Essays of an Information Scientist, Vol:8, p.320-326, 1985
Current Contents, #34, p.3-9, August 26, 1985

Current Comments®

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

Origins of *Current Contents*, ISI, and Computer-Aided Information Retrieval. How It All Began at the Welch Medical Library Indexing Project

Number 34 August 26, 1985

Perhaps the question I am most often asked is, "How did you get into the information business?" Or, as I once phrased it, "What's a nice boy like you doing in a business like this?"[1] Alternatively, readers often ask, "How did you get the idea for *Current Contents*?" The only reasonable explanation is to provide a capsule account of the Welch Medical Library Indexing Project.

Earlier this year, in a discussion of shorthand systems, I mentioned my experience as a chemical stenographer.[2] In early 1951 my career as a physical chemist came to a halt. My colleagues and I, however, did not see our work published until 1954, which illustrates the delays in publication in those days.[3] Thanks to a series of serendipitous events, I met James W. Perry, one of the pioneers of information retrieval. Jim introduced me to Sanford V. Larkey, director of the William H. Welch Medical Library at Johns Hopkins University, Baltimore, Maryland. I was asked to join the Welch Project, then in its third year of operation. The professional staff consisted of Williamina A. Himwich, a physiologist, Helen G. Field, a librarian, and San Larkey, a physician and historian.

In the pages that follow, I've tried to condense the history and *raison d'être* for this project. This story is directly related to many other reports I've provided on medical information retrieval but especially the *Index Medicus*.[4] MEDLINE and many other systems we take for granted today are direct descendants of this project.

The first contents-page service I ever published was started in 1952 at the Welch Project. It was called *Contents in Advance* and covered library-science and documentation journals. It continued to appear for a while after the demise of the project. The first version of *Current Contents*® (*CC*®) did not appear until 1955.

Today we take for granted electronic online services like MEDLINE or *SCISEARCH*®. These and hundreds of other databases can provide virtually instantaneous access to current research data. But not so long ago, the retrieval of scientific and medical literature was a completely manual operation. Even today, human indexing largely dominates. Thirty-five years ago, however, the use of automation just to organize, sort, and retrieve bibliographic data was more theory than reality. The possibilities of machine indexing had been barely explored. When I first used *Chemical Abstracts* in 1950, its indexes, like most others, were three to five years late. Biomedical and other indexes not only lacked currency, but there was also general chaos in bibliographic control.

In 1948, in response to these problems, an advisory committee to the Army Medical Library was formed. This group eventually recommended awarding a research grant of about $25,000 per year to the Welch Library at Johns Hopkins to establish what came to be known as the Welch Medical Library Indexing Project. An early observer of these events was Scott Adams, deputy director, National Library of Medicine (NLM), Washington, DC.[5] Paraphrasing the mandate of the committee, the project was to investigate "...the size of the

universe of biomedical literature and the extent of its coverage by the existing services, the common and disparate features of subject headings among the services, and the application of machine methods to medical indexing."[6] When I joined the project in 1951, it had made progress in the first two objectives. The last was yet to come.

To appreciate the concerns of the advisory committee and the purpose of the Welch Project, it is necessary to consider briefly the history of medical indexes. As I reported in my earlier account of *Index Medicus*, its origins started with John Shaw Billings in 1879.[4] In his book *The Great Medical Bibliographers*, John F. Fulton, then at Yale University School of Medicine, New Haven, Connecticut, tells how Billings was placed in charge of the small medical library in the office of the Surgeon General of the Army at the end of the Civil War.[7] (p. 70) Since there were few libraries that carried significant collections of medical literature, Billings set out to expand the library's collection. Within eight years, he had increased the holdings of medical volumes and pamphlets by many thousands.[7] (p. 72)

Billings's task was made more difficult by the precipitous rise in the volume of medical publishing in the latter half of the nineteenth century. As Frank B. Rogers,[8] retired director of the NLM points out, there were only 33 American medical periodicals being published in 1847. Thirty years later, Billings counted 262 American periodicals among 864 medical titles published worldwide.[9]

Billings and his staff introduced *Index Medicus* in 1879 as a periodical index to the current medical literature.[9] The *Index Medicus*, according to Estelle Brodman, then of the NLM, listed the contents of many of the books, journals, and pamphlets received by the library.[10] One year later, the same material was reorganized and combined with the library's medical materials from earlier periods to form the *Index-Catalogue*. The *Index-Catalogue* appeared thereafter in one volume per year. The *Index Medicus*, on the other hand, was issued once a month. It listed about 18,000 articles in its first year of publication. Within 50 years, its annual coverage had trebled.[9] Consider, however, that we now list the same number of articles in *CC* in one month.

