

What Is Bibliometrics and Why Should You Care?

Melissa K. McBurney
melissa.mcburney@pnl.gov

Pamela L. Novak
pam.novak@pnl.gov

Pacific Northwest National Laboratory
P.O. Box 999
Richland, WA 99352

Viewed in the aggregate, a company's publications represent one of its greatest business assets. Can professionals in the information and communication fields help their company understand the return on its investment in publishing? The field of bibliometrics studies publication patterns by using quantitative analysis and statistics. Bibliometrics can be either descriptive, such as looking at how many articles your organization has published, or evaluative, such as using citation analysis to look at how those articles influenced subsequent research by others. Counting publications can be useful for doing some comparisons, but citation analysis allows you to look at the impact those articles have had on others by determining how often they are cited. Citation analysis can also show what journals, organizations, and even countries have high impact in different fields of research.

The Institute for Scientific Information (ISI) has been a leader in the citation analysis field since 1961, when ISI published the first Science Citation Index. Pacific Northwest National Laboratory (PNNL) has been using data from ISI for both descriptive and evaluative purposes. This data is used to track what the researchers at the Laboratory are writing and then comparing research groups within the organization over a period of years to identify trends and opportunities. PNNL has also used citation analysis to explore what organizations and academic institutions are doing research in certain fields for partnering opportunities.

Keywords: bibliometrics, citation analysis, impact factor

Introduction

Bibliometrics - the use of statistical analyses to study publication patterns - covers a wide range of laws and methodologies. For the purposes of this paper, we will be focusing on the use of publication counts, citation analysis, and impact factor to show how Pacific Northwest National Laboratory (PNNL) has been using bibliometrics. If you'd like to learn even more about the field, a bibliography of further reading is provided at the end of this paper.

Currently, PNNL is working with data from the Institute for Scientific Information (ISI) to determine how many articles Laboratory authors are publishing, what journals and fields they are publishing in, and how much impact (i.e., readership and influence) those journals have. In the future, we'd like to take this a step further and look at the impact of individual articles and contributors within the Laboratory. This paper will provide some background on citation analysis, the Institute for Scientific Information, and impact factors. We will then focus on PNNL's current use of bibliometric data and our plans for future study in this area.

Citation Analysis: "An Association of Ideas"

Citation analysis looks at the relationship between a paper and the papers it has cited, based on the assumption that articles citing each other are somehow related. Relationships studied by citation analysis include field, authors, institutions, and countries.

Citation analysis would not be possible without citation indices to gather the information together. The original idea of citation indexing is credited to *Shephard's Citations*, which was first published in 1873 as an index to law literature; Eugene Garfield, founder of ISI, is credited with taking this idea and applying it to the science literature.

"Here, in Garfield's own words, is the essence of the idea: 'Almost all the papers, notes, reviews, corrections, and correspondence published in scientific journals contain citations. These cite—generally by title, author, and where and when published—documents that support, provide precedent for, illustrate, or elaborate on what the author has to say. Citations are the formal, explicit linkages between papers that have particular points in common. A citation index is built around these linkages. It lists publications that have been cited and identifies the source of the citations. Anyone conducting a literature search can find from one to dozens of additional papers on a subject just by knowing one that has been cited. And every paper that is found provides a list of new citations with which to continue the search.'" (Cronin & Atkins, 2000)

Garfield's original idea to create "an association of ideas index" (Garfield, 1955) has grown into a huge business with the original data being built upon, repackaged and used in ways that he never imagined. While some of these uses have created controversy, "[t]he ISI has always acknowledged the limits of citation analysis. But the lure of numbers has proved irresistible to those charged with judging scientists' work." (Adam, 2002) Today citations are used "to rank journals, research groups, laboratories, universities, and nation states. And not infrequently, they are factored into the promotion and tenure decision-making process, a perennial topic of lively debate." (Cronin & Atkins, 2000) Some feel that the use of citation analysis is more objective

than the peer review process: “bibliometric performance indicators allow substantial improvement of the peer review based evaluation by supplying new elements in terms of objective analysis of research output and impact.” (van Raan, 1996)

When using citation analysis, several caveats must be respected. “There has been a tendency over the years to reify the meaning of citations. Citation counts are not measures of quality in and of themselves. When the measure is used as an indicator of the quality of an individual’s work without examining the ‘meaning’ of the citations to the work, particularly the type of citation and the characteristics of those who cite the work, the probability increases of drawing inappropriate inferences about the impact or quality that an individual’s scholarly and scientific work has had on his specialty or field.” (Cole, 2000) For example, there are “different publication and citation characteristics in the different fields of science,” (van Raan, 1996) there is a need for “cleaning and unification of authors names, and in particular addresses of author affiliation,” (van Raan, 1996) and “corrections for self-citations and/or ‘in-house’ citations.” (van Raan, 1996) The reason for the citation also needs to be taken into account. Sometimes a citation to someone’s work is not at all an endorsement! It may instead be a harsh critique of the methodology or conclusions.

