Essays of an Information Scientist: Of Nobel Class, Women in Science, Citation Classics, and other Essays, Vol:15, p.75, 1992-93 Current Contents, #23, p.5-13, June 8, 1992

Current Comments®

EUGENE GARFIELD
INSTITUTE FOR SCIENTIFIC INFORMATION®
3501 MARKET ST., PHILADELPHIA, PA 19104

Contract Research Services at ISI—Citation Analysis for Governmental, Industrial, and Academic Clients

Number 23 June 8, 1992

Over the last few years, there has been an increasing use of citation data by government, industry, and academia for measuring scientific performance. The unique, massive citation databases of the Institute for Scientific Information® (ISI®) have played a central role in the growth of the field called "scientometrics." Under the direction of Henry Small, ISI's Research Department provides customized contract scientometric services. Clients include a variety of institutions in need of research performance evaluations and other studies.

Varied Scope of Data

A recent example of the scientometric use of ISI's citation data figured prominently in a special survey of European research appearing in the April 24 issue of *Science*.[1] The survey relied in part on ISI data prepared by David Pendlebury of our Research Department, who is the editor of *Science Watch®*.

While citation analysis offers a unique perspective on scientific research, like any tool, it can be misused. An example of a simplistic misuse of citation data is the series of unrefereed news reports on uncitedness in various scholarly fields—including not only social and natural science, but also the arts and humanities.[2,3] A number of commentaries followed those reports including a rebuttal by Pendlebury.[4]

Henry, David, and their colleagues in the Research Department draw on specially created data files to perform analyses that range from simple statistical calculations to sophisticated analyses of research fronts. These files include the nine-million-records *Integrated Citation Database,* the *Hot Article Database,* and the *Research Front Database.* These special files are derived from ISI's main databases—the *Science Citation Index®*, the *Social Sciences Citation Index®*, and the *Arts & Humanities Citation Index®*.

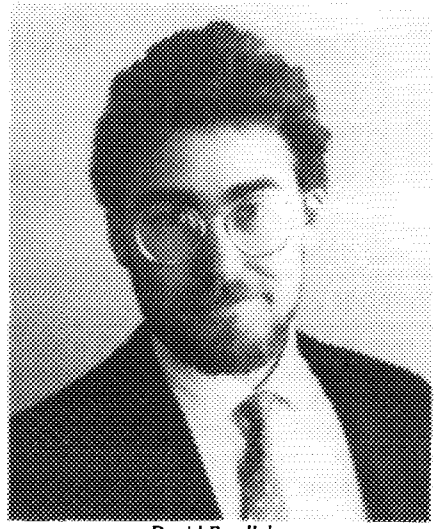
From Henry's group, clients can order just data and make their own analyses, or they can order both data and their appropriate interpretation.

The department can deliver the data from these studies in a variety of formats—print, magnetic tape, diskette, tape cartridge, and, in one recent large-scale study, even on CD-ROM. When required, the data sets can be updated regularly. One of our greatest challenges is to provide an informed perspective on the wide variety of client questions that accompany requests for data.

Small earned his PhD in chemistry and the history of science at the University of Wisconsin (1971). In 1987, he was the recipient of the Derek de Solla Price Medal for his work in scientometrics and information science. A pioneer of co-citation theory,[5] he is the author of more than 60 papers and reports. And, being an amateur jazz musician, Henry is adept at improvisation.

Since the ISI databases are of considerable interest to government and nonprofit agencies in Washington, we are represented there by a full-time specialist. Elizabeth Aversa, ISI's Contract Research Marketing Representative, has prepared the article below, describing the department's capabilities for doing academic, corporate, or

Henry Small



David Pendlebury

government studies. Already this year, our research staff has provided reports and data to more than 25 clients, among them medical institutes, universities, foreign governments, and industrial research labs.

These sophisticated reports are quite varied in scope. In one contract, we were asked to compare the research productivity and impact of a variety of departments at several universities. Another study involved the examination of research productivity at a corporate lab. Still another involved a country-by-country comparison of several dozen research fronts. Backed by a staff of skilled programmers and analysts, ISI was able to deliver not only data, but also informed, balanced interpretations.

