(4) Maynard, J. T., *J. Chem. Doc.* **10**, 285–9 (1970).
(5) Meyer, E., *Angew. Chem.* **82**, 605–11 (1970).
(6) Rössler, S., and Kolb, A., *J. Chem. Doc.* **10**, 128–34 (1970).
(7) Schier, O., Nübling, W., *et al.*, *Angew. Chem.* **82**, 622–8 (1970).
(8) Skolnik, H., and Hays, *J. T., J. Chem. Doc.* **10**, 243–7 (1970).
(9) O'Dette, R. E., Klingsberg, E., Bloemeke, M. J., Garfield, E., Wigington, R. L., *J. Chem. Doc.* **9**, 183–90 (1969).
(10) Shtejn, V. S., *Zh. Vses. Khim. Obshchest. D. I. M.* **16**(1), 68–73 (1971).
(11) "Dokumentation eingestellt," *Nachr. Dok.* **22**, 206 (1971).
(12) "Zweckentfremdet," *Nachr. Dok.* **22**, 219 (1971).
(13) Sabiniewicz, S., *Przem. Chem.* **49**, 461–5 (1970).

Multipackaging of Data at Source

CHARLES L. BERNIER
State University of New York at Buffalo, Buffalo, N.Y. 14214

Received February 4, 1972

Prompt handbooks and computer stores of data can be provided through cooperative multipackaging of data by authors and publication through existing organizations. Data scattered throughout primary and secondary literatures are costly to find, extract, evaluate, organize, report, and share. Techniques, media, and systems for making organized data available exist, but have not been used extensively enough. Reasons for failure to use these are explored and recommendations made for extraction by authors at time of writing, centralized organization, and publication through existing organizations. Datum signatures are described. Centralized extraction is shown to be unfeasible.

Authors embed data in documents. Extraction later is inconvenient, mainly because of scattering. Comprehensive, concise compilations of current, processed data (hereafter called simply "data") are not available despite awareness of need and despite what seem to be practical techniques, media, and systems for dissemination of data (Cf. all references). This paper explores reasons for failure to use these techniques, media, and systems for the benefit of all specialists (such as engineers, professional people, and scientists), and presents possible solutions to the problems involved in this failure.

For at least a century, the scientific community has been well aware that useful data are scattered throughout its primary and secondary literatures. Retrieval of data requires undue amounts of time of scientists to: search indexes, find documents (including computer tapes), copy data, evaluate it, organize it, report it, and sometimes share it by publication.[1,2]

Specialists waste valuable hours searching the wildernesses of their primary and secondary literatures for data. In view of the existence of handbooks and of computer stores of data, they do not understand why prompt, comprehensive handbooks and computer stores of data have not been provided to them.[3,4] Their time, including overhead, may average $200 a day. Handbooks of current data have not kept pace with outpourings of technical presses.[2] Specialists are well aware of the convenience of handbooks, despite need to use indexes to them, as contrasted with the inconvenience and cost of searching the primary literature for data, for example. Access to handy handbooks is a matter of seconds; access to handy computer stores is a matter of minutes; access to data of unknown location in the primary literature is a matter of hours, days, weeks, or even months. The cost of providing data conveniently and promptly is small compared with the cost of developing, calculating, recording, checking, and publishing the data in the first place. The cost is negligible compared with the combined value of such data to our civilization. Such value can be estimated from Gross National Products, longevity, morbidity, mortality, national incomes, etc. Making data more convenient seems to lie in the fields of economics and of psychology. We experience what can be termed, "fractured economics"—dollars (e.g., 21–24 billions per year) made readily available for research, while cents (e.g., hundreds of millions of dollars) are provided for convenient communication of the results. In the United States, we seem heavily to favor getting information from experience (e.g., experiment) over getting it from the record (e.g., the library).

PRIMARY PUBLICATIONS

Primary publications are not convenient sources of data because of:

Word Dilution. The primary paper or report is written to enable repetition of the work described without consulting the author. It is ideal for this purpose. Papers and reports are unsuitable for skimming—the rapid, convenient location of data.[5] Data are diluted by too many words. Words are seductive. Place a word in front of a scientist, and he will read it. He will not know whether

or not it is relevant to his interests until he has read it. By then, his time may have been wasted.

