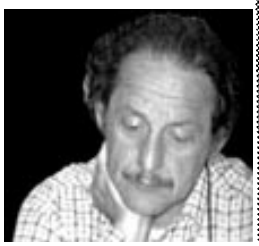
Essays of an Information Scientist, Vol:2, p.550-555, 1974-76 Current Contents, #34, p.5-10, August 23, 1976



"Current Comments"

On the Literature of the Social Sciences and the Usage and Effectiveness of the *Social Sciences Citation Index*[1]

Number 34 August 23, 1976

In the early days I suggested that the *Science Citation Index*® (*SCI*®) might be used as a retrieval system for the social sciences.[2] Some ten years before that I had been interested in the possibility of a citation index to the Bible,[3] and although it may not be generally known, the ancestor of present *Current Contents*® editions—*Current Contents of Social and Management Sciences*—was running in 1955. In short, information services for social scientists were among the earliest of ISI® 's activities. For a period, partly due to the momentum established by the *Genetics Citation Index*, we became more involved in the natural sciences; perhaps also my interest in the social sciences was a little before its time. Latterly an international upsurge in interest, with the realization that the proper study of mankind is man, has been accompanied by a matching level of interest at ISI. This culminated in the introduction of the *Social Sciences Citation Index*® (*SSCI*®) in 1973 coupled to a program of introducing earlier years as quickly as possible.

We have in fact now published editions back to 1970, 1969 is in press and 1967 and 1968 have been authorized. I suspect we will go back at least to 1965 or even 1961 depending upon user reactions.

The current 1975 edition covers over 1400 journals completely and over 2300 journals selectively in the Social Sciences as we broadly define them. Some of the *SSCI* subjects and publication practices--for instance in psychology, one of the dominant areas--closely resemble the *SCI*. The *SSCI* psychology coverage is an extension of the *SCI* coverage. In other areas the situation is quite different. For example over 120 law journals are covered in the *SSCI* but none in the *SCI*.

The journals selected for the *SSCI* were based on lists provided in a number of investigations and directories, by editorial board recommendations, and by citation analysis as described by Weinstock.[4]

Several studies of the social sciences literature have been carried out, notably in the UK. Earle and Vickery[5] found that in the social sciences about four times as many books and one third as many periodicals were cited compared to science. In 1965 the output of books was about the same in the two areas but there were half as many periodicals in the social sciences. Furthermore the largest social sciences journals do not compare in size to such journals as *Science, Nature, Doklady* etc.

The usual controversy about using journal citation counts as a usage indicator comes up in the Earle article but since usage in this case relates to the British Lending Library where well over half the social sciences demand is in the business area, there is little useful information about this subject.

In a later article by Broadus[6] individual social science subject studies were reviewed and as might be expected rather different results were reported from those obtained by Earle and Vickery, which were for the whole area.

The most comprehensive survey so far

attempted must be the one carried out at Bath University and described in a series of reports. The 1975 report[7] in particular is pre-eminent. According to this survey social sciences serial titles were growing at about 3% a year in 1970, and 30-40% were published in English. About 150,000 monographs and books were published. An important conclusion reached in this study is that much published work in the social sciences is of little more than local relevance. This is accompanied by a growing body of methodological, conceptual, and theoretical work of international value.

We may surmise that a fairly small proportion of all serials and monographs. Bath estimates that about 200,000 significant social sciences journal articles are published a year--have an international impact, and that a rather high proportion of these items are likely to be in English. The bulk of social sciences publications are made up of the sum total of locally relevant material, nearly always in the language of the particular country.

A different aspect of the social sciences literature has been discussed by Price,[8] who suggested that the age of references might be used to distinguish between what he calls the soft and hard sciences. For a literature in which the scholar digests all that has gone before and distills words of wisdom about the same questions, "Price's Index" has a low value--that is, many of the references will be to items published more than five years ago. On the other hand for literature which is mainly research front without much archive, a characteristic of much "hard" science, "Price's Index" will be high.

Taking the Bath information and Price's Index together we might expect that the literature of the natural sciences would be characterised by articles mainly citing relatively recent journal items, while social sciences articles would tend to cite relatively old non-journal items.

