G Model
RESPOL-3100; No. of Pages 13

ARTICLE IN PRESS

Research Policy xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

New linked data on research investments: Scientific workforce, productivity, and public value

Julia I. Lane[a,*], Jason Owen-Smith[b], Rebecca F. Rosen[c], Bruce A. Weinberg[d]

[a] American Institutes for Research, USA, IZA, BETA University of Strasbourg, CNRS, France and University of Melbourne, Australia
[b] University of Michigan, USA
[c] American Institutes for Research, USA
[d] Ohio State University, IZA and NBER , USA[1]

ARTICLE INFO

Article history:
Received 31 December 2014
Accepted 31 December 2014
Available online xxx

JEL classification:
C8
O3
J4

Keywords:
IRIS
UMETRICS
STAR METRICS
Science of science policy
Linked data
Scientific workforce
Scientific networks
Occupation classification

ABSTRACT

Longitudinal micro-data derived from transaction level information about wage and vendor payments made by Federal grants on multiple US campuses are being developed in a partnership involving researchers, university administrators, representatives of Federal agencies, and others. This paper describes the UMETRICS data initiative that has been implemented under the auspices of the Committee on Institutional Cooperation. The resulting data set reflects an emerging conceptual framework for analyzing the process, products, and impact of research. It grows from and engages the work of a diverse and vibrant community. This paper situates the UMETRICS effort in the context of research evaluation and ongoing data infrastructure efforts in order to highlight its novel and valuable features. Refocusing data construction in this field around individuals, networks, and teams offers dramatic possibilities for data linkage, the evaluation of research investments, and the development of rigorous conceptual and empirical models. Two preliminary analyses of the scientific workforce and network approaches to characterizing scientific teams ground a discussion of future directions and a call for increased community engagement.

"The ITG undertook a literature review to determine the state of the science to date. A questionnaire was circulated to Federal agencies to ascertain what methods are currently in use for programmatic investment decisionmaking; as well as to ask what tools and resources are needed by Federal agencies that are currently unavailable. The ITG found that. . .the data infrastructure is inadequate for decisionmaking" (National Science and Technology Council, 2008) *emphasis added.*

"The working group was frustrated and sometimes stymied throughout its study by the lack of comprehensive data regarding biomedical researchers. The timeframe and resources of the study did not allow for comprehensive data collection or the implementation of a comprehensive model of the biomedical workforce. It is evident from the data-gathering and analyses undertaken by the working group that there are major gaps in the data currently being collected on foreign-trained postdoctoral researchers and those who work in industry." (NIH Biomedical Research Workforce Working Group, 2012).

1. Introduction

Internationally, public support for science and thus the details of science policy have come to depend on evaluating the results of research. In addition to measures of productivity, establishing the economic impact and public value of investments in R&D is of particular concern. The Research Assessment Exercise in the United Kingdom places tremendous emphasis on scholarly production, as does the Excellence in Research Australia program (Jensen and Webster, 2014; Owens, 2013). The United States has focused both on measuring scientific and economic impact. The policy focus in

* Corresponding author. Tel.: +1 3017181621.
E-mail address: jlane@air.org (J.I. Lane).

[1] This research was supported by NSF SciSP Awards 1064220 and 1262447; NSF Education and Human Resources Award 1348691; NSF NCSES award 1423706; NIH P01AG039347; the US Census Bureau, and the Ewing Marion Kaufman and Alfred P. Sloan Foundations. Data were generously provided by the Committee on Institutional Cooperation and its member institutions. We thank Wei Cheng, Cameron Conrad, Russ Funk, Christina Jones, Felix Kabo, Evgeny Klochikhin, Yulia Muzyrya, Joe Staudt and Michelle Yin for research support, Greg Carr, Marietta Harrison, David Mayo, Mark Sweet and Stephanie Willis for help with data issues, and Barb McFadden Allen, Jay Walsh, Roy Weiss and Carol Whitacre for their overarching support. The research agenda draws on work with many coauthors, but particularly work with Paula Stephan, Jacques Mairesse, and Michele Pezzoni.

http://dx.doi.org/10.1016/j.respol.2014.12.013

Please cite this article in press as: Lane, J.I., et al., New linked data on research investments: Scientific workforce, productivity, and public value. Res. Policy (2015), http://dx.doi.org/10.1016/j.respol.2014.12.013

Japan has been on rebuilding public trust in the value of science (Arimoto and Sato, 2012).

The conceptual framework for implementing impact evaluations in most policy areas is well understood – there needs to be a theory of change and a well-defined counterfactual (Gertler et al., 2012), but by and large no consensus framework exists. In the emerging science of science policy field, one key theory of change contends that funding interventions affect the complex interactions of scientists, which shape collaborative networks. The structure, composition, and content of those networks in turn influence discovery and training to provide the mechanism whereby scientific and economic impact is achieved. The approach is an advance relative to both bibliometric analysis and accounting frameworks. Bibliometric analysis, which uses sophisticated techniques to study documents, has neither an explicit theory of change nor a counterfactual and was not designed to be used for research evaluation (Cronin and Sugimoto, 2014; Lane, 2010). Accounting approaches, which attempt to tie results to individual grants in order to calculate a straightforward return on investment, confuse the intervention, funding, with the object of interest itself, and thus inherently muddy efforts to define counterfactuals (Lane and Bertuzzi, 2011).

