conference, Barcelona, Spain, 12 to 17 July 2009, pp. 505–509; http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5190811.
19. J. N. Cummings, S. Kiesler, *Soc. Stud. Sci.* **35**, 703 (2005).
20. The opinions expressed are those of the authors and may not reflect the policies of NSF or NIH.

10.1126/science.1201865