Aversa holds a PhD in information systems from Drexel University, granted in 1984. Her doctoral thesis involved citation analysis for patterns of literature aging. Prior to joining ISI in 1990, she taught at several universities and did consulting in library and information science.

The title Elizabeth has chosen for her article is reminiscent of the festschrift edited by Maurice Goldsmith and Alan Mackay on *The Science of Science*.[6] This festschrift was published in honor of John D. Bernal's pathbreaking book *The Social Function of Science*.[7] As reported previously in *Current Contents®*,[8] Bernal significantly influenced quantifiers like Derek de Solla Price. Both these pioneering scientometricians would have been avid users of the new ISI databases. So too would have been historians of science like George Sarton.[9]

Indeed, the work of Henry's group is the culmination of a long line of citation analysts over the last 60 years. Space does not permit a comprehensive listing, but the work of Jack Westbrook on industrial evaluations in 1960[10] is but one of many that could be noted.

For more information, call or write: Henry Small, Director, Research Department, Institute for Scientific Information, 3501 Market Street, Philadelphia, PA 19104. Tel: (215) 386-0100, ext. 1307, or fax (215) 386-6362. Elizabeth Aversa can be reached at (410) 997-3745, or by fax at (410) 740-2335.

* * * * *

My thanks to Paul R. Ryan and Eric Thurschwell for their help in the preparation of this introduction.

REFERENCES

1. **Anderson A.** U.S. juggernaut overwhelms divided European elite. *Science* 256(5056):460-4, 24 April 1992.
2. **Hamilton D P.** Publishing by—and for?—the numbers. *Science* 250:1331-2, 1990.
3. ----------. Research papers: who's uncited now? *Science* 251:25, 1991.
4. **Pendlebury D A.** Letter to editor. (Science, citation, and funding.) *Science* 251:1410-1, 1991.
5. **Small H.** Co-citation in the scientific literature: a new measure of the relationship between two documents. *J. Amer. Soc. Inform. Sci.* 24:265-9, 1973. (Reprinted in: *Current Contents* (7):7-10, 13 February 1974 and *Essays of an information scientist.* Philadelphia: ISI Press, 1977. Vol. 2. p. 28-31.) [See also: **Small H.** Cogitations on co-citations. Citation Classic. Commentary on *J. Amer. Soc. Inform. Sci.* 24:265-9, 1973. *Current Contents/Social & Behavioral Sciences* 24(10):10, 9 March 1992, and *Current Contents/Arts & Humanities* 14(6):20, 16 March 1992.]
6. **Goldsmith M & Mackay A,** eds. *The science of science.* London: Souvenir Press, 1964. 235 p.
7. **Bernal J D.** *The social function of science.* New York: Macmillan, 1939. 482 p.
8. **Garfield E.** J.D. Bernal—the sage of Cambridge. 4S award memorializes his contributions to the social studies of science. *Current Contents* (19):5-17, 10 May 1982. [Reprinted in: *Essays of an information scientist.* Philadelphia: ISI Press, 1983. Vol. 5. p. 511-23.]
9. ----------. George Sarton: the father of the history of science. Parts 1 & 2. *Current Contents* (25):3-9, 24 June 1985; (26):3-8, 1 July 1985. (Reprinted in: *Essays of an information scientist: ghostwriting and other essays.* Philadelphia: ISI Press, 1986. Vol. 8. p. 241-53.)
10. **Westbrook J H.** Identifying significant research. *Science* 132:1229-34, 1960.

Research on Research: Customized Citation Analysis for Governmental, Industrial, and Academic Clients

by

Elizabeth Aversa

ABSTRACT

ISI®'s Research Department is engaged in monitoring, evaluating, and planning the research enterprise for clients in government, universities, and corporations worldwide. These activities include creating customized institutional, journal, country, and topical publication and citation data sets; providing quantitative analyses of these data; and developing new methods and tools for assessing research productivity and trends through the use of bibliographic and citation data. Departmental resources, such as the *Science Indicators*, *Hot Article*, and *Research Front Databases*, are described. SCI-MAP and *Science Watch*® are also reviewed.

