

Symposium on Cooperative Efforts in Information Processing. Introduction*

BERNARD S. SCHLESSINGER

Southern Connecticut State College, New Haven, Conn. 06515

Received January 18, 1973

The symposium papers which follow are an outgrowth of a previous paper¹ by the author in which the duplication of efforts in the area of physicochemical property data handling was pointed out, and the need for cooperative efforts was stressed.

The need for cooperative efforts is even more strongly indicated by some recent work by the author and Steven Hildrich (unpublished work) in which a random sample of 145 abstracts of journal articles in *Chemical Abstracts* was checked for duplicate processing in several other abstracting and indexing services. Preliminary results indicate that the intellectual effort of processing these articles is repeated for better than half the cases by another abstracting and indexing service. Extrapolating these preliminary results to the total volume of publication in *Chemical Abstracts*, and using conservative figures for costs and time spent in quality abstracting alone leads to a total expenditure in duplicate costs of roughly two million dollars annually, and in duplicate time of 75 man-years annually.

Considering the present climate of economic belt-tightening and the lack of good information personnel, such duplication of effort is unforgivable. The papers that follow were chosen to be representative of the present laudable efforts to eliminate some of the processing duplication in various areas of chemistry by cooperation, or to detail how cooperative effort has led to a more useable product. These examples of cooperation in the areas of abstracting and indexing (the first paper by C. M. Flanagan), information storage and retrieval systems (the next three papers by S. J. Martinez, S. A. Rossmassler, and P. Urbach, respectively), and structural data handling (the last two papers by C. E. Granito and W. J. Wiswesser *et al.*) are encouraging developments in working toward the ultimate goal of the most efficient processing of information to make maximum use of money and talent available to the information community.

LITERATURE CITED

*Presented before the Division of Chemical Literature, 164th Meeting, ACS, New York, N. Y., Aug. 27, 1972.

- (1) Schlessinger, Bernard S., "Physicochemical Property Data Handling," *J. Chem. Doc.* **9**, 20-4 (1969).

Coordination—A Detailed Review of the Relationships Among the Publications and Services of BIOSIS, CAS, and Ei*

CAROLYN M. FLANAGAN

Engineering Index, Inc., 345 East 47th Street, New York, N. Y. 10017

Received January 18, 1973

In April 1970, the BioSciences Information Service of Biological Abstracts (BIOSIS), the Chemical Abstracts Service (CAS), and Engineering Index, Inc. (Ei) announced a joint five-part study with the following objectives: to determine the amount of overlap among the lists of primary journals each of the services regularly monitors, to find out how the same journal articles are selected for analysis, to study differences in the editorial policies and procedures employed by the three services, to measure the degree of compatibility or interconvertibility of the computer-readable files of each of the services, and to determine the similarities and differences of indexing policies. This paper reports the findings of the first two parts of this study.

In 1963, the Report of the President's Science Advisory Committee, stated that "Since the entire information system is a network of separate subsystems, rapid and effi-

cient switching between the different elements of the system is essential. Such switching will be fully effective only if the different subsystems adopt uniform practices toward abstracting and indexing."¹

Again in 1969, the Satcom Report noted that "The increasing volume of material to be covered has caused the

* Presented before the Division of Chemical Literature, 164th Meeting, ACS, New York, N. Y., Aug. 27, 1972.

cost of such services (abstracting and indexing) to rise rapidly. At the same time, increasing subscription costs have discouraged individuals from subscribing to them Special care is necessary to ensure the broad usefulness of the product, and maximum responsiveness to the progress of science and technology. Two paramount issues in this context are sensitivity of management, particularly in regard to scope of coverage and adequacy of abstracting and indexing, and availability of abstracts for reprocessing."²

The history of the abstracting and indexing services has been documented for us in bibliographies, monographs, and brochures of individual services. In attempting to evaluate the mass of informative material available to us in this area, we find that most of the bibliographies have been created by librarians for librarians as a training or research tool and that from the early 18th century to date there have been some 2000 abstracting services to contend with. Some of these have followed the life cycle from wild enthusiasm through disillusionment to oblivion. Many still remain and are further evaluated in monographs sometimes created because of a need to publish. It is the individual service brochure that is most informative and revealing in that we begin to see the changes that have and are taking place to serve our user audiences.