Scatter. Data are scattered not only within, but also among papers and reports. Time taken to locate, extract, evaluate, organize, categorize, index, and disseminate data has largely fallen upon the shoulders of users of the technical literature. These users, being specialists in other subject areas, often have neither the training nor the patience to carry out these operations, except for evaluation. As a result, these operations are often undertaken with reluctance and even with resentment. Difficulties are experienced in search. Searches may be incomplete. Searches may not be undertaken or may be abandoned. Time of very valuable people is wasted.[6]

Delays. The promptness with which stock-market quotations appear can be a goal for other sources of data. Financial people do not tolerate delays accepted routinely in other fields. Striving for promptness does not mean advocacy of rushing into print and repenting at leisure, but only advocacy of prompt dissemination of data once it is deemed publishable. The concept that speed in the publication process fosters errors is fallacious. Stock-market quotations in the newspapers are fantastically accurate. Accuracy of published data is controlled by routinizing, training, and checking; and it has but little to do with speed.

Supposed Lack of Evaluation. Skepticism, a virtue for scientists, engineers, medical men, and other specialists, causes them to read papers to re-evaluate the data in them. Such reading and re-evaluation take time. Careful evaluation already provided by authors, and the additional (sometimes cursory) evaluation by editors and reviewers or referees may be minimized. A specialist may accept reasonable data from a handbook without question, but may balk at accepting data from a primary publication although it may have received the more careful evaluation. We do not have adequate measures for, or records of, evaluation that are published as a part of data. Data considered here are largely labeled numbers; the labels should carry not only the names of the numbers, but also confidence limits and records of the evaluations.

Language Problem. The seventy or so languages in which an increasing number of papers are now published[7] present formidible barriers to efficient, prompt extraction and use of data by people who do not read these languages.

Access Problem. Technical literature, published in hundreds of countries and in tens of thousands of periodicals, is not promptly available to all who need it. Indexes of quality are late, incomplete, or nonexistent. Periodicals may be at binderies. Borrowers may hold the material. Issues may be small and soon depleted. Interlibrary loan is sometimes slow. Secrecy is a major barrier to access. Even the largest libraries do not have all of this material.

ABSTRACTS

Published abstracts are not convenient sources of current data because of:

Omissions. There are little or no data in indicative or descriptive abstracts. Even informative abstracts omit data—often in such a way that the omissions are invisible to the user. Measures of validity and confidence limits of data are too frequently omitted. The labels may be incomplete. Records of evaluation are nearly always missing.

Delays in Publication. Delays of months to a year are not uncommon.

Word Dilution. Even abstracts, which may average 10% of the length of the originals, take too long to search for data.

Validity. During abstracting, errors creep in; incorrect data are published. Few professional workers are so reckless as to rely upon abstracts for data that they use in their research and publications. Originals are consulted with waste of time and effort.

Language Problems. Although abstracts may be available in the language of the user, the desire to validate data may take him into the primary literature and into languages that he does not read.

Access Problems. The same problem exists here as for the primary literature owing to the desire to validate. Delays in indexes of high quality to abstract journals may delay reliable access to these journals by months or years.

Scatter. Similar data are scattered among abstracts through the years and decades because abstracts on the same subjects cannot be published side-by-side. The related, original documents, from which abstracts are derived, are published at unpredictable times and in unpredictable places.

HANDBOOKS

Techniques for preparing handbooks of data exist. Updating procedures can include use of decimal-numbered pages to promote interfiling. Loose-leaf data books are not new. Techniques for acquiring documents and for extracting, organizing, indexing, and publishing data are well-known as are techniques for computer storage, accumulation, processing, organization (e.g., in listing), retrieval, and updating of data. The specific causes of failure to use these techniques comprehensively now seem to be: lack of organized extraction of data at source and re-evaluation and reformatting of data prior to dissemination.