In fact this trend is born out if the citation characteristics of articles covered by the *SCI* and *SSCI* are analysed. Large enough samples of cited items in the *SCI* and *SSCI* 1975 editions have been taken to indicate the situation for the whole population. The percentage of citations to journal and non-journal items is 85 and 15 for the *SCI* and 47 and 53 for the *SSCI* respectively.

Age statistics are available for the 1974 editions. The percentage of citations to literature less than 5 years old is 48.45 and 44.85 for the *SCI* and *SSCI* respectively, while the percentage of citations to literature more than 15 years old is 18.57 and 21.03.

These differences are not as large as might be expected from Bath and Price. Almost certainly the reason is that the *SCI* extends into the social sciences, while the *SSCI* extends into the natural sciences on the one hand and into the humanities on the other. This is a reflection of ISI's policy, which holds that an indexing policy which puts documents into mutually exclusive boxes labelled *Biochemistry, Behavioural Sciences, Urban Studies, Demography,* etc., is out of step with the cross-disciplinary trends in modern sciences. Information should be retrievable, unimpeded by such arbitrary divisions within the limits of economic feasibility; our definition of science and the social sciences is considerably broader than conventional definitions for this reason.

Turning now to quite a different topic--and perhaps the most important consideration for the database provider--does the service meet the needs of the user?

We have recently talked to two sets of people at all establishments which subscribe to the *SSCI* in the UK. At every place we located the member of the library staff--usually in the periodicals or reference section of a University library-

most knowledgeable about the *SSCI* and its use. We obtained a good impression of typical attitudes and opinions. End users--that is students and faculty members--were identified by library staff and we talked to them as well, although some librarians felt that they should not give us names and did not do so.

We had to remember that it is unreasonable to expect librarians to have the same high level of interest in the *SSCI* that we have ourselves. After all the *SSCI* is only one among many indexes held in a large library. At a few places we talked to known members of the Institute for Information Scientists. Nearly always these people expressed much greater interest and were more aware of the potential of the system.

The general opinion is that the *SSCI* is a useful or very useful index used as much as and in some cases more than other indexes although in many places it is quite new. Several libraries felt that new indexes take quite a long time to become well used, and that use would increase still further. There was an almost universally held opinion about the need for user education, and several librarians were self-ciritical of the amount they were able to impart. ISI's printed material and tape-slide presentation were considered to be well done, but the need for a short simple explanation of how to use the system was often emphasized. We are taking this suggestion seriously.

It is probably unfair to contrast the greater enthusiasm of the other group of people we talked to--the end users--with that of librarians. The librarians were typical, the end-users were not. We were usually referred to end-users by librarians who had remembered people who had used the *SSCI* frequently. Consequently these end-users turned out to be students and faculty members who were unusually information-conscious. In our experience the average person is not information-conscious--he or she needs information but is not disposed to make a substantial effort to obtain it.

Most of the end users with whom we talked had started with a casual look at the *SSCI*, had become interested, and had ended up with a good realisation of the potential of the system. Some were unreservedly enthusiastic. We received comments like (and I quote) "the *Source Index* is an incomparable bibliographic record" and "the *SSCI* is a marvellous resource".

We obtained some interesting examples of completed searches. One user achieved good results in the subject, "Maintenance and reliability of plant and equipment" by using *Permuterm® Subject Index* (*PSI*) word entries. We recommended the *SCI* as a preferred index. A social science undergraduate used the cited author approach to find articles about "wage bargaining and the theory of inflation".

Other cases included a sociologist who used the *PSI* for the subject "inter-racial mixed marriages" a psychology lecturer who used several approaches for investigating "criminal responsibility for the insane" and "psychological aspects of abortion," and a professor of sociology who used highly sophisticated search methods for several subject areas including "innovation and innovative people" and "Black English."

We conclude from the comments received that citation indexes are as effective in the social sciences as they are in the natural sciences.

Other points which came up were the usefulness of the system for locating book reviews--of obvious importance in sociology because of the predominance of books as information sources in this area--and of the value of using the system to identify schools of major workers in a subject.

We found it rather surprising that li-

brarians do not usually attempt to keep any kind of record of the amount of usage of indexes, nor of user reaction. We did not get the impression, even in these days of 'tight budgets, that any conscious attempt was being made to justify a subscription in terms of usage or alternatively to eliminate little-used indexes. We formed the impression that this was because it is difficult to maintain records of this kind, and because a good reference section is considered to be of such great importance that it is unlikely to be pruned even in times of tight budgets.