Journal Impact Factors and Their Abuses

Another tool that is used in bibliometrics is the journal impact factor published by ISI in an annual report called *Journal Citation Reports (JCR)*. The impact factor is calculated by dividing the number of current citations a journal receives for articles published in the two previous years by the number of articles published by that journal in those same years. As an example, in the 2000 edition of *JCR* the impact factor for a journal would be calculated by adding up the citations received in 2000 to articles published in 1998 and 1999 and dividing by the number of articles published in 1998 and 1999. An impact factor of 1 would mean that on average, every article published in that journal over the two-year period in question was cited once. Obviously then, the higher the impact factor, the more frequently a journal’s articles are cited.

As with citation analysis, there is much controversy about how impact factors should and should not be used. Researchers are justifiably concerned that policy and funding decisions will be made using impact factor scores by people who do not understand how those scores are calculated. According to Amin and Mabe, “The use of the absolute value of an impact factor to measure quality should be strongly avoided . . . because the long-term average trends in different fields vary.” (Amin & Mabe, 2000) For example, the impact factor should not be used to compare journals in different fields because authors in one field may cite fewer articles in their papers than those in another field. “In general, fundamental and pure subject areas have higher average impact factors than specialized or applied ones. The variation is so significant that the top journal in one field may have an impact factor lower than the bottom journals in another area.” (Amin & Mabe, 2000) The type of journal makes a difference as well -- those journals with mainly review articles will be cited more often. Because of these same caveats, ISI and others also warn against using the impact factor to evaluate the work of individual authors. “‘These scores were never designed by ISI to be proxies for the influence of papers or, when aggregated, the work of individuals,’ says David Carter, the company’s vice-president for corporate communications.” (Adam, 2002)

So how should the impact factor be used? “Impact factors, as one citation measure, are useful in establishing the influence journals have within the literature of a discipline.” (Amin & Mabe, 2000) “Publishers have eagerly latched onto these data, using favourable impact factors in promotional materials for their journals, and librarians have found them to be a convenient guide in deciding which journals to subscribe to.” (Adam, 2002) The impact factor is often used instead of paying for the more expensive citation data for individual authors.

In more controversial uses, some universities in Germany are heavily weighting impact factor when allocating departmental funding; in Finland, “government funding for university hospitals is partly based on publication points, with a sliding scale corresponding to the impact factor of the journals in which researchers publish their work.” (Adam, 2002)

When any such quantitative measure is used in decision-making to the exclusion of important qualitative factors, those decisions are vulnerable to dispute. When impact factor scores are used to determine what research will and will not be funded, the results can be very damaging to the progress of science. Unfortunately, such abuses have given bibliometrics a bad name in many circles. Yet bibliometric data, when used primarily to describe rather than to evaluate the research trends of an organization, can be helpfully revealing.

Bibliometric Studies at PNNL

Pacific Northwest National Laboratory began purchasing publication data from ISI in 1996 to assist in tracking a Laboratory-level goal on publishing. These data have since been used both for publication counts for articles authored by PNNL staff and to help correct citation data in an internal publication tracking system. We decided to start with the available data in conducting publication trend analyses. Our primary interest was to create, if possible, a “portrait” of the Laboratory’s scientific reputation and determine how this portrait has changed over time.

The data for fiscal years 1996-2001 (October 1996 to September 2001) were used for the analyses. During this time the Laboratory published in 817 separate journals or conferences for a total of 3,256 articles. Impact factors for all years were assigned to these titles using the 2000 edition of the *JCR* (i.e., we did not match up the 1996 impact factor with the 1996 articles). Only 2,708 of the articles ended up with an impact factor since some of the publications are not assigned one by ISI. Using the *JCR*, the journals with impact factors were assigned to as many as four different subjects for a total of 118 subjects. If a journal was in the top ten in the subject areas assigned, the rank was also noted. For ease of analysis, the 118 subjects were then grouped into eight broad subject areas: chemistry, computer science, earth sciences, engineering, life sciences, multidisciplinary, physics, and statistics/mathematics.

The data was analyzed using linear regression analysis. Image plots and line graphs were produced that help to visually display PNNL’s publication activities since 1996. Trends over time were analyzed to determine statistical significance. These analyses identify areas in which PNNL is improving, remaining constant, or decreasing. As we share these with PNNL management and research staff, those measures of significance will be critical in avoiding hasty

Reflections on Communication

conclusions and decisions about how to “improve” the Laboratory’s publishing portrait in future years.