The demand for quantitative indicators to measure and evaluate research activity and performance has increased in recent years. This has occurred at a time when two antithetical views of the research enterprise have emerged. On the one hand, support for science—and especially for technology—is now widely accepted as a strategic and necessary investment in a nation's intellectual and economic vitality. On the other hand, financial constraints are being imposed on research producers worldwide. These constraints come in many forms, such as strictures on overhead costs and a variety of other belt-tightening practices by funding agencies. Governmental, institutional, and corporate R&D managers and decision-makers the world over are looking for ways to set research priorities and maximize the return on their research expenditures. To that end, more and more scientists are seeking to measure both inputs and outcomes in the research arena.

One measurable output of research is journal article publication. In addition, one can measure the impact of these articles by determining how often, and by whom, they are cited. ISI®'s Research Department is the chief source for tailor-made data sets of output and impact statistics and of sys-

tematic studies for the assessment of research outcomes.

In the Research Department at ISI, information scientists, computer programmers, and analysts work as a team for clients on a wide range of projects, all of which ultimately rely on special extractions and manipulations of ISI's publication and citation data. Our clients include university administrators, government officials, corporate managers, research planners, and science-policy experts. Unlike other types of research at ISI—for example, on new product development or advanced production methods—the work of the Research Department can perhaps best be described as "research on research."

The Need for a Global View

While those who evaluate research outcomes need both quantitative and qualitative indicators of performance, others such as research planners, policymakers, and publishers need to keep track of new developments in their own and in related fields. This task is increasingly difficult because of the rapid accumulation of new knowledge. What is needed is a global view of research activities, a view that highlights the most prominent features of the ever-changing scientific landscape, one that reveals fast-moving or "hot" fields and new, emerging specialty areas. Innovative analyses based on ISI's database can provide that global view by showing trends in publication and citation rates at the article, specialty, or field level.

Customized Data Analysis

While many reports offer glimpses of research performance in different fields or by particular universities or organizations,[1-2] few offer sufficient detail to meet the needs of managers and decision-makers. It is here that ISI's Research Department can help. We tailor our services to a client's specific needs by providing, for example, data on a single department or lab within an organization, on a group of scientists, or on a field or topic defined by a set of journals, keywords, or cited references.

Because ISI's publication and citation data are international and multidisciplinary in scope, both foreign and domestic research activities can be assessed across all fields of science.

For universities, we routinely provide data and analyses for the evaluation of research efforts at specialty, field, department, and individual levels. Profiles of peer institutions enable university administrators to compare their institution's performance with those of other institutions. Rankings of the citations per published paper (or "impact") from different institutions can produce a provocative picture of research performance in a field. While extreme caution is advised in the use of citation data to indicate the impact of individuals,[3] when performed properly—with expert interpretation, and recognition of potential artifacts and limitations—studies of individuals can also be revealing and instructive.

For corporations, we routinely develop data sets for managers to assess performance, plan strategically to fill gaps in research, and monitor competitor organizations' research outputs. Although publication and citation measures have been termed "lagging indicators" because they are an evaluation of work already completed, they also should be viewed as the scientific community's opinion on the utility of earlier research. As such, these data tell us as much about current science and what is important in the eyes of today's investigators.

Government agencies and private foundations that fund science can use ISI data to indicate how well their research dollars are being spent. The performance of grant applicants versus nonrecipients, the productivity of the reviewers themselves, and the volume and impact of research from the use of government or privately sponsored facilities and data centers are measurable through the use of ISI's data.

On a more strategic level, government bodies often need to target research priorities, compare national or regional research outputs, study transborder collaborative efforts, or assess the state of an entire discipline. ISI's Research Department provides such groups with data and statistical summaries to support their studies.

