From the Science Citation Index to the Journal Impact Factor and Web of Science: interview with Eugene Garfield

12

Eugene Garfield

I had always visualized a time when scholars would become citation conscious and to a large extent they have, for information retrieval, evaluation and measuring impact… I did not imagine that the worldwide scholarly enterprise would grow to its present size or that bibliometrics would become so widespread.

Eugene Garfield

The most remarkable contribution of Eugene Garfield to science is the creation of the *Science Citation Index (SCI)*. In the early 1960s, he founded in Philadelphia the Institute for Scientific Information (ISI) (now part of Thomson Reuters), which became a hotbed for developing new information products (Figure 12.1). Based on the concept of using articles cited in scientific papers to categorize, retrieve, and track scientific and scholarly information, the *SCI* was further developed to create the Web of Science, one of the most widely used databases for scientific literature today. *Journal Citation Reports*, which publishes the impact factors of journals, and *Essential Science Indicators*, which provides information about the most influential individuals, institutions, papers, and publications, are also based on the *SCI*. The concept of "citation indexing," propelled by the *SCI*, triggered the development of new fields such as bibliometrics, informetrics, and scientometrics.

Figure 12.1 The Institute for Scientific Information at 3501 Market Street, Philadelphia, now part of Drexel University.

Eugene Garfield's career is marked also by the development of other innovative information products that include *Index Chemicus*, *Current Chemical Reactions*, and *Current Contents*. The latter included the tables of contents of many scientific journals and had editions covering clinical medicine, chemistry, physics, and other disciplines. Citation indexes for the *Social Sciences Citation Index* (*SSCI*) and the *Arts & Humanities Citation Index* (*A&HCI*), *Index to Scientific & Technical Proceedings and Books (ISTP&B)*, and *Index to Scientific Reviews* were later added to this list.

Garfield started the *SCI* and *Current Contents* (*CC*) in a chicken coop in New Jersey (shown in Figure 12.2). For almost every issue of *CC* he wrote an essay that

Figure 12.2 Dr. Garfield in front of the chicken coop in New Jersey where *Current Contents* and the *Science Citation Index* were created.
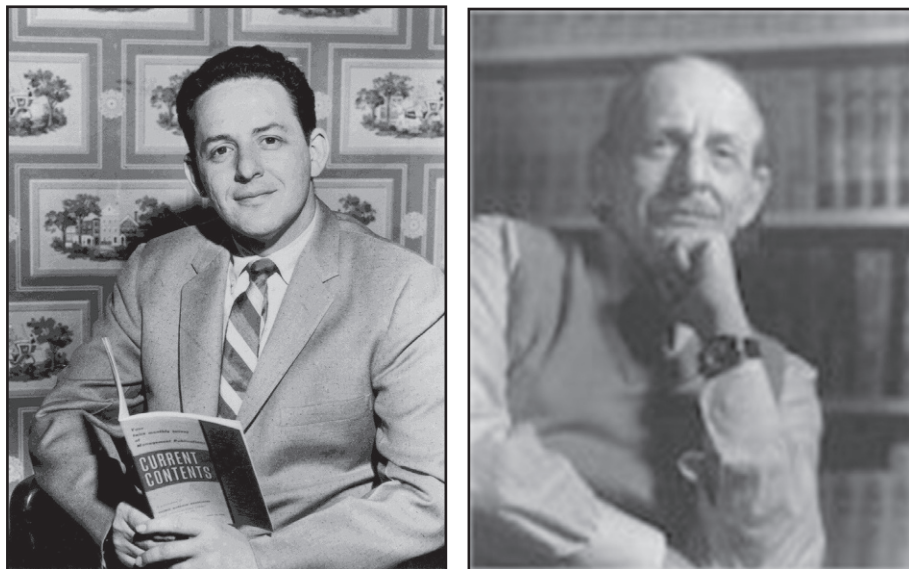
Figure 12.3 Pictures of Eugene Garfield published with his essays in editions of *Current Contents*.

was accompanied by his picture (Figure 12.3). These essays, devoted to a broad range of topics, have fascinated scientists and information specialists from all over the world.