Other indexes were created to meet the ever-rising volume of medical publications. In 1916 the American Medical Association (AMA) began publication of the *Quarterly Cumulative Index to Current Medical Literature*. In 1927 it was merged with the *Index Medicus* to form the *Quarterly Cumulative Index Medicus* (*QCIM*).[11] For five years, the *QCIM* was published jointly by the AMA and the Library of the Surgeon General's Office. In 1932 the AMA assumed complete responsibility for the *QCIM*, while the library concentrated on the *Index-Catalogue*. The manual preparation of these volumes, particularly of the author and subject indexes, was an arduous task. Brodman quotes Fielding Garrison, who became associate editor of the *Index Medicus* in 1903, on the difficulty of assembling each issue: "...drudgery of the most devitalizing kind," noted Garrison, "ruinous to the eyesight, with consequent impact upon the nervous system, and wearying to the flesh."[10]

During the 1940s, World War II impeded coverage of the growing volume of world medical literature, and the *QCIM*, plagued by increasing costs and printing slowdowns, accumulated a huge backlog of articles to be indexed. It fell two years behind schedule.[10]

To maintain current awareness, more and more researchers turned to the Army Medical Library's *Current List of Medical Literature*, started by Atherton Seidell. As I've noted previously, some might call the *Current List* the predecessor of *CC*.[4] It was a weekly 20-page pamphlet listing the typed tables of contents of newly received periodicals available from the library on microfilm. Like *CC*, it was small enough to be kept in a coat pocket. Although the *Current List* became an official government publica-

tion and underwent expansion and improvement, its coverage was not entirely consistent or complete.

These three indexes—the *Index-Catalogue*, the *QCIM*, and the *Current List*—covered biomedical literature in the late 1940s. Both the AMA and the Army Medical Library sought to improve the currency, journal coverage, and uniformity of subject headings in the three indexes. In 1946-1947, Larkey and others at the Welch Library were considering a training program for medical librarians.[12] It was felt that the program should directly involve research on the problems of medical indexing. The Army Medical Library and the Library of Congress both expressed interest in such a program.

In 1948, at the suggestion of Joseph H. McNinch, director, Army Medical Library, the Surgeon General appointed a Committee of Consultants for the Study of Indexes to Medical Literature Published by the Army Medical Library. The group became known as the Committee on Indexing. Under a contract between the Army Medical Library and the Johns Hopkins Institute for Co-operative Research, Sanford V. Larkey was asked to lead a research project on the problems of medical indexing.[11]

"One of the great difficulties in using the present indexes," wrote Larkey, "is the lack of standard terminology or of a standard approach."[12] He went on to outline the main aims of the project: to study, individually, the various subject-heading lists in medical indexes and to survey the possible uses of machine methods in scientific and medical bibliography. "The use of machine methods may appear somewhat utopian," Larkey continued, "but one must look to the possibilities of the future."[12]

Specifically, according to Adams, the project members set out to address the following questions: What are the indexing requirements of modern medical science? Does the *QCIM* or any of the current indexes meet these requirements? What are the characteristics of a good index with reference to periodicity, format, and printing characteristics? What are the possibilities of using mechanical devices to accelerate indexing?[6] Project members also had to decide whether to continue the *Index-Catalogue*. For a variety of reasons, including an archaic system in which one alphabetic volume had to be published before another could be added, the project recommended in 1950 that the *Index-Catalogue* be discontinued.

As part of its investigation into current medical indexes, the project conducted many interviews with doctors, research specialists, students, and librarians. These interviews, as the staff stated in its 1951 report, confirmed that "...no one is satisfied with the present situation" in regard to medical indexes. "Few if any of the users are aware of how complete or incomplete the coverage of journals is," continued the report, "or of how best to use the maze of subject headings presented to them in the various indexes."[13] The project team hoped to tackle the existing inconsistencies in subject headings, as well as the confusing tangle of "*see*" and "*see also*" references faced by the user of medical indexes. The ultimate aim was to "take the first step toward a standard authority list."[13] According to the American Library Association, an authority list establishes the authoritative forms of headings to be used in a set of bibliographic records as well as the references to be made to and from the headings.[14]

To achieve this goal, the project first set up alphabetic files on three-by-five-inch cards of subject headings for the *Index-Catalogue* and the *QCIM* and for the medical headings from the Library of Congress list of subject headings. These files would later be converted so that they would be arranged by categories. By 1950, a combined file of nearly 30,000 cards had been assembled for subject headings from the *Index-Catalogue* and the *QCIM*.