Another important user group consists of publishers of journals or scientific and

technical books. These clients need custom data on publications that are of specific interest to them. The comparison of one journal's impact to others, the pinpointing of key topics or key researchers for editorial decision-making, or the identification of journals for acquisition or development, are some of the applications the Research Department supports.

For all types of clients, we can provide three levels of service. At one end of the spectrum, custom data sets are delivered for further analysis by the client, either in print, or in tape or diskette formats. At the other end of the spectrum, we build a customized data set for the client, and then analyze, summarize, and interpret the data in the form of a report. Most frequently, however, clients request a combination of data and interpretation.

Elizabeth Aversa

Resources in the Research Department

The types of contract projects undertaken by the Research Department can be envisioned more clearly by describing the information resources that our staff can bring to bear on the client's project.

Our most important resources are the large, specialized mainframe-computer-mounted files or databases.

Three main files serve as the backbone of the department's work: the *Science Indicators Database*, the *Hot Article Database*, and the *Research Front Database*.

Science Indicators Database

The *Science Indicators Database*, also called the *Integrated Citation File*, is the most versatile of our data resources and provides the widest range of applications. This database consists of a compilation of 11 years of ISI's publication and citation data stored on a mainframe computer under a data management system. The file currently consists of about nine million bibliographic records, representing all items indexed in all of ISI's products, including the various editions of *Current Contents*® (*CC*®), the *Science Citation Index*® (*SCI*®), *Social Sciences Citation Index*® (*SSCI*®), and *Arts & Humanities Citation Index*®. It covers the period 1981 through 1991, and is updated semiannually.

The *Science Indicators Database* is stored in an integrated format, meaning that each bibliographic item is linked to other items it cites, as well as to items that cite it. This feature provides a network searching capability unlike any available in an existing online system.

For statistical studies, the advantage of storing ISI's data in this integrated format is that full bibliographic information is available for cited items that have been covered by ISI in earlier years as source items. This makes possible the direct computation of citation counts or other impact measures for *all* authors, organizations, nations, or any other attributes displayed in the ISI source document record.

An item's citation count is an intrinsic property in this integrated format since it points to all papers that made reference to it in more recent items. The citation count for an organization is simply the sum of the citation counts for all bibliographic items on which the organization appears as part of an author address. The citation time-series for an organization can then be constructed by aggregating the time-series of individual items.

For each paper, the *Science Indicators Database* can identify all authors, author addresses (including institutional affiliation, department, city, state, and country), title, journal, volume, page, and year of publication. Annual time-series and cumulated citation counts are available for each paper.

From this file, a variety of data sets and analyses can be generated. Sets of papers by authors from particular institutions, countries, or in specific subject areas or journals can be extracted. Then, different statistics and summaries can be developed: time-series citation impacts (average citations per paper), percentages of cited and uncited articles, total publications and citations, and so on.

The *Science Indicators Database* has been used as a source of data for:

- Comparative analyses of institutional research performance.
- Time-series studies of an institution's contribution to a specific research specialty.
- Assessments of the productivity of scientists at a corporate lab.
- Tabulations of an organization's most-cited authors.
- Summaries of an institution's publications with citation counts for each item.
- Comparisons of one country's research output and impact with those of other nations.
- Identification of a journal's most-cited papers.

Comparisons of citation impacts over time among several institutions or countries can be graphically plotted. To get a picture of changes in citation impact over time, clients are frequently advised to look at a series of overlapping periods for both cited and citing articles. For example, citation impact can be calculated by dividing the number of citations received from 1981 through 1985 by the number of papers published in that same period. As a second data point, the mean citations per paper for papers published from 1982 through 1986 and cited during the same time frame are computed, and so on from 1987 through 1991. To normalize citation impact, we divide the actual mean by the corresponding mean for a baseline file, constructed to represent the total population from which the individual "samples" were in effect drawn. This provides time-series trends in terms of relative citation impact for a particular institution or country, which can then be compared to other entities (Figure 1).