The *SCI* preceded the search engines, which utilized the principle of citation indexing to create algorithms for relevancy of documents. "Citation linking," a concept that is central to the *SCI*, was on Sergey Brin's and Larry Page's minds when they published the paper in which Google was first mentioned (Brin and Page, 1998).

The *SCI* has been used in ways that its creator never envisioned. The Journal Impact Factor (IF), derived from the *SCI* and created for the purpose of evaluating journals, was (incorrectly) extrapolated to measure the quality of research of individual scientists. IF has been used by information scientists, research administrators, and policy makers to see trends in scientific communication and to compare countries, institutions, departments, research teams, and journals by their productivity and impact in various fields. Sociologists and historians of science have been studying processes, phenomena, and developments in research using data from the *SCI*. Librarians often use the IF to select and "weed out" their collections. Editors monitor their journal's impact and citation, and publishers use it to market their publications and decide whether to launch new journals or discontinue existing ones.

It is not an exaggeration to say that the creation of the *SCI* was one of the most significant events in modern science, in general, and in scientific information, in particular. The celebration of the golden anniversary of the *SCI* in May 2014 was also a celebration of the incredible legacy of Eugene Garfield.

Svetla Baykoucheva: At the time when you created the Science Citation Index (SCI) *(Garfield, 1964), you could not have imagined that it would have such ramifications for science. How did you come up with the idea of using citations in articles to retrieve, organize, and manage scientific information and make it discoverable?*

Eugene Garfield: When I was working at the Welch Medical Library at Johns Hopkins University in 1951–1953, I was advised by Chauncey D. Leake to study the role and importance of review articles and decided to study their linguistic structure. I came to the conclusion that each sentence in a review article is an indexing statement. So I was looking for a way to structure them. Then I got a letter from W. C. Adair, a retired vice president of *Shepard's Citations*. He told me about correspondence he had back in the 1920s with some scientists about the idea of creating an equivalent to *Shepard's Citations* for science. *Shepard's Citations* is a legal index to case law. It uses a coding system to classify subsequent court decisions based on earlier court cases. If you know the particular case, you can find out what decisions, based on that case, have been appealed, modified, or overruled.

In *Shepard's*, these commentaries were organized by the cited case, that is, by citation rather than by keywords as in traditional science indexes. I began to correspond with Mr. Adair. I was then a new associate editor of *American Documentation*. So I suggested that he wrote an article for the journal, and he did. Then, I wrote a follow-up article in *Science* in 1955 on what it would mean to have a similar index for science literature.

While *Shepard's* dealt with thousands of court cases, the scientific literature involves millions of papers; so for practical reasons, an index for the sciences was orders of magnitude greater. Since I could not get any government support, I left the indexing project. In 1953, I went back to Columbia University to get my master's degree in library science. I was told by NSF that I could not apply for a grant, because I was not affiliated with any educational or nonprofit institution.

After I wrote the article in *Science*, I got a letter from Joshua Lederberg who won the Nobel Prize in 1958. He said that for lack of a citation index, he could not determine what had happened after my article was published. He then told me that there was an NIH study section of biologists and geneticists. He suggested that I wrote a proposal for an NIH grant. NIH gave ISI a grant of $50,000 a year for three years. However, after the first year, when Congressman Fountain of North Carolina questioned NIH policy, they decided that grants could not be made to for-profit companies. So NIH transferred the funds to the NSF and that's how it became an NSF-funded project.

We eventually published the *Genetic Citation Index*, which is described in the literature. We printed 1000 copies and distributed them to individual geneticists. We produced the 1961 *Science Citation Index* as part of the *GCI* experiment. It covered 613 leading journals. We asked NSF to publish it, but they refused. At that time, we were very successful with *Current Contents*, so ISI used company funds to finance the launch of the 1961 *Science Citation Index* and began regular quarterly publication in 1964.