Following the UK survey we carried out a very similar one in the United States. We took a sample of 22 subscribers and asked the same kind of people at each place the same questions. The consensus of opinion in the US was found to be remarkably similar to that in the UK. Some US periodical or reference librarians resembled UK information scientists--that is, they were professionally interested and aware of the potential of the system. Again we received very enthusiastic comments from both librarians and social scientists whose curiosity had extended beyond simple lookup concepts.

Several additional points of interest came up in the US survey. One user, engaged in research in child cognisance, was concerned about timeliness and referred to the triannual editions as soon as published. This was one reason he used *SSCI* in preference to other indexes--it is more timely. Three users expressed their views about 'noise' in a rather sensible way-- in one case noise turned out to be necessarily high because of the broadness of the user's interest. Each said that he regarded noise as a minor problem which does not usually slow up searches. One user, engaged in a longitudinal study of Adult Socialisation said that he found "30 or 40 hot articles in 2 to 3 hours--much the most rewarding search I have carried out in any index".

At one library it was interesting to find on-line access to *SSCI* via Tymshare/ Lockheed running more or less in competition with *SSCI* manual lookups. The librarian said that her staff tended to operate the terminal on behalf of the user when they felt that the co-ordinating power of the machine would enable the search to be carried out much more quickly. It is time consuming to use a printed index when wanted articles need to be defined in terms of a set of attributes.

Returning to the question of journal coverage, we questioned all persons in UK about this subject. A number of libraries felt that there was a US bias and one thought the *SSIC* was weak on Eastern European journals. End-users however usually felt that the dominance of Anglo-American journals was justified. Another one said he had observed a number of minor US journals, but very few minor non-US journals.

There is, in fact, a greater Anglo-American journal dominance in the *SSCI* than in the *SCI*. In the *SSCI* the figures are 55% US, 19%UK, 26% all other countries, and for the *SCI* 39% US, 17% UK and 44% all other countries.

While we have no doubt that we cover in both systems a high proportion of the world's journals above some arbitrary level of importance, is the proportion of Anglo-American journals covered unjustifiably high?

On the basis of assessment by citation impact coverage is not unjustifiably high either for the *SCI* or *SSCI*. We are confident about our own and several other analyses indicating the dominance of English language journals. The majority originate in the US or the UK. However, there are also an appreciable number originating from other countries. Baughmann[9] shows that 95% of the cited sociological literature is in English, and Fletcher[10] that 79% of economics litera-

ture non-journal citations is to Anglo-American literature.

However the question must be asked to what extent do language barriers and national prejudices limit the extent to which significant material receives international exposure?

Suppose ISI was to increase its *SSCI* journal coverage by 10% by adding suitably identified foreign language journals, what benefit would that provide for *SSCI's* international user? The above-mentioned citation analyses are biassed since the citing sources are mainly English language journals. If a number of foreign language sources were analysed would some heavily cited foreign language journals emerge as candidates for *SSCI* coverage? Would appreciable Anglo-American parochialism be revealed?

We are aware of arguments which support these ideas. For instance in social psychology[11] it has been suggested that non-American social science is permeable to American ideas while American social science is impermeable to non-American ideas.

We conclude that an increase in prime foreign language journals in the *SSCI* in some subject areas ***might*** improve the dissemination of new ideas. However, we have no hard substantiating evidence. We believe that most authors--regardless of their nationality--who have something internationally significant to say will strive ro publish in international journals (usually English language journals) because they believe that by so doing they will receive maximum exposure and recognition. However, if it can be shown to our satisfaction that the inclusion of any foreign language journal not currently covered would result in any appreciable benefit to *SSCI* users, we will certainly consider adding it to our list. In conclusion I will briefly mention another application of citation analysis. I refer to the identification of high quality work by research communities, the interrelationships between research specialties, and studies of the structure of science through the analysis of its literature. I suggested that this might be a rewarding topic for investigation some fifteen years ago.[12]

Work has proceeded apace in the natural sciences[13,14] and Henry Small at ISI recently used a similar approach in the social sciences. We have devised a method for automatically identifying sets of articles which tend to be frequently cited together. These heavily co-cited articles have been identified as the core literature of a scientific specialty--as perceived by the citing authors currently working in the specialty. Natural-science specialties include *reverse transcription, cyclic AMP, sickle-cell disease, plate tectonics,* and *X-ray sources.*

The same automatic procedure has been applied to the social sciences, and when comparing the results with the natural sciences some interesting differences are seen. Firstly, the most heavily cited items in the social sciences are nearly all books, while in the natural sciences they are nearly all articles. Secondly the most heavily cited items are mainly in psychology. Fifteen out of the 26 most heavily cited items are in this area, although the most heavily cited item of all is B. J. Winer's "Statistical principles in experimental design".

