9-1-2012

Guilding investments in research: Using data to develop science funding programs and policies

Norman S. Bravemen
Braveman BioMed Consultants LLC

Follow this and additional works at: https://www.researchtrends.com/researchtrends

Section 3: Guiding Investments in Research

Using Data To Develop Science Funding Programs and Policies*

Norman S. Braveman, Ph.D.
President, Braveman BioMed
Consultants LLC

One important goal of organizations that provide funds for biomedical and behavioral research is to encourage and support research that leads to more effective health promotion, better disease prevention, and improved treatment of disease. They do this in order to provide a scientific evidence base. In order to ensure that an organization or its programs are effectively moving science forward toward this objective, organizations that fund research must continually assess and re-assess their goals, directions and progress. While there are a variety of ways that funding organizations can carry out these program assessments, there are several discrete and interlinked components common to all approaches including development of: a strategic plan that identifies organizational values, mission, priorities and objectives; an implementation plan listing the timelines, benchmarks, mechanisms of implementation, and the sequence of events related to the elements of the strategic plan; a logic model, based on information gained from all stakeholders, which identifies inputs or available resources that can be used along with expected outcomes from the organization's activities; a gap analysis, an assessment of progress in reaching organizational goals as well as in carrying out the implementation plan by addressing questions about the current position of the organization in relation to where it expected or wanted to be.

In the process of conducting a gap analysis the organization also addresses specific questions about the current state-of-the-science along with pathways to scientific advancement in terms of what is needed to move science ahead, along with identifying barriers to and opportunities for progress. Nevertheless, most program assessments by funding organizations use what I call 'demographic information', that is information that answers questions on the number of grants in a portfolio, how much is being spent on a particular funding program,

the mix of grant mechanisms (e.g. basic vs. translational vs. clinical research; investigator initiated vs. solicited research; single project grants vs. large center multi-center grants), and the number of inventions or patents resulting from research supported by any individual portfolio of grants or group of grant portfolios.

While these kinds of measures may be excellent indicators of progress of an organization, with the exception of information about inventions and patents, they seem at least one step removed from measuring the impact of an organization's grant portfolios on the content of science itself. In order to maximize the impact of organizational activities and programs on progress in science, the analysis should use science itself as the data that guides the planning, development and implementation of programs and policies. It's what the scientists whose research the organization may be supporting do in justifying the next step in their research.

There are times when organizations analyze the science of grants in their portfolios by capturing key words in the titles, abstracts, progress reports, and/or grants or grant applications. These are generally tracked over time by program analysts. While the program analysts are typically highly experienced and/or trained, they carry out the analysis by hand and from time to time, from document to document, or from person to person the algorithm they use in classification and categorization can shift in small ways. Such shifts introduce a source of variability that can reduce the reliability and perhaps even the validity of the final results. Moreover, analyzing science by hand is a long, tedious, and expensive task. So our tendency is to do this kind of detailed analysis infrequently…clearly not in 'real time' as seems to be what is needed in this age of fast-paced discovery.

Scientific Fingerprinting

Fortunately, the technology now exists that will allow us to analyze the content of science in a valid, reliable and timely way that overcomes many of the problems that crop up when we do it by hand. More than that, because this approach is computer-based, and therefore fast and reliable, its use allows us to carry out assessments often and on a regular basis. The approach I'm referring to involves the formal textual analysis of scientific concepts and knowledge contained within documents such as strategic and implementation plans, grant applications, progress reports, and the scientific literature. We refer to the output of the textual analysis of a document as a 'scientific fingerprint' or simply a 'fingerprint'.

Without going into the details of the underlying processing, a fingerprint is a precise abstract representation of a text that allows us to look into the text or content rather than only looking at the metadata or demographics. Because fingerprinting is concept driven and not keyword driven and because it uses an ontology (i.e. A is an example of B) as its base, it is not necessary to have a term appear in a document in order for it to be part of a fingerprint. For example, it is possible for a document to contain all of the diagnostic characteristics of a disease but not the name of the disease in order for the disease name to appear in the fingerprint. The only requirement is that the diagnostic characteristics be identified as examples of the named disease somewhere in the scientific literature that makes up the database that is searched. Further, the concepts or weights given to individual concepts comprising a scientific fingerprint of textual content can be adjusted to fit the views of experts in the field. Thus they are not adopted blindly or without validation by experts.

Fingerprinting uses as its base for comparison and analysis the entirety of the Elsevier Scopus database consisting of 45.5 million records, and the information used to develop a fingerprint can be captured relatively easily. Because it is computer-based the textual analysis of the grant portfolio is also much faster than when it is done by hand, thus allowing us to continually assess and reassess an organization's scientific grant portfolio. It is applicable to any document and can be used at any stage of evaluation and program development. In short, it is possible to carry out continual ongoing real time assessments using fingerprinting. Finally, as science changes, as reflected in the scientific literature itself, the fingerprint profile of a given area of science will change, and those changes can be documented and used in an analysis of progress both in science and in the organization's grant portfolio.

While the use of this approach to textual analysis is in its infancy, fingerprinting can allow organizational decision making to be based on the state-of-science, help align organizational goals with the current state-of-the-science, and clarify the organization's role and contributions within a specific area of science. As such, when coupled with demographic data charting the organization's performance it can provide a fuller picture of the current role of the organization in moving science forward, and in the possible role that the organization can play in future scientific development.