SB: You have founded the Institute for Scientific Information in Philadelphia, which has become the most important center for scientific information in the world, but many people who now use Web of Science don't even realize that it is based on the SCI. *How did the SCI "underwrite" such resources as the Web of Science,* Journal Citation Reports, *and* Essential Science Indicators?

EG: When we first started the *SCI*, it was a print product. From the very beginning, we used punched cards. So, in a sense, it was machine-readable. Later, we also published the *SCI* on CD-ROM and we released our magnetic tapes to various institutions, including the NSF. The CD-ROM version evolved, and then in 1972, *SCI* went online. The ISI databases *SCI* and *SSCI* were among the first to be on DIALOG. *SSCI* was database #6. That's how ISI databases evolved into the electronic form known today as *Web of Science*. Each year, we increased the source journal coverage. We started with 615 source journals in 1961. I think that more than 10,000 are now covered.

This is how the *SCI* works: whatever you cite in your published papers will be indexed in the database. This is true whether or not the cited journal or book is covered as a source in the database. However, even if an article is covered as a source item, but is never cited, it will not show up in the "Cited Reference" section. You need to differentiate between the source index and the citation index. You use the source index to find out what has been published in the journals that we cover and you use the citation index to see what has been cited. Web of Science includes both the source index and the citation index.

SB: *One by-product of the* SCI *was the Journal Impact Factor, which has become a major system for evaluating scientific impact. You have been quoted saying that the impact factor is an important tool, but it should be used with caution. When should the IF not be used?*

EG: It is not appropriate to compare articles by IF, because IF applies to an entire journal. While the IF of the journal may be high, an individual article published in it may never be cited. You publish a paper, for example, in *Nature*, which has very high IF. The article may never be cited, but if it is published in a high-impact journal, it indicates a high level of quality by being accepted in that journal. I have often said that this is not a proper use of the IF. It is the citation count for an individual article and not the IF of the journal, which matters most. The Journal IF is an average for all articles published in that journal.

SB: *Another unique model that you have created for managing scientific information was the weekly journal* Current Contents *(Garfield, 1979), which included the tables of contents of many scientific journals. Browsing through these tables of contents allowed serendipity. We now perform searches in databases, but we don't browse (through shelves of books, tables of contents of journals). Are we missing a lot of information by just performing searches and not browsing?*

EG: The *Current Contents* has been displaced by the publisher alerts you can now receive gratis from journals. You can receive these alerts by email and browse them to see the articles published in the latest issues. You can also receive citation alerts from Thomson Reuters and others based on your personal profile.

SB: *Your essays in* Current Contents *have educated scientists and librarians on how to manage scientific information. This task has become much more difficult today. How do you manage information for your personal use?*

EG: I don't do research anymore. I follow the Special Interest Group for Metrics (SIG-Metrics) of ASIS&T (the Association for Information Science and Technology). You don't need to be a member of ASIS&T to access this Listserv. I also follow other Listservs such as CHMINF-L (the Chemical Information Sources Discussion List).

SB: Managing research data is becoming a major challenge for researchers and users. How will the organization, retrieval, and management of literature and research datasets be integrated in the future?

EG: If everything becomes open access, so to speak, then we will have access to, more or less, all the literature. The literature then becomes a single database. Then, you will manipulate the data the way you do in Web of Science, for example. All published articles will be available in full text. When you do a search, you will see not only the references for articles, which have cited a particular article, but you will also be able to access the paragraph of the article to see the context for the citation. By the way, that capability already exists in *CiteSeer*—the computer science database located at Penn State University. When you do a search in *CiteSeer*, you can see the paragraph in which the citation occurs. That makes it possible to differentiate articles much more easily and the full text will be available to anybody.

SB: Some of the new methods for evaluating research use article-level metrics. Is the attention to smaller chunks of information going to challenge the existence of the scientific journal? Are there going to be DOIs for every piece of information published—a blog, a comment, a tweet, a graph, a table, a picture? And how will this tendency for granularity in scientific communication affect the ability of users to manage such diverse information?

