Tech Mining Cited References to Understand the Influence of Journal Articles on reports of the US National Research Council.[1]

Jan Youtie[*], Barry Bozeman, Sahra Jabbehdari, Andrew Kao

[*]*jan.youtie@innovate.gatech.edu*
Enterprise Innovation Institute, Georgia Institute of Technology, 75 Fifth Street, Suite 300, Atlanta, Georgia 30308 (USA)

Introduction and Background

Science and technology indicators based on citation have long attracted attention of policy makers seeking a way to discern what research is or is likely to be of greatest quality. This focus on citation occurs in spite of caveats from the citation analysis community that citations are not always a good measure of research quality (Glänzel et al., 2006). But even though citations are widely used to evaluate scientific research, they have not attracted much attention in terms of their application in scientific and technical policy studies.

The aim of this study is to address this gap by examining the extent to which scientific and technical research papers are used in policy studies. We propose a tech mining method to extract and code citations of scientific and technical information (STI) in a given policy report. For purposes of this study, we are using the term STI in a manner somewhat narrower than is typical in the literature (see McClure, 1988; Walker and Hurt, 1990). We are concerned here with open scientific and technical literature appearing in peer-reviewed academic journals. We draw on public values literature (Bozeman 2002, 2007) oriented toward the value of bringing the "best" information to bear on critical public decision-making. Of course, there is not always a consensus on the need to bring STI into public decision making about science policy. In scholarly journal articles, STI is the predominant information source cited, with the exception of some articles citing newspaper articles which are used in part to indicate the relevance of the study to a broader audience (Hicks and Wang 2013). In policy reports, however, STI is used alongside a diversity of other information bases such as: personal opinion of policy maker (which may draw on expressed political values, self-interest, experiential knowledge, or other sources) and informal and formal communications (through the committee hearing process) with persons knowledgeable about the area, as well as newspaper articles, other policy reports, information in Websites, and the like. However, it is the rare professional scientist or engineering researcher who does not feel that STI should have a prominent, and perhaps, even a privileged place and the corpus of information sources related to public decision-making. Thus, there is a long history of lamentations that STI should have a wider use in policymaking.

This paper focuses on use of STI in arguably a quite important science and technology policy domain in the United States: the National Research Council (NRC). The NRC performs research work for the production of reports on science and technology issues within the private non-profit National Academies. The NRC's social, political and organizational aspects have received surprisingly little research attention.

Data Sources and Methodology

The dataset for examining the use of STI in NRC policy reports is 589 NRC reports published from 2005-2012. This sample is directed toward single-issue empaneled studies because such studies are more likely to use STI. We obtained a list of NRC reports from the annual reports on the National Academies website for each of the years under analysis and downloaded PDFs of these reports. For each of these reports, we collected and coded information available from the PDFs about the study (e.g., size of the report, report policy area), about the committee chair and members (e.g., affiliation with academia, business, government), and about the references (e.g., STI journal articles, total number of references). We have specifically explored two indicators of STI usage: one based on the share of the total number of references in the report that are STI and a second based on the number of STI references divided by the total number of pages in the report.

We obtained this information through a combination of (1) automated cleaning and matching methods (using Excel macros) to highlight and match report references to a thesaurus of journal articles and (2) manual checking and coding of the references, by two separate coders, to determine whether or not the references were STI. Some NRC reports had a separate list of cited references; these lists presented the references in standardized form and were therefore most easy to clean, match and code as STI. Many, particularly older reports, used a footnote convention. This convention required extracting the footnotes before coding them as STI. In all, there were more than 120,000 references in these 589 reports.

Results

NRC reports ranged from 16 pages to 650 pages with a median of 164 pages and a mean of 188 pages. All but three NRC reports had references and 88% of the reports had references involving STI. The median NRC report had 126 total references and 30 STI references while the mean was 205 total references and 89 STI references, suggesting skewedness with some reports extensively using STI (i.e., one report used 2,330 total cited references of which 1,440 referenced journal articles). We normalized the measure by reporting the percentage of citations that are STI to address these distributional issues. The median and mean proportion of STI were 0.26 and 0.30 respectively

We note that the use of STI varies substantially by policy area. It is less prevalent in the defense area and more prevalent in the natural resources/environmental area. NRC reports are developed through a committee structure and the use of STI also differs based on the extent of academics (versus private industry, government or other sector representatives) on NRC committees. These results suggest that STI does play a significant role in the science policy making process of the NRC.

References

Bozeman, B., 2002. Public-value failure: when efficient markets may not do. *Public Administration Review* 62 (2), 145–161.

Bozeman, B., 2007. *Public Values and Public Interest: Counterbalancing Economic Individualism*. Georgetown University Press, Washington, D.C.

Glänzel, W., Debackere, K., Thijs, B., & Schubert, A. (2006). A concise review on the role of author self-citations in information science, bibliometrics and science policy. *Scientometrics*, *67*(2), 263-277.

Hicks, D., & Wang, J. (2013). The New York times as a resource for mode 2.*Science, Technology & Human Values*, *38*(6), 851-877.