A key task for the project staff was to find a system for examining and classifying these disparate terms. The staff reviewed such systems as the Universal

Decimal Classification and the Library of Congress classification system to determine whether they might serve as a framework for examining related subject headings.[15] While any *a priori* established system will provide useful historical links between terms, the standard or universal classification systems were largely useless and obsolete.

We found the subject headings fell into certain natural groups. The four main groups that emerged initially were identified as "anatomic terms," "pathologic conditions," "biologic organisms," and "chemical substances." Instead of a universal classification system, a simple system of essentially mutually exclusive "categories" was created. This categorization scheme, as it turned out, was one of the major accomplishments of the Welch Project. Eventually, the original 4 categories were expanded to 15, to cover disease entities and syndromes, public health terms, physical agents and phenomena, and so on.

These categories were used in a study of the subject headings in the cumulative index to the *Bulletin of the Johns Hopkins Hospital*, which the project staff had helped to prepare. This, in turn, led to a study of subject headings in the *QCIM*. The alphabetic card file of the *QCIM* was rearranged in category order. This task identified many more subject headings than in the earlier *Bulletin* study, so additional categories and subcategories became necessary. By heuristic methods the categorization system was gradually expanded.

When I spoke recently with Mina Himwich, now retired, she reminded me that the use of new technology for studying subject headings was driven by the desire to put each term in as many categories as required. Instead of duplicating headings and subheadings, we keypunched as many category codes as needed for each term. It is unfortunate that the major project paper on categorization as a basis of machine coding was never published.[15]

Another of the project's key tasks was to investigate the use of machine methods in medical indexing. As Larkey pointed out in a 1953 report, machines had been used in the production of some printed indexes but not in the medical field up to that time.[16] In 1951 we began a detailed study of subject headings used in the *Current List of Medical Literature*. Instead of three-by-five-inch cards, we used IBM punched cards to record the 4,300 main headings.[15]

These terms were categorized and compared with similar groups created from the study of *QCIM* headings. As we sifted through "new" subject headings, further revisions were made in the categorization system. For example, out of "pathologic conditions" we created subcategories such as "wounds and injuries" and "poisoning." The larger category of chemical terms needed the attention of a chemist. So it was not surprising that we also studied the indexing methods used by *Chemical Abstracts*. I even became one of their volunteer abstractors.

This work did not proceed in a vacuum. There was regular contact with the staff of the *Current List*. Not only were Brad Rogers and Seymour Taine involved in this work, but also Robert Hayne, Thelma Charen, and Stanley Jablonski, among others. During the two years I was with the project, the subject-heading analysis would proceed during the day. Larkey and I spent hours discussing terms and generalities. During the night I would work out the "programs" needed to wire the punched-card machines we used to prepare and tabulate subject-heading lists.[17] These machines were quite limited in the kinds of searching they could perform. Standard sorting machines, for example, could read only 1 of the 80 columns on a regular punched card at a single pass. Once the coded information was punched in, an entire file of cards might have to be sent through the sorter several times to search for various combinations of data. The IBM 101 sorting machine could read all 80 columns simultaneously but was still inadequate for the project's needs. To solve this problem, I

devised a "superimposed" wiring scheme for the IBM 101 that allowed it to sort for as many as 48 seven-character codes in various combinations at one time.[18] The logic of these "programs" was not unlike that used in our *Sci-Mate®* system of text searching.[19]

Eventually we had a relatively sophisticated hierarchical notation for the category system. To describe an article on one punched card, we simply keypunched a five-digit category code, followed by a two-digit subheading code. For example, a card listing a journal article on skin disease would be punched with the five-digit numeric code denoting a main heading, like "skin diseases" or "psoriasis," followed by the code for a subheading like "diagnosis" or "therapy." Other appropriate headings selected to index the same article, such as "antibiotics" or "vitamins," would also be coded on the card.[18] This searching research was separate and distinct from the work on systematizing headings for the day-to-day indexing needs of the *Current List* staff.

To account for all subheadings and all "tracings"—"*see*" and "*see also*" references and the like—several sets of cards were prepared. After several preliminary runs and further revisions in the categorization system, we were able to use machine methods to produce updated subject-heading authority lists at will. In constructing thesauri it is important to know all related terms going to and from a particular term you have chosen.

The authority list that the staff developed, which became known officially as the *Subject Heading Authority List* (*SHAL*), served as a prototype for a later list, called *Medical Subject Headings*, or *MeSH*. *MeSH* is the subject-heading authority list currently used by the NLM in the preparation of the *Index Medicus*.