For faster moving fields, shorter time "windows" might be desirable, while, for other fields, longer time frames might be required.

Another type of impact study analyzes the attributes of papers that cite a given set of publications. For such a set of citing articles, ranked listings of authors, institutions, countries, or journals can be compiled. This kind of analysis helps publishers know where their journals are being read and cited, and enables science managers to see which institutions or countries are most influenced by their institution's work.

Another type of analysis reveals intercitation patterns or citation "flows." Such studies are possible since the *Science Indicators Database* contains all links between *both* the citing and cited items. The result is what is known as an intercitation matrix, with citing attributes listed along the rows and cited attributes along the columns. Interesting examples of intercitation analyses are the reciprocal patterns of citations among nations, and the interaction patterns of citing nations and cited language, from which language-barrier investigations can be pursued. These can be weighted for the size of each entity.

Data sets from this file are available in print, magnetic tapes, cartridge tapes, and floppy diskettes. Our staff regularly advises clients regarding the most appropriate and manageable formats for their data.

The *Hot Article Database*

The second major ISI data resource we use is the *Hot Article Database*. This consists of a set of approximately 1,000 articles that are culled every two months from a cumulative three-year file of the *SCI*. For each of the articles selected, a citation count is maintained and updated every two months. To identify a "hot paper," the system first groups articles by journal and tabulates a citation frequency distribution. Different distributions are constructed for

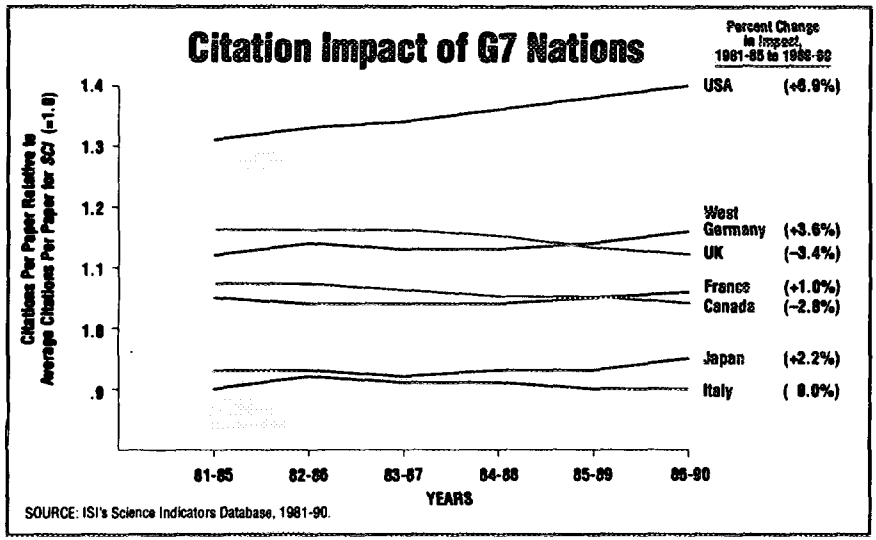


Figure 1. A time-series chart produced by the Research Department at ISI, showing citation impact (citations per paper) of each of the Group of Seven (G7) nations relative to the world average. Papers by US researchers rose 6.9% in relative citation impact, whereas papers by UK researchers declined by 3.4%.

articles "born" in particular bimonthly cohorts—for example, all 10- to 12-month-old articles in *Science*, so that the age of the article as well as the journal in which it appears can be taken into account. A variable citation threshold is then set to select the top 1,000 articles as determined by their position in the journal and bimonthly distributions, using the article's current bimonthly citation count.

Thus, the hot article set includes papers that have attracted more citation "attention" than other papers in the same journal and published at the same time. They become "hot" when a number of current authors cite them soon after their publication. The *Hot Article Database* can be searched in a variety of ways, including by keyword, journal, author name, institution, subject category, etc.

The *Hot Article Database* is available on diskette for installation on a personal computer. The file is updated every two months, and each update includes bibliographic records and citation counts for about 1,000 papers from all fields of science.