When performing the automatic procedures, which we call clustering, to identify specialties and to show how specialties relate to each other, the progressive raising of certain selection parameters gradually eliminates all but the areas of most intense activity. At the lower levels, activity is centered in a number of large groupings including finance and monetary theory, management, politics, law, sociology, psychology, leadership, sexual behaviour, psychoanalysis, personality, and drug abuse. As thresholds are raised

certain areas such as economics, law and political science disappear, leaving areas of high publication concentration in subjects such as memory and learning, and social and experimental psychology, with psychology dominating.

There are two major differences between social-sciences and the natural-science clusters. First, subjects are more tightly connected in the social sciences--the whole area hangs together and disciplinary boundaries blur more than in the natural sciences. Second, the mean age of the specialties--that is of the cited core documents is greater in the social sciences--6.5 years compared to 3.5 years in the natural sciences.

We are about to produce specialty information for a number of successive years in the social sciences, as we have already done for the natural sciences. This will show how specialties are growing and fading, and how relationships between different areas are changing.

☆ ☆ ☆ ☆ ☆

Shortly after this paper was presented at St. Andrews, a list of *SSCI* clusters, along with a cluster-map of high-activity research areas in the social sciences appeared as one of these editorials.[15]

REFERENCES

1. **Garfield E.** On the literature of the social sciences and the usage and effectiveness of the *Social Sciences Citation Index*. Paper presented at the Biennial Conference of the Institute for Information Scientists, 18-20 June 1976, St. Andrews, Scotland. Reprinted in *Current Contents®* (*CC®*) No. 34, 23 August 1976, p. 5-10.
2. **Garfield E.** Citation indexing; a natural science literature retrieval system for the social sciences. *Amer. Behav. Sci.* 7:58-61, 1964.
3. ----------. Citation index to the Old Testament. Paper presented at the Annual Meeting of the American Documentation Institute, Philadelphia, Pennsylvania, November 2-4, 1955.
4. **Weinstock M.** ISI's *Social Sciences Citation Index;* a new comprehensive multidisciplinary information retrieval system for social sciences literature. Paper presented at the Annual Meeting of the American Society for Information Science, Washington, D.C., October 23-26, 1972.
5. **Earle P & Vickery B.** Social science literature use in the UK as indicated by citations. *J. Documentation* 25:123-41, 1969.
6. **Broadus R N.** The literature of the social sciences; a survey of citation studies. *Internat. Social Sci. J.* 23:236-43, 1971.
7. **Bath University.** Size, growth, and composition of social science literature. Design of information systems in the social sciences (DISISS). Bath University Research Report, Series A, Number 2, March 1975.
8. **Price D J D.** Citation measure of hard science, soft science, technology and non-science. In: *Communication among scientists and engineers,* ed. by C.E. Nelson & D.K. Pollock (Lexington, Mass: D.C. Heath, 1970), p. 3-22.
9. **Baughmann J C.** A structural analysis of the literature of sociology. *Library Quarterly* 44:293-309, 1974.
10. **Fletcher J.** A view of the literature of economics. *J. Documentation* 28:283-95, 1972.
11. **Finison L J & Whittemore C L.** Linguistic isolation of American social psychology. *Amer. Psychologist* 30:513-16, 1975.
12. **Garfield E.** Citation indexes in sociological and historical research. *Amer. Documentation* 14:289-91, 1963.
13. **Small H & Griffith B C.** The structure of scientific literatures. I. Identifying and graphing specialties. *Science Studies* 4:17-40, 1974.
14. **Griffith B C, Small H G, Stonehill J A & Dey S.** The structure of scientific literatures. II. Toward a macro- and microstructure for science. *Science Studies* 4:339-65, 1974.
15. **Garfield E.** *Social Sciences Citation Index* clusters. *CC* No. 27, 5 July 1976, p. 5-11.