Machine methods were also used in another of the project's tasks: a survey of the world's medical serials. The staff collected information on more than 6,000 periodicals in the fields of medicine and related sciences. Data on each publication, including language, country, frequency, and major and minor subject fields, were encoded on punched cards. We also determined which journals were "covered" by numerous indexing and abstracting services. After analyzing the collected data, the staff concluded that there was a great deal of wasteful duplication of coverage among the major abstracting and indexing services. At the same time there were major gaps.

The staff's report on world medical serials concluded that better coordination of effort among these services, possibly through the use of machine methods, was called for.[20] But these studies were inconclusive. Ten years later, as an extension of the *Genetics Citation Index*, ISI® performed an article-by-article analysis of indexing and abstracting services. It was found that selection criteria were quite subjective and inconsistent. "Coverage" by an individual abstract journal could never be conclusive. This was one of the guiding factors in designing the *Science Citation Index®* (*SCI®*). I vowed that we would never again cover journals "selectively."

In June 1953, the Welch Project was terminated. From 1953 to 1954, while at the Columbia University School of Library Science on a Grolier Fellowship, I wrote up my project work.[17,18] I also produced my first paper on citation indexes.[21] It is a poignant and nevertheless sad but true fact that I had to publish the results of my research without Larkey's blessing. Since he is not alive to tell his view of the story, it is sufficient to note here that even the closest of collaborators may have very different perceptions of the central events surrounding scientific history.

Later, I was a consultant to the National Library of Medicine Index Mechanization Project.[11] This reflected their desire to apply what we had learned about machine methods to the production of the *Current List*.[22] Using increasingly sophisticated sorting equipment and high-speed cameras to process and photograph punched cards containing typed index entries for the *Index Medi-*

cus, NLM was able to cut drastically the time required to produce the *Current List*.

In assessing the Welch Project, Rogers expresses the view that the Welch Project's work on subject headings and on categorization represents its greatest achievement.[9] One measure of the project's impact in this area is that the current version of the *MeSH* list largely owes its form and organization to the Welch Project. However, equally important to the eventual success of MEDLINE were our primitive efforts at mechanization. Each step in the evolution to digital computers is easily traced back to the project.

During my tenure on this project, I was also involved in a number of other parallel studies. One included an interview project in cooperation with Saul Herner, then at the Applied Physics Laboratory, Johns Hopkins University, Silver Spring, Maryland, on the information-gathering habits of scientists.[23] I learned from those interviews the importance of timeliness in delivering information. In those days of slow indexing and slow publication, attending conferences was the only solution to keeping up. While one should not underestimate the value of personal contacts, meetings should not become a substitute for literature control. I also learned from that experience that organized information is in some sense almost antithetical to the browsing methods scientists prefer in their quest for information. I learned the distinction between information discovery and information recovery. Research administrators too often forget that both types of activity are necessary. I also learned the importance of keeping in touch with users—the research scientists as well as those who serve them.

I have discussed elsewhere that my experience on the Welch Project also laid the foundation for my interest in citation indexing, as well as the importance of reviews, and their relationship to the history of science and medicine.[24] Through this project I met such giant figures as John Mauchly, Chauncey B. Leake, Verner W. Clapp, Ted Herdegen, Peter Luhn, H. Bentley Glass, and others too numerous to mention here.[25]

One of the proudest achievements of the project was the organization of the First Symposium on Machine Methods in Scientific Documentation,[26] attended by 300 persons. The $3.00 fee included project reports and a marvelous buffet lunch. When it was all over, Ralph Shaw[27] wrote me, "As a documentalist you make a great caterer!" I like to think that being an information caterer is not the worst legacy.

I hope this will satisfy those who would like to know where it all began. Over five years ago when I gave the financial story of ISI's origins in *CC*,[28] I omitted this account of the company's intellectual beginnings. If the financial story is what you expected here, please write for a reprint of that essay. It more or less picks up the story where this one ends.

* * * * *

My thanks to Terri Freedman and Christopher King for their help in the preparation of this essay.

REFERENCES

1. **Garfield E.** What's a nice boy like you doing in a business like this? or: what it takes to be an information scientist. *Essays of an information scientist.* Philadelphia: ISI Press, 1977. Vol. 2. p. 32-4.
2. ----------. Is shorthand the route to success in science or anything else? Parts 1&2. *Current Contents* (1):3-12, 7 January 1985 and (2):3-12, 14 January 1985.
3. **Bernhard S A, Garfield E & Hammett L P.** Specific effects in acid catalysis by ion exchange resins. III. Some observations on the effect of polyvalent cations. *J. Amer. Chem. Sci.* 76:991-2, 1954.
4. **Garfield E.** Some reflections on *Index Medicus*. *Essays of an information scientist.* Philadelphia: ISI Press, 1981. Vol. 4. p. 341-7.