It is the individual service that has progressed through the evolutionary stages of management, editorial, production, and economic change, that today, is creating the data base required for a highly specialized research and development community. The individual service thinks it knows its audience and what that audience wants. The service tailors its products and resulting prices to a market that, in its complexity, changes more rapidly than the capability of the service to meet the market demands. The literature is replete with criticism, suggestion, and experience concerning the needs of the information community and as individual services we have listened and tried to meet these needs in our own sphere of influence. In recent years, we have discovered that no service is unique in its problems.

In April of 1970 an agreement was drawn up and duly signed by the responsible management directors of the Biosciences Information Service of Biological Abstracts, the Chemical Abstracts Service, and Engineering Index, Inc., to conduct a review of the relationships between and overlap in printed and computer-readable publications and services.

The study is designed to provide information needed by the three services for planning cooperative programs, as well as to reconcile differences in policies and practices to make their publications and services more useful to their user communities.

This is not just another study. It is designed to supply the basic data required by the three services to coordinate information in transdisciplinary fields. It is the result of a directive from managements which recognize the need for closer cooperation; the same managements that have brought about innovative changes in new and existing products and have used the developments in the computer field to mechanize their services to a point where the present study can extract information from the separate data bases in a compatible mode.

The study will be carried out in five phases with the following objectives:

1. To determine the amount of overlap among the lists of primary journals each of the services regularly monitors
2. To find out how the same journal articles are selected for analysis by more than one of the three services
3. To study differences in the editorial policies and procedures employed by the three services as they relate to

the form, format, and content of bibliographic descriptions and abstracts

4. To measure the degree of compatibility or interconvertibility as well as the reliability of the computer-readable files of each of the three services

5. To determine the similarities and differences in indexing policy by comparing access points or index entries selected by the three services for the same or related concepts.

Phases 1 and 2 have been completed, and the results to date are or soon will be available.³

Phase 1 compared the listings of the primary journals each service had monitored for coverage as of May 1, 1970. This list included only primary journals. Patents, books, conference proceedings, technical reports, theses, and standards are not included.

Computer-produced lists were supplied by the three services. BIOSIS—7460, CAS—9172, and Ei—2181 journal entries. Each list had been prepared using similar data elements including CODEN which was used as the identifier to create a master file.

Analysis of the file showed that 14,592 different journals were being covered collectively by BIOSIS, CAS, and Ei. Of these 10,511 or 72% were covered by one service, 3,941 or 27% by two of the services, and 140 or 1% by all three services.

In sorting the cards for the journals monitored by only one of the three services, it was determined that the BIOSIS list contained 4317 entries for journals that were not monitored by either CAS or Ei. The CAS list contained 5122 entries monitored solely by CAS, and the Ei list provided 1072 entries for journals that neither BIOSIS nor CAS monitored.

To determine two-way overlap, the 3941 cards with two source codes were sorted. This analysis revealed that BIOSIS and CAS were both monitoring 2972 of the same journals, none of which Ei monitored and an additional 140 journals each that Ei did monitor. The analysis identified 31 journals that were being monitored by both BIOSIS and Ei, but not by CAS. Thus, the overlap between BIOSIS and Ei was comprised of these 31 journals plus the 140 journals monitored by all three services. CAS and Ei were monitoring 938 journals that were not listed by BIOSIS, bringing the CAS-Ei overlap to 1078 journals when the 140 three-way overlap journals were added.

In reviewing the two-way overlap of monitored journals, the combined BIOSIS and CAS lists were found to contain entries for 13,060 different currently published scientific and technical journals. Of these, 3112 or 23.8% were being monitored by both BIOSIS and CAS. Together CAS and Ei monitor 10,815 different journals, of which 1078 or 10% are monitored by both. Only 1.8% or 171 of the 9470 different journals covered by BIOSIS and Ei were being monitored by both processors.

The 140 journals that were monitored by all three services constituted less than 1% of the 14,592 journals monitored by the three services when the study was begun in May 1970.

The second phase of the study determined overlap at the level of individual journal articles. Each service examined its selection records to determine the number of articles it had selected for coverage from the issues bearing publication dates of July 1969 through June 1970, inclusive. Once the number of articles selected for coverage by each service from the overlap journals for the one-year period was determined, it was possible to compare one service's selection against one or both of the others to determine the overlap at the article level.