The Research Department uses the *Hot Article Database* as a tool to identify breakthrough papers in various fields. The Top 10 listings of hot papers are also a regular feature of our newsletter, *Science Watch*®, which is discussed below.

Research Front Database

The third major database we use is the *Research Front Database*. Research fronts are specialty areas of the sciences and social sciences defined by sets of highly cited "core" papers and corresponding sets of current-year papers that cite the core group. The core papers represent the foundation literature for a research topic on which the set of current citing papers relies. Research fronts are generated annually from a combined *SCI* and *SSCI* database. Annual research front files are available for each year from 1983 through 1991.

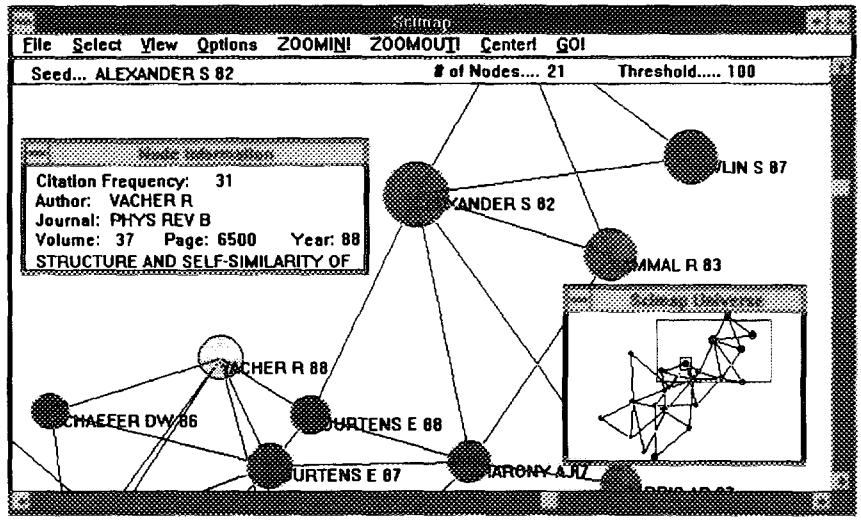


Figure 2. ISI's new SCI-MAP software performs cluster analyses on large sets of bibliographic data. Users can create graphical representations of the relationships among journals, articles, and authors, plus retrieve information for research planning and evaluation, studies on the structure of literatures, or data for library and information center decision-making. Displayed here is a portion of a recent cluster on fractals and self-similarity, showing a number of key papers and their relationship to one another.

Core article sets are generated by a clustering algorithm applied to pairs of highly cited documents that are linked by co-citation. Co-citation is the number of times a pair of papers is jointly cited by current-year articles.[4-6] Citing articles are then tagged by what core they cite.

After generating the citing article sets, statistical indicators are generated for each research front, such as the number and percentage of articles from different nations and organizations. Other statistics are calculated, such as the percent of core papers published within the most recent three-year period. The latter is an immediacy measure that can be used to indicate how new or rapidly changing the area is, whether it is based on current or older findings.

The annual clustering identifies about 8,000 specialized research fronts, containing about 60,000 core papers and representing about 300,000 citing papers.

For each research front, we produce lists of the citing papers (ranked by the number of links to the core papers), authors who most often publish on the subject, and the most frequently occurring title words in the citing papers. We can also produce maps that illustrate the relationships among core cluster articles. Successive cluster maps can be used to investigate the direction research on a topic is taking over time. Higher level maps show how clusters relate to each other or form entire disciplines or fields of science, and how the fields relate to each other.

In addition to working with these three specialized data files to produce custom data sets for clients, our staff also is actively engaged in designing new tools and services for use by clients. An example of these efforts is SCI-MAP, a microcomputer-based clustering and mapping system. (Figure 2).