RE-EVALUATION

Recommendations to omit re-evaluation at the start of a program for extraction, organization, and publication of data—so as to get on with the task of making data more convenient—have been published. The arguments to take the loads of search and extraction off the shoulders of users are persuasive.[2] Eventual re-evaluation—by other than the ultimate user—may be found desirable, but can wait. Re-evaluation is largely a professional task that may be repeated by many users despite all previous evaluations and re-evaluations.[1] Evaluation by authors, colleagues, editors, referees, and reviewers before publication may be superior to that of compilers of handbooks whose product may often be used without challenge. Search, extraction, and organization are services that can aid re-evaluation by facilitating comparisons. The efforts involved in these three services may be foreign to re-evaluation, take the time of the re-evaluator and distract him from his complicated, highly-specialized task. So, if we temporarily omit re-evaluation by authorities—in realization of the urgent need to provide organized, comprehensive data promptly and conveniently to specialists—we will have removed one reason for failure to provide convenient sources of prompt data (manual or computerized). We are assured that in omitting re-evaluation by authorities, that we are not eliminating all evaluation. We know that the author's technical reputation is at stake in

everything he publishes; so he is eager to maintain and improve validity of data he presents. Editors, reviewers, referees, journal publishers, and professional societies are all jealous of their professional reputations and so promote dissemination of only valid data. All of this helps to protect the user (as well as author) of data. The professional user should know that data, re-evaluated or not, can be inappropriate, incorrect, or incomplete. He should be constantly aware of this in his experimental work and in his use of the primary literature, handbooks, and other secondary sources of data. There is nothing new or startling about all of this.

EXTRACTION

The other principal block to provision of convenient sources of prompt data is having it extracted. Authors may be viewed as reluctant to extract their own data so that colleagues and peers can find it more conveniently and promptly. Concern for authors, may I suggest, is misplaced; it is users about whom we should be concerned. It is obviously to the advantage of authors to have users find their data conveniently and find it helpful. Authors receive credit for what good they do—in the minds of colleagues, peers, subordinates, and superiors.[9] Authors can be requested without hesitation to extract data by rules, just as they are now asked to abstract, categorize, and write their papers according to rules. Extraction of data can be a prerequisite to publication, to acceptance of a paper for presentation at a meeting, and to acceptance of an official, terminal report. There is nothing new or startling about all of this. Specialist readers of their literatures can, in addition to authors, be requested to extract data according to simple rules from papers they have chosen to read. This is to complete what authors have left undone until editors come to require all authors to extract.

Incentives and planning for extraction of data have been proposed[6,8] and explained as unnecesary.[5] Experience in publishing *Chemical Abstracts* and *Biological Abstracts*, as well as other abstracting journals, shows that specialists who are well qualified to abstract are readily available gratis or for honoraria. Incentives for part-time abstracting and extracting by subject authorities are: working cooperatively on a project that benefits all, keeping currently informed in a field in an especially effective way, keeping up in other languages that might become rusty, prestige in working for a respected organization, recognition by colleagues, investment of spare time in a worthwhile enterprise, travel, and pin money.[6]