While we are not yet ready to share the actual results of our analysis, we can report our conclusions from this initial activity:

- Even though no specific goals or targets for publishing have been provided to research staff, the results show a strong and growing PNNL presence in several scientific disciplines that are central to our mission. This correlates well with the interests of individual research staff to publish as a means of gaining recognition among their peers. Research staff decide which journals to publish in based in large part on the journals they themselves read and esteem. In other words, the system is somewhat self-sustaining and works without specific direction from upper management.
- Some researchers publish primarily in the journals of other disciplines, and so their work was not as obvious in the results of this analysis. Statistics and computer science are two excellent examples; both of these disciplines tend to be applied to problem-solving in other fields such as engineering or environmental science. In order to more accurately gauge PNNL’s portrait in these applied disciplines, we will need to analyze the publication history of individual PNNL researchers in these disciplines.
- Some of PNNL’s work is being published in newer journals because the fields are just now emerging. Such journals may be very influential but do not have impact factors assigned by ISI because there is simply not enough citation data yet. Again, we will need to examine data at the individual publication level to understand PNNL’s position in these new fields. We will also have to wait a few years for ISI to collect the necessary citation data.

Future Directions in Bibliometrics at PNNL

In addition to the analysis of individual researchers’ publication metrics described above, we recognized that having some external benchmarks would greatly help us validate our conclusions. We hope during the next year to enlist several of our sister national laboratories in performing the same types of bibliometric analysis. We also plan to enlist research staff directly in validating our assumptions about how publishing works at PNNL. Our hope is that the research staff will join with us to propose a plan for using bibliometric analysis effectively. We are intrigued by the unique perspective we have gained through this study, and we believe that as long as we avoid the pitfalls of using the results to evaluate or prescribe, bibliometrics can be a valuable tool for describing and promoting our scientific productivity.

Acknowledgments

We are gratefully indebted to Brett G. Amidan and Thomas A. Ferryman of PNNL’s Data Mining & Inferential Analysis Group for their expertise in developing the methodology and conducting the statistical analysis for this study. Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by Battelle Memorial Institute under contract DE-AC06-76RL01830.

References

- Adam, D. The Counting House. *Nature* 415: 726-729.
- Amin, M. and M. Mabe. Impact Factors: Use and Abuse. *Perspectives in Publishing*. No. 1, Oct.2000.
- Cole, J. R. "A Short History of the Use of Citations as a Measure of the Impact of Scientific and Scholarly Work." In B. Cronin and H.B. Atkins, eds. *Web of Knowledge: A Festschrift in Honor of Eugene Garfield*. Medford, New Jersey: Information Today, Inc., 281-300, 2000.
- Cronin, B. and H.B. Atkins. "The Scholar's Spoor." In: B. Cronin and H.B. Atkins, eds. *Web of Knowledge: A Festschrift in Honor of Eugene Garfield*. Medford, New Jersey: Information Today, Inc., 1-7, 2000.
- de Solla Price, D.J. Networks of Scientific Papers. *Science*. 149: 510-515, 1965.
- Garfield, E. Citation Indexes for Science: A New Dimension for Documentation through Association of Ideas. *Science*. 122: 108-111, July 1955.
- Garfield, E. Citation Indexing for Studying Science. *Nature*. 227: 660-671, 1970.
- Hertz, D.H. "Bibliometrics, History of the Development of Ideas In." In B. Kent, ed. *Encyclopedia of Library and Information Science*. New York: Marcel Dekker, 144-219, 1987.
- Osareh, F. Bibliometrics, Citations Analysis and Co-Citation Analysis: A Review of Literature I. *Libri*. 46: 149-158, 1996.
- Osareh, F. Bibliometrics, Citations Analysis and Co-Citation Analysis: A Review of Literature II. *Libri*. 46: 217-225, 1996.
- van Raan, A. Advanced Bibliometric Methods as Quantitative Core of Peer Review Based Evaluation and Foresight Exercises. *Scientometrics*. 36: 397-420, 1996.

About the Authors

Melissa McBurney is the Assistant Director for Electronic and Technical Services for Pacific Northwest National Laboratory's Hanford Technical Library. Her department acquires and catalogs library materials and licenses electronic journals and databases for PNNL staff to access. Melissa has a Masters of Science in Library Science from The University of North Carolina at Chapel Hill and a Bachelors in Political Science from North Carolina State University. She has authored a book on electronic resources for engineering and presented papers on electronic documents collections.

Pam Novak is the manager of PNNL's Scientific and Technical Information Department, whose 140 staff provide publishing services, operate the site library, and manage the company's records.

Reflections on Communication

Pam holds a Masters in English Literature from the University of Virginia and a Bachelors in English from the University of Rochester. She has published on topics ranging from management to corporate web policy, and in 2000 co-authored an invited article on advanced publishing technologies and services for the Wiley Encyclopedia of Electrical and Electronics Engineering.