

# Current Comments

## Discipline-Oriented Citation Indexes & Data Bases—Bridging the Interdisciplinary Gap via Multidisciplinary Input

Number 3

January 19, 1981

This year ISI® will develop a series of disciplinary information services to be commercially available in late 1981 or early 1982. The initial products will cover mathematics, biochemistry, plant sciences, and earth sciences. Each will consist of four basic components. The first printed component will cover 30 years of literature from 1951 to 1980. The most recent five or ten year period will be covered initially. The second component will be a printed annual cumulation. The third component will provide on-line access to the data base, supplemented by a newly developed software system now under testing. For mathematics, the fourth component will be a monthly current awareness publication similar to *Current Contents*® (CC®).

Anyone who has heard me describe the rationale for the *Science Citation Index*® (SCI®) knows how much I stress its value as an interdisciplinary tool.<sup>1</sup> Most fields today are to some extent multi- or interdisciplinary. Pharmacologists, for example, are in touch with synthetic chemistry at one stage of drug development, experimental biology at another, and with clinical medicine at yet another. Even a field as "clearly" defined as organic chemistry today overlaps heavily with physical or analytical chemistry. Obviously, the typical CC reader needs coverage of many fields.

The difficulties that traditional discipline-oriented indexes had in dealing

with the problem of defining fields was one of the reasons for starting the SCI. Among other things, it solved the problem of indexing such multidisciplinary journals as *Nature* and *Science*. SCI has by now become an important tool for researchers, who, among other reasons, want to avoid the coverage uncertainties of discipline-oriented indexes.

For economic and other practical reasons, SCI's "unified index" approach means that it is used almost exclusively in large multidisciplinary libraries. This has limited the use of citation indexes because the typical university scientist relies mainly on departmental or personal libraries for most of his or her information needs. Use of the main library is often inversely proportional to its distance.

How then can we get the SCI into the typical departmental library? We do provide a lower departmental second-copy rate for those institutions which already subscribe. But oftentimes department libraries do not have adequate space for SCI. If this is not the case there are often psychological and political barriers to this approach.

Specialists object that 50 percent or more of the broad-based SCI is not focused on their particular needs. And if the chemistry department is going to have the SCI what about the physics or mathematics departments? On-line computer searching is increasing and is often thought to be a solution, but the printed index is still the preferred information

retrieval medium. So for now, the only practical solution is to develop a series of printed indexes aimed at narrower audiences. Lest this be misunderstood, I am still referring to rather large areas of literature. *SCI*'s coverage is massive—over 500,000 articles yearly—but even in mathematics and computer science today there are about 50,000 articles published per year. The same is true for biochemistry and dozens of other slices of the *SCI*. And there are hundreds of smaller specialties one can identify such as polymer and macromolecular science, analytical chemistry, etc.

How then does one decide what slice to cover in an index to mathematics, biochemistry, earth sciences, or whatever? We faced this problem back in 1958 when we did our first experiments on the *Genetics Citation Index*.<sup>2</sup> It proved impossible to create such an index just by processing classical genetics journals. Even 20 years ago, we had to deal with the then rapidly emerging field of molecular biology. Had we stuck to classical genetics, we could not have produced a reasonable product. Today, the volume of literature in molecular biology is quite large. As it now stands it is more logically treated as a major branch of biochemistry, as it will be in the *Biochemistry Citation Index*. It is possible, however, that the volume of molecular biology literature may one day equal that of "classical" biochemistry. Oddly enough the maturation of molecular biology makes it less difficult today to create a citation index based on the "classical" genetics literature.

When we created the *Genetics Citation Index*, we knew we needed multidisciplinary journal input to cover such a widely dispersed field. This meant that all important journals of science and medicine had to be processed for input. We did just that and then established detailed citation and other selection criteria. We were able to select 95 percent or more of the relevant literature of

genetics. This was verified by a group of experts. By a similar procedure we will compile our disciplinary citation indexes. I must confess that my preoccupation with this problem made it more difficult for me to grapple with the problems of marketing separate indexes. But I will never forget the librarian who told me he would buy the *SCI* if we would only call it the *Chemistry Citation Index*. Even though we are planning separate indexes for biochemistry, etc., it is not impossible that we may have to provide such a comprehensive companion to the standard chemical indexes.

The first selection criterion we will use is the list of core journals for each field. From countless studies, we feel confident that we can identify the core journals of any field. These core journals will publish about 30 percent to 50 percent of the literature of most fields. In some fields, the Bradford distribution<sup>3</sup> is such that 90 percent of the relevant literature is published in less than 100 journals. All core journals for each field will be indexed completely and comprehensively. They can be identified both by citation analysis and expert opinion.

The second selection criterion is also based on citation indexing. We will select any article that cites any core journal. In this way we can select relevant articles from multidisciplinary or non-core journals. In addition to these major selection criteria, we may also select articles on the basis of title words, institutions producing the article, or individual authors.

This, then, is our simple solution to the disciplinary-interdisciplinary dichotomy. Conventional abstracting and indexing services rely on human judgment to make individual article selections. But our choices will be based on well-defined algorithmic procedure. We have in fact been doing this to the satisfaction of users of the *Social Sciences Citation Index*® (*SSCI*™) for

several years. In this way they obtain the benefit of our coverage of science and medicine. Similarly, users of the *Mathematics Citation Index* will find out about articles appearing in the social as well as the natural sciences.