It is estimated that the total number of journal articles abstracted and/or indexed by the three services during the 1969-70 study period was 301,000.

Statistical analysis indicated that only 822 of the same articles were covered by BIOSIS, CAS, and Ei, and it was determined that this did not constitute a large enough overlap to warrant additional effort. In the two-way article overlap, the overlap between BIOSIS and CAS is approximately 43,000 and between CAS and Ei approximately 17,500 journal articles. The methodology used to determine the overlap at the journal article level and detailed tabular data have been published in Part 1;³ Part 2 will be published soon.⁴

Cost figures for abstracting and indexing an individual article vary so widely that no true figure can be quoted at this time, but it is obvious that we are considering an overlap of some 50,000 journal articles which is substantial enough to warrant further bilateral investigation.

The three services are continuing the study to obtain the information they need for future cooperation to make their publications and services more useful to the information community.

Each of the three services has agreed to arrange to finance its contribution to the study. If it becomes neces-

sary, the duration of the study will be extended so as not to place an undue financial burden on any of the participants.

LITERATURE CITED

- (1) "Science, Government, and Information," report of the President's Science Advisory Committee, 52 pp., Jan. 10, 1963.
- (2) National Academy of Sciences, "Scientific and Technical Communication," report of the Committee on Scientific and Technical Communication, Washington, D. C., National Academy of Sciences, 1969.
- (3) Wood, J. L., Flanagan, Carolyn, and Kennedy, H. E., "Overlap in the Lists of Journals Monitored by BIOSIS, CAS, and Ei," *J. Amer. Soc. Inform. Sci.* **23**, 36-8 (1972).
- (4) Wood, J. L., Flanagan, Carolyn, and Kennedy, H. E., "Overlap among the Journal Articles Selected for Coverage by BIOSIS, CAS, and Ei," *J. Amer. Soc. Inform. Sci.*, to be published.

A Cooperative Information Storage and Retrieval System for the Petroleum Industry*

SAMUEL J. MARTINEZ

Information Services Department, The University of Tulsa, Tulsa, Okla. 74104

Received January 18, 1973

The *Petroleum Abstracts* bulletin was first published in January 1961, as a cooperative information storage and retrieval service for the exploration, development, and production segment of the petroleum industry. Information retrieval from retrospective files was first accomplished by an inverted subject card file. In 1965, this system was replaced by controlled-vocabulary, subject indexing. Currently under development is the PASS direct-access computer system (Petroleum Abstracts Search System) which allows real-time intercommunication with the disk-pack stored data base, through a telephone-linked portable teletype console. This system is designed to operate on a Xerox Sigma 6 computer.

The *Petroleum Abstracts* bulletin and associated Information Storage and Retrieval Services were initiated in January 1961 by a group of petroleum producing companies. Prior to this time, many of these companies had been monitoring the technical literature and abstracting articles and patents pertinent to their specific interests. Such operations were conducted by internal information services departments, usually in close association with the companies' technical libraries. Recognizing that this practice represented considerable duplication of effort, a group of these companies delegated The University of Tulsa to administer a cooperative abstracting service for the benefit of the exploration and production segment of the petroleum industry. The cost of conducting this service was divided among the various participating companies, each individual contribution being proportional to the company's capitalization. The operation of the abstracting service was guided by a Subscriber Advisory Board, consist-

ing of representatives from each of the participants. The purpose of this group was to provide user feedback and to help establish policies and operational guidelines.

As a result of this cooperative venture, each company was able to obtain much greater coverage at much lower cost than had proved possible with their previous individual abstracting efforts. A similar abstracting and retrieval service, administered by the American Petroleum Institute, provides parallel coverage for the refining and petrochemical segments of the petroleum industry.

The cooperative information services provided by The University of Tulsa have greatly expanded and diversified since 1961.¹ The *Petroleum Abstracts* bulletin itself has expanded from 10,488 abstracts in 1961 to 16,924 abstracts in 1971. Much of this increase may be attributed to expanding technological developments and diversified new fields of interest—such as offshore exploitation methods, environmental pollution, etc. More sophisticated retrieval techniques also have been developed to provide faster, more comprehensive searching of the rapidly growing data base.

*Presented before the Division of Chemical Literature, 164th Meeting, ACS, New York, N. Y., Aug. 27, 1972.