The SCI-MAP System

This program runs under Microsoft Windows and enables users to cluster large data files on their own personal computers. Based on similar principles to the *Research Front Database*, SCI-MAP allows the user to select and vary the level and strength of the links among the entities to be grouped. Using SCI-MAP, clients can create graphical representations of the relationships among journals, articles, authors, and spe-

cialty areas, and they can retrieve information for research planning and evaluation, library collection development, and for carrying out studies of the structure of literatures in any discipline.

The linking methods used in SCI-MAP are not limited to co-citation. Bibliographic coupling (patterns of shared references) and citation flows also can be implemented.

ISI's Research Department uses SCI-MAP for a variety of tasks, such as performing searches for key articles, exploring the boundaries of research areas, and classifying sets of papers. And, while SCI-MAP is used in-house as a research tool, it is available to clients as a software package combined with either a standard or customized data set. The data sets to use with SCI-MAP can be extracted to suit the client's needs—for example, to focus on a particular subject area, time period, or geographic region.

Science Watch

The Research Department uses results from all of the above resources and tools in producing its newsletter of trends and performance in science, *Science Watch*. Announced in *CC* in January 1990,[7] *Science Watch* is now in its third year. Editor David Pendlebury says the newsletter is written for science policymakers, research administrators at universities and in industry, science journalists, and also for "anyone who requires incisive overviews of key developments in scientific research today." Recent issues have included a series on scientific impact of the G7 nations (Figure 1); citation impact rankings of universities in chemistry, immunology, clinical medicine, electrical engineering, and in other fields; interviews with leading scientists; and even a story on the world's most prolific scientists.

For assessing research efforts, studying a nation's or a university's scientific output and impact, evaluating journals, and identifying structures, trends, and patterns in scientific research, ISI's Research Department can provide data, methods, and consultation on how to design the most effective study. ISI plans to remain at the forefront of "research on research."

REFERENCES

1. **US Congress, Office of Technology Assessment.** *Federally funded research: decisions for a decade.* Washington, DC: US Government Printing Office, 1991. OTA-SET-490.
2. **Holmfield J D.** Broadening the use of quantitative information in science policy. (Meredith M O, Nelson S D & Teich A H, eds.) *Science and technology yearbook—1991.* Washington, DC: American Association for the Advancement of Science, 1991. p. 285-301.
3. **Garfield E.** How to use citation analysis for faculty evaluations, and when is it relevant? Parts 1 & 2. *Current Contents* (44):5-13, 31 October 1983; (45):5-14, 7 November 1983. (Reprinted in: *Essays of an information scientist.* Philadelphia: ISI Press, 1984. Vol. 6. p. 354-72.)
4. **Small H.** Co-citation in the scientific literature: a new measure of the relationship between two documents. *J. Amer. Soc. Inform. Sci.* 24:265-9, 1973. (Reprinted in: *Current Contents* (7):7-10, 13 February 1974 and *Essays of an information scientist.* Philadelphia: ISI Press, 1977. Vol. 2. p. 28-31.) [See also: **Small H.** Cogitations on co-citations. Citation Classic. Commentary on *J. Amer. Soc. Inform. Sci.* 24:265-9, 1973. *Current Contents/Social & Behavioral Sciences* 24(10):10, 9 March 1992, and *Current Contents/ Arts & Humanities* 14(6):20, 16 March 1992.]
5. **Garfield E.** *Citation indexing: its theory and application in science, technology, and humanities.* Philadelphia: ISI Press, 1983. 274 p.
6. **Small H & Garfield E.** The geography of science: disciplinary and national mappings. *J. Inform. Sci.* 11:147-59, 1985. (Reprinted in: *Current Contents* (43):4-14, 27 October 1986 and *Essays of an information scientist: towards scientography.* Philadelphia: ISI Press, 1988. Vol. 9. p. 325-35.)
7. **Garfield E.** Announcing *Science Watch*: a unique newsletter tracking trends and performance in scientific research. *Current Contents* (4):3-6, 22 January 1990. (Reprinted in: *Essays of an information scientist: journalology, KeyWords Plus, and other essays.* Philadelphia: ISI Press, 1991. Vol. 13. p. 25-8.)