CENTRALIZED EXTRACTION

Decentralized extraction of all highly processed data is necessary because of the arithmetical and logistical absurdities of comprehensive, centralized extraction. The professional literatures of articles may now approximate two million items a year. These are written in about 70 languages and appear in at least sixty thousand periodicals (The Biological Sciences Communication Project has identified 55,000 journals in the life sciences. *Chemical Abstracts* covers 10,000 journals with some overlap. *Engineering Index* covers 3500 journals with some overlap.) Each item is written in two languages—the native or acquired tongue of the writer and in the jargon of the specialist's field. For selection and accurate extraction, the extractor must know both languages. A specialist cannot keep up in a field broader than one he can read and digest. If the average specialist digests an average of about five documents each working day, then he can digest about one thousand per year. This means that all specialized knowledge must be divided into roughly an average of 2000 subfields. Seventy languages times 2000 subfields means 140,000 subfields in all languages. This means about 140,000 specialists for optimum extraction. If each specialist can handle an average of two languages, this means 70,000 specialists needed for extraction. And if each specialist can extract from two other subfields adjacent to his as well as from his own (a total of three subfields) without experiencing serious difficulties and making too many mistakes, then about 23,000 specialists will have to be assembled. Keeping this number in a central location, paid, happy, and each up in his own subfield presents a very difficult, if not impossible logistical problem. Assuming an average salary plus overhead of $20,000, the annual budget for salary comes to almost five billion dollars. Extraction will not keep these professional people busy all of the time. A specialist will average fewer than 100 documents a year. At 20 documents extracted per day, extraction will keep each specialist busy an average of fewer than five days of the year. When secrecy (personal, proprietary, military, and political) is also considered, the problem of centralized, comprehensive extracting becomes even more difficult to solve. Decentralized extraction by groups of specialists is a more attractive possibility. Even more attractive is completely decentralized (individualized) extraction by authors, colleagues, peers, referees, reviewers, editors, and specialist users at times of writing or reading a document. Rules for extracting data are in use for card files, handbooks, and computer stores of data.

DATUM SIGNATURES

A datum signature is a complete datum expressed linearly in a preferred order. A number, by itself, is not a datum. For example, "100," is not a datum. At least, "100 of what," must be expressed—e.g., "Boiling point of normal water at 760 mm. Hg is 100°C." This sentence is a datum signature. It is complete except for supplying data on confidence limits, etc., for citing the source of the signature, and for the name of the extractor.

Preferred order enables datum signatures to be arranged in lists so that users of lists can be directed by simple rules as to how to find a datum signature. Indexes to datum signatures can be prepared. Preferred order within the signature is arbitrary, as is order in the Universal Decimal Classification or in the Colon Classification. Datum signatures can be entered into a computer store which is programmed to rearrange their components to bring any component to the fore, and to print a list in order of that component. For example, in the above example on boiling point of water, the signature can be rearranged to have, in addition to property (boiling point), as starting component: substance (water), units of measurement (° C.), conditions of measurement (760 mm. Hg), confidence limits, references, etc. Each list serves a different purpose: A list of properties creates, e.g., a boiling-point table. A list of substances gives a table of all properties of the same substance (as in the famous handbooks of chemistry). The lists can also be prepared without the aid of a computer, as has been done for decades. Components of signatures that are of less importance in a given list at a given time and for a given purpose can be omitted. Other components can be abbreviated or coded to save space. Lists of abbreviations, codes, and defini-

tions are then provided. For example, the citation can be coded as a serial number. Such coding is commonly used in tables of data.

Data are often tabulated or expressed as graphs, charts, or diagrams for greater usefulness and for more rapid intelligibility. The table, graph, chart, or diagram is labeled with some or all of the other components of the datum signatures. Computerized tabulation, graphing, charting, and diagramming is, of course, standard practice and time-saving. A master list of all properties and other labels for data are unnecessary because authors supply labels that are commonly used and understood. Authors also supply definitions for new labels. "See" cross references, such as are commonly used in subject indexes, can be used to prevent scattering of identical data among synonyms. "See also" cross references can be used to guide users among closely related properties. These standard indexing practices have been used successfully by *Chemical Abstracts* for more than half a century.

MULTIPACKAGING OF DATA AT SOURCE

The process of supplying organized collections of highly processed data much more promptly and comprehensively than at present includes:

Securing cooperation of editors, authors, users, reviewers, scientific or professional societies, and others in extraction of data and in organizing and administering its extraction

Devising simple, self-learned rules for extraction in each subject field

Keeping track of extracted documents to avoid duplication

Persuading organizations (e.g., industrial, scientific, professional, and governmental) to accept responsibility for publication or for dissemination within security

Centralized organization, categorization, and publication (or internal dissemination)

Selling published data with subsidy at the start to reduce the economic barrier and to improve opportunity to sample service

Selling without subsidy to continue service

Systematic feedback (e.g., by questionnaire or interview) to determine: value of the publications (multipackages) in eyes of users, effect on organizations, and—in the long