EG: Back in the 1950s, Saul Herner published an article, "Technical information: too little or too much?" (Herner, 1956). And the same question is being asked today. You can create a personal profile to receive the information that you need. As long as there are tools for refining searches, I don't think there will be too much information for a scientist who will be focusing on a narrower topic. I don't think, really, that anything has changed in that respect. I still browse the same way, I still search the same way, and I want to go to a deeper level.

SB: There is a lot of hype about altmetrics now, but there are currently mixed opinions on its potential as a credible system for evaluating research. Is academia going to be receptive to these new ways of measuring scientific impact?

EG: This is a big educational problem—to educate administrators about the true meaning of citation impact. Scientists are smart enough to realize the distinctions between IF, h-index, and other indexes for citations. There are now hundreds of people doing bibliometric or scientometric research. I think there will be more refined methods, and I'm not worried about that—as long as people continue to get educated. Proper peer review from people who understand the subject matter will be important. I don't think altmetrics is that much different from the existing metrics that we have. Human judgement is needed for evaluations.

SB: A recent article in Angewandte Chemie *was titled "Are we refereeing ourselves to death? The Peer-Review system at its limit" (Diederich, 2013). It discussed the challenges peer review is presenting to journals, authors, and reviewers. Even highly respected journals such as* Science *and* Nature *had to retract articles that had gone through the peer-review process. Is peer review going to be abolished?*

EG: Joshua Lederberg discussed this problem about 30 years ago—that scientists will publish open access and that we will have open peer review. If you are a scientist

and if you are an honest scientist, you don't have to worry about open peer review. People will rely on your reputation. Commentary could come from anywhere and from anyone who wants to contribute to the subject. I don't think peer reviewers will go away. Peer review will evolve to open peer review. I don't believe in the secretive type of peer review. Those scientists who commit fraud will be exposed. If they want to take a chance doing it, that's their problem. But, eventually, they will face worse hurdles for being exposed. And their research won't get cited very much.

SB: People now work in big collaborations, and it is rare that a single person would publish a paper. How is authorship going to evolve and how will individual authors be recognized?

EG: It is a very difficult question, but I think the professional societies have to decide about that. For each paper that is published, somebody will have to take responsibility for that paper. At ISI, we processed all authors equally. For the foreseeable future, all authors will have equal weight. The problem remains for administrators, because they want numbers to evaluate people for tenure.

SB: To follow up on the previous question—researchers who are at a high level, they don't work in the lab. So how could they take responsibility for a paper when someone else has done the actual work?

EG: That depends on the author. It comes down to personal judgement.

(All images accompanying this chapter were provided by Eugene Garfield and reproduced with his permission)

More information about Eugene Garfield is available at www.garfield.library. upenn.edu.

References

Brin, S., Page, L., 1998. The Anatomy of a Large-Scale Hypertextual Web Search Engine. Retrieved September 8, 2014, from http://infolab.stanford.edu/~backrub/google.html.

Diederich, F., 2013. Are we refereeing ourselves to death? The peer-review system at its limit. *Angew. Chem. Int. Ed.* 52 (52), 13828–9. http://dx.doi.org/10.1002/anie.201308804.

Garfield, E., 1964. "Science Citation Index"—new dimension in indexing. *Science* 144 (3619), 649–54. http://garfield.library.upenn.edu/essays/v7p525y1984.pdf.

Garfield, E., 1979. *Current Contents*. Its impact on scientific communication. *Interdiscip. Sci. Rev.* 4 (4), 318–323. http://garfield.library.upenn.edu/essays/v6p616y1983.pdf.

Garfield, E., 2006. The history and meaning of the Journal Impact Factor. *JAMA* 295 (1), 90–93. http://garfield.library.upenn.edu/papers/jamajif2006.pdf.

Herner, S., 1956. Technical information: too much or too little? *Sci. Mon.* 83 (2), 82–6. http://garfield.library.upenn.edu/papers/herner1956.pdf.

This page intentionally left blank