ISI's files are now approaching ten million source articles published over a 20 year period. Printed indexes cover 1961 to date. We are adding to this data base at a rate approaching one million articles per year. For each article, we have created machine records for every paper or book it cites.<sup>4</sup> Thus, in the source index for each discipline we can include complete bibliographic data for each article you have published. We can also include a "citation abstract," that is, the list of all the references or authors cited or quoted. We do this already in the printed and on-line *SSCI*. Space or storage limitations have precluded doing so in the *SCI* and its on-line version, *SCISEARCH*<sup>®</sup>. And for the same reason we can print the new indexes in much larger type.

We can also include cross-references for all authors cited. The publication of so-called "all-author" data will greatly simplify use of citation indexes for a number of bibliometric and sociometric purposes. A feature of these indexes that will simplify the problem of defining the boundaries between fields is a device we already use in the *SSCI* and *SCI*. When an article is cited in both indexes a cross-reference is provided indicating where more references can be found. In the same way we can inform the user of the annual math or biochemistry indexes to check the *SCI* for additional references.

The program outlined above cannot be executed overnight. We shall give certain disciplines higher priority so that we can produce the appropriate cumulated indexes at the earliest possible date. In biochemistry, for example, our files are already so complete that we can fill in data for earlier years with minimum effort. On the other hand, we

will have to work harder to complete our back files in the earth sciences. But since ISI already processes close to 900,000 articles and book chapters per year from current and back-year journals, it would only require about 50 percent additional capacity to process 500,000 additional back-year articles per year. Over the next two decades we can process every issue of every significant journal of science both past and present, which is not yet in our files. It is a massive undertaking, but one I believe will be justified over the years. The value to historians alone would be enough to justify the effort. I'll be saying more about this when we officially announce our ten year cumulation for the *SCI* covering 1955-64.

The first two disciplines to be covered are mathematics and biochemistry. Our coverage of math will be pure and applied including fields like computer science, operations research, and statistics. The introduction of the *Mathematics Citation Index* and its on-line version, *ISI MATHSEARCH*, is particularly relevant to the earlier discussion. Most mathematicians are not aware of *SCI*'s extensive and timely coverage of mathematics. It is not available in most math libraries. Our market surveys also show that pure mathematicians in particular want us to cover at least 25 years of literature. In applied mathematics, there is a keener interest in the more recent literature with an emphasis on the application of mathematics to problem solving. For this the citation index concept is ideal. Whether it is a statistical technique or a computer algorithm, the citation index can retrieve applications of mathematics over the whole range of science and technology. While the *SCI* has been most heavily used in academia to date, the *Mathematics Citation Index* should find wide use in industry and government. It will cover about 40,000 articles per year. The first monthly issue should appear in 1981 and the first five year cumulation covering 1976-80

should be ready in late 1981 or early 1982. The on-line files should be up in 1981.

A similar schedule will be worked out for the *Biochemistry Citation Index*. Since the rate of publication for biochemistry has been studied intensively, we may also decide to begin with a five year cumulation covering 1976-80—about 200,000 articles alone. If we include the applications of biochemistry, the coverage will easily exceed 50,000 per year. Indeed, biochemistry is so pervasive we have to provide precise criteria to preclude it from becoming a life sciences citation index. Clearly we will want to study the possibility and need for indexes in immunology, microbiology, and other identifiable markets in the life sciences.

Once each discipline is defined and a data base selected, it will be put on-line. On-line searching will be available only to those institutions that subscribe to the printed indexes at rates that will encourage on-line use. We are in the business of creating data bases. We are not committed to any particular technology—print, electronic, or otherwise. Each user will decide which is the preferred medium. But all users must pay a

fair share of the cost of creating the data base.

We at ISI have long been committed to developing information services that are at once both comprehensive and selective. With this new approach to creating disciplinary citation indexes and on-line data bases it will be possible for you and your colleagues, whatever your "invisible college," discipline, or mission, to use ISI's unique files and methods to produce data bases for your own field. We are particularly anxious to work with professional societies who want to supplement their existing information programs with selective indexes of the type I've described. If you believe that there are enough potential users for a printed or on-line index to your specialty, please contact us. As this program develops we expect to assign product managers with specialized knowledge to each group of fields. In the meantime you may wish to contact George Malanga, vice president of institutional marketing, or Morton V. Malin, vice president of professional relations, ISI, 3501 Market Street, University City Science Center, Philadelphia, Pennsylvania 19104.

© 1981 ISI

## REFERENCES

1. **Garfield E.** *Citation indexing: its theory and application in science, technology, and humanities.* New York: Wiley, 1979. 274 p.
2. **Garfield E & Sher I H.** *Genetics citation index.* Philadelphia: Institute for Scientific Information, 1963. 864 p.
3. **Garfield E.** Bradford's law and related statistical patterns. *Current Contents* (19):5-12, 12 May 1980.
4. .... Everything you always wanted to know about ISI data bases but were afraid to ask. *Current Contents* (45):5-17, 7 November 1977.\*

\*Reprinted in: **Garfield E.** *Essays of an information scientist.* Philadelphia: ISI Press, 1980. 3 vols.