5. ----------. Scott Adams and *Medical bibliography in an age of discontinuity*—a tribute to a visionary leader in the field of medical information. *Essays of an information scientist.* Philadelphia: ISI Press, 1984. Vol. 6. p. 197-203.
6. **Adams S.** The way of the innovator: notes towards a prehistory of MEDLARS. *Bull. Med. Libr. Assn.* 60:523-33, 1972.
7. **Fulton J F.** *The great medical bibliographers.* Philadelphia: University of Pennsylvania Press, 1951. 107 p.
8. **Garfield E.** Bringing the National Library of Medicine into the computer age: a tribute to Frank Bradway Rogers. *Essays of an information scientist.* Philadelphia: ISI Press, 1985. Vol. 7. p. 5-11. (Reprinted from: *Current Contents* (2):3-9, 9 January 1984.)
9. **Rogers F B.** The origins of MEDLARS. (Stevenson L G, ed.) *A celebration of medical history.* Baltimore, MD: Johns Hopkins University Press, 1982. p. 77-84.
10. **Brodman E.** The contribution of the Army Medical Library to the bibliographic control of medical literature. *Spec. Libr.* 43:48-54; 73-6, 1952.
11. **Langner M C**, ed. The National Library of Medicine Index Mechanization Project. (Whole issue.) *Bull. Med. Libr. Assn.* 49(1, part 2), 1961. 96 p.
12. **Larkey S V.** The Army Medical Library Research Project at the Welch Medical Library. *Bull. Med. Libr. Assn.* 37:121-4, 1949.
13. **Welch Medical Library, Johns Hopkins University.** *Terminal report of Indexing Research Project.* Unpublished report, 1 November 1948-31 January 1951. (Contract No. W-49-007-MD-467.) 65 p.
14. **Young H**, ed. *The ALA glossary of library and information science.* Chicago: American Library Association, 1983. p. 16.
15. **Field H G, Himwich W A, Garfield E, Whittock I M & Larkey S V.** *Final report on subject headings and on subject indexing.* (Welch Medical Library Indexing Project.) Baltimore, MD: Johns Hopkins University, 1955. 124 p.
16. **Larkey S V.** The Welch Medical Library Indexing Project. *Bull. Med. Libr. Assn.* 41:32-40, 1953.
17. **Garfield E.** The preparation of subject-heading lists by automatic punched-card techniques. *J. Doc.* 10:1-10, 1954.
18. ----------. Preliminary report on the mechanical analysis of information by use of the 101 statistical punched card machine. *Amer. Doc.* 5:7-12, 1954.
19. ----------. Introducing *Sci-Mate*—a menu-driven microcomputer software package for online and offline information retrieval. Parts 1&2. *Essays of an information scientist.* Philadelphia: ISI Press, 1984. Vol. 6. p. 80-7; 96-106.
20. **Himwich W A, Field H G, Garfield E & Larkey S V.** *Survey of world medical serials and coverage by indexing and abstracting services.* (Welch Medical Library Indexing Project.) Baltimore, MD: Johns Hopkins University, 1954. 126 p.
21. **Garfield E.** Citation indexes for science. *Science* 122:108-11, 1955.
22. ----------. The preparation of printed indexes by automatic punched-card techniques. *Amer. Doc.* 6:68-76, 1955.
23. **Herner S.** Information gathering habits of workers in pure and applied science. *Ind. Eng. Chem.* 46:228-36, 1954.
24. **Garfield E.** *Citation indexing: its theory and application in science, technology, and humanities.* Philadelphia: ISI Press, 1983. p. 6-7.
25. ----------. Mapping the multi-disciplinary world of NFAIS! (Miles Conrad Lecture presented to the 27th Annual Conference of the National Federation of Abstracting and Information Services (NFAIS), 3-6 March 1985, Arlington, VA.) *NFAIS Newslett.* 27(2):44-52, 1985.
26. **Hyslop M R.** Documentalists consider machine techniques. *Spec. Libr.* 44:196-8, 1953.
27. **Garfield E.** To remember Ralph Shaw. *Essays of an information scientist.* Philadelphia: ISI Press, 1980. Vol. 3. p. 504-10.
28. ----------. How it all began—with a loan from HFC. *Essays of an information scientist.* Philadelphia: ISI Press, 1981. Vol. 4. p. 359-62.