The empirical framework for impact evaluations is also equally well understood – a typical approach is to build a longitudinal dataset that measures baselines, mediating and moderating factors, and outcomes; that dataset is then used and augmented by a community of practice. This special issue of Research Policy moves such a framework forward for the field of science of science policy by identifying hitherto unexamined data and by informing the scientific community about new initiatives. It is likely that future empirical advances can then, in turn, inform a new conceptual framework for science and innovation policy.[2]

Developing that framework is an urgent matter. In the US, the Science of science policy has been active since 2006 (Marburger, 2005).[3] Yet the United States House Committee on Science, Space and Technology has questioned decisions made on individual grants in political science (Mole, 2013) and the merits of science funding as a whole are regularly challenged (MacIlwain, 2010).

This paper surveys the current landscape. It also describes a large scale, open resource that is being built in the United States, called UMETRICS and sketches two exemplary use cases for these data that is based on using new measures of the workforce to map out the scientific networks that underpin federally funded research. It highlights the engagement of a community of practice in the design of the data infrastructure, particularly in classifying occupations and analyzing collaboration networks. It concludes by discussing the importance of engaging the larger community of scholars and practitioners in the establishment of an institute that provides a sustainable data infrastructure to support scientifically rigorous and practically applicable science and innovation policy research.

2. Conceptual framework and the current landscape

This section outlines the conceptual framework that underlies our efforts and the current data landscape. The goal of the framework is to answer questions like "What have we learned about NSF-funded research?" and "What is the economic impact of research funding?" (Walsh, 2013). Establishing key descriptive facts about the research enterprise is a necessary step toward validating a conceptual framework that can be responsive to the needs of science policy makers and of the research community. For example, the E-Government Act of 2002 (§207), although honored more in the breach than in the observance, requires Federal agencies documenting R&D investments to develop policies to better disseminate the results of research performed by Federal agencies and federally funded research and development centers.[4] The ability to systematically measure and assess the dissemination and use of findings is a compelling interest for policy-makers, domain scientists, and researchers concerned with social and economic returns.

2.1. Conceptual framework

The conceptual framework we elaborate here identifies individual researchers (or the research community consisting of networks of researchers) as the analytical unit of interest. Here, the theory of change is that there is a link between funding (WHAT is funded) and the way in which networks and teams are assembled by the strategic actions of researchers (WHO is funded). We next link features of networks and teams to the process of science, its products, and their transmission. The transmission of scientific products (discoveries) through the movement of people or the expedient of publication and patenting in turn generates social, economic and workforce "impacts." Institutions both administer grants and provide the material and intangible infrastructure necessary to produce science. Of course, the activities of researchers can be aggregated in multiple ways, since they act both on their own or as members of larger teams and communities, to produce, communicate and utilize scientific knowledge and discoveries.

The framework we propose stands in sharp contrast to that commonly used by science funders, who – consistent with their mandate to manage research investments rather than document their returns – emphasize individual grants to the virtual exclusion of people, teams, and the later use of scientific products. Hence their primary unit of analysis is the grant, and research administrators spend much time and energy trying to link research grants to research outputs by requiring scientists to acknowledge specific grants and report results on a grant-by-grant basis. The science of science policy framework recognizes that the social organization and work practices of cutting edge science do not fall cleanly within individual projects bounded by particular goals and clear starting or ending dates. Most of the work of discovery and training takes place in collaborative groups that encompass multiple overlapping projects. In practice, the work of individuals and teams is supported by and integrates a pastiche of grants that serve multiple purposes and often span several funding agencies. Even though the primary lever for policy makers to influence the character, goals or uses of science is funding individual projects, the implications and effects of new funding arrangements or incentives can only be fully understood in the context of the individual and collective careers that are the cornerstone of contemporary science and training. Misunderstanding this basic view will lead to misspecification of any analysis.

While it shares a substantive focus on the diffusion and utilization of discoveries and substantial concern with publications and patents as important scientific outputs, the framework we propose also contrasts sharply with the bibliometric literature. That work largely focuses on counting and evaluating the impact of written artifacts that formally codify discoveries. In this model the publica-

[2] By the science and innovation enterprise, we mean the science and innovation ecosystem writ large – from funders (public and private) to researchers (in academia, government, and industry) to the organizations that hire people with scientific training and/or draw on science and innovation to produce commercial products.

[3] The US Science of Science Policy combines a Federal interagency group on the Science of Science Policy charged with identifying policy questions and a scientific research program at the National Science Foundation charged with advancing the Science of Science and Innovation Policy.

[4] http://www.gpo.gov/fdsys/pkg/PLAW-107publ347/pdf/PLAW-107publ347.pdf

Please cite this article in press as: Lane, J.I., et al., New linked data on research investments: Scientific workforce, productivity, and public value. Res. Policy (2015), http://dx.doi.org/10.1016/j.respol.2014.12.013