Research Trends

Volume 1 | Issue 30 Article 3

9-1-2012

Big data: Science metrics and the black box of science policy

Julia Lane
American Institute of Research, US

Section 2: Big Data

Science Metrics and the black box of Science Policy

Julia Lane
Senior Economics Director,
American Institute of Research

The deluge of data and metrics are generating much heat but shed little light on the black box of science policy. The fundamental problem is conceptual: metrics that connect science funding interventions with numbers of documents miss the key link. Science is done by scientists. Dashboards of metrics that don't link back to scientists are like dashboards missing the link of cables to the engine. They do not provide policy makers with information on how or why funding changed the way in which scientists created and transmitted knowledge. In other words, while bibliometricians have made use of the data deluge to make enormous advances in understanding how to manage scientific documents, the science policy community needs to use the data deluge to make enormous advances in understanding how to manage science (1).

Missing causal links matters

If the focus of funding agencies turns to forcing scientists to produce scientific papers and patents, then they will do so. But if, as the evidence suggests, the essence of science is the creation, transmission and adoption of knowledge via scientific networks, then by missing the causal links, the agencies may distort and retard the very activity they wish to foster. Funding agencies must develop "the ability to define a clear policy intervention, assess its likely impact on the scientific community, find appropriate measures of scientific activities in the pre- and post-period, and define a clear counterfactual." (2) This is no different from Louis Pasteur's swan flask diagram (see Figure 1) that illustrates the fact that spontaneous generation is impossible and that life can never be created out of non-life (3). Like any scientist, we must develop the appropriate conceptual framework that enables us to write down the theory of change of how science policy interventions work – describing what makes the engine run (4).

A sensible organizing framework has been provided by Ian Foster, which identifies individual scientists (or the scientific community consisting of the networks of scientists) as the "engine" that generates scientific ideas. In this case the theory of change is that there is a link between funding and the way in which those networks assemble. Then, in turn, there is a link between scientific networks and the way in which those ideas are created and transmitted, and hence used to generate scientific, social, economic and workforce "products".

Big Data offer science funders a tremendous opportunity to capture those links, precisely because the causal links are often so long and tenuous that relying on manual, individual reporting is, quite simply, destined to fail. The Science of science policy community has been developing a body of knowledge about how to think about and identify those links, rather than just saying,

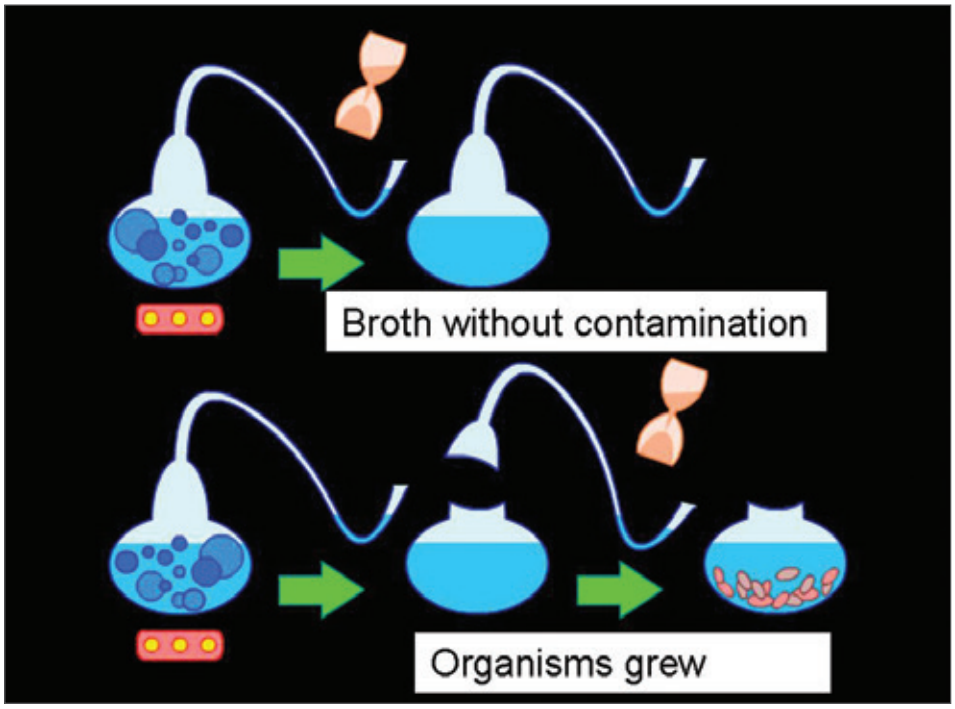


Figure 1: Illustration of swan-necked flask experiment used by Louis Pasteur to test the hypothesis of spontaneous generation.

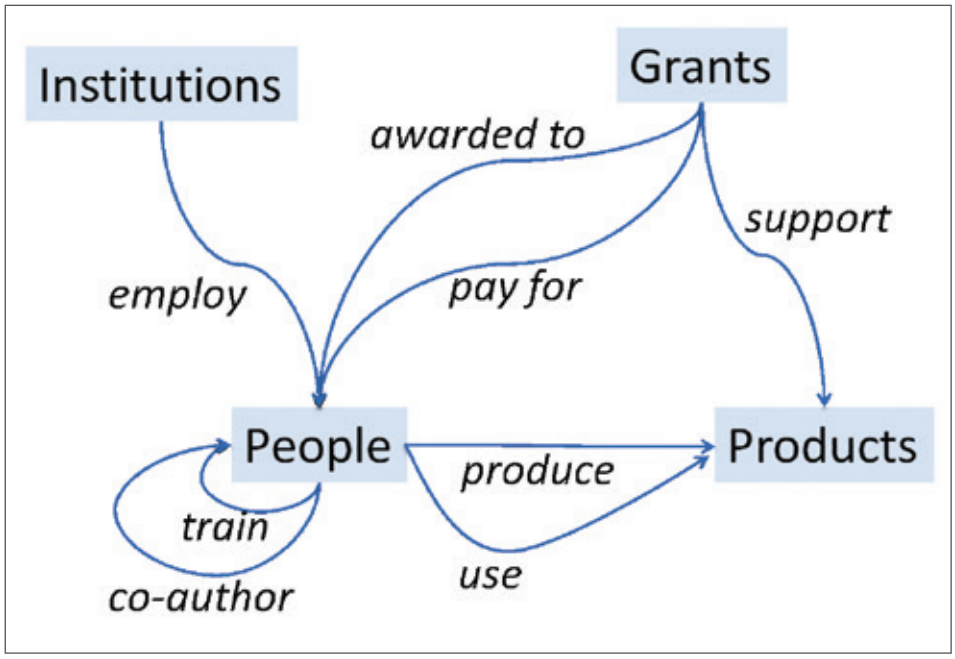
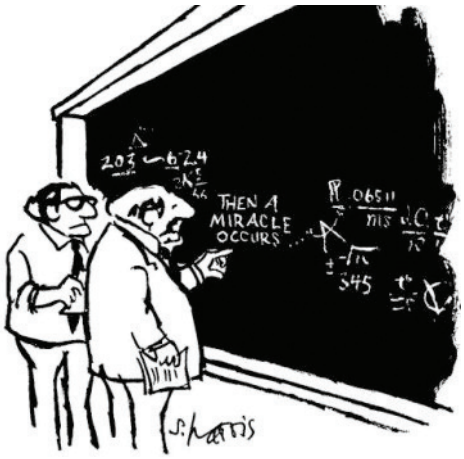


Figure 2: The practice of science (source: Ian Foster).



Copyright 2006 by Sidney Harris

as the cartoon (left) would have it "that a miracle occurred". The Summer issue of the Journal of Policy Analysis and Management (5), from which the quote was drawn, features articles that document what a science of science policy means in practice; namely, bringing the same intellectual set of models, tools and data to science policy as have been brought to labor, education, and health policy (and many others) (6). The September NSF SciSIP Principal Investigator conference (7) will demonstrate how far this new scientific field has come in moving towards more theoretically grounded metrics – in many cases by both building on the impressive work done by bibliometricians, and working with experts in the field. And the STAR METRICS program has built on the efforts of that community to begin to provide a linked data infrastructure on which those metrics can be founded (8).

In summary, Big Data offers an enormous opportunity to advance the science of science policy. Making the links, so that science funders have new understanding of what is needed to foster science, will enable new light to shine on what has hitherto been a rather black box within which miracles occurred.

References:

1. Lane, J. (2010) "Let's make science metrics more scientific", Nature 464, 488–489.

2. Furman, J. L., Murray, F. & Stern, S. (2012) "Growing Stem Cells: The Impact of Federal Funding Policy on the U.S. Scientific Frontier", J. Pol. Anal. Manage. 31, 661–705.

3. File: Experiment Pasteur English.jpg - Wikipedia, the free encyclopedia. at <http://en.wikipedia.org/wiki/File:Experiment_Pasteur_English.jpg>

4. Gertler, P.J., Martinez, S., Premand, P., Rawlings, L.B., & Vermeersch, C.M.J. (World Bank 2011) "Impact Evaluation in Practice", at http://siteresources.worldbank.org/EXTHDOFFICE/Resources/5485726-1295455628620/Impact_Evaluation_in_Practice.pdf

5. Journal of Policy Analysis and Management - Volume 31, Issue 3 - Summer 2012 - Wiley Online Library. at <http://onlinelibrary.wiley.com/doi/10.1002/pam.2012.31.issue-3/issuetoc>

6. Lane, J. & Black, D. (2012) "Overview of the Science of Science Policy Symposium", J. Pol. Anal. Manage. 31, 598–600.

7. NAS CNSTAT SciSIP Principal Investigator Conference. (2012).at <http://www7.nationalacademies.org/cnstat/SciSIP%20Invitation.pdf>

8. Largent, M. A. & Lane, J. I. (2012) "Star Metrics and the Science of Science Policy", Review of Policy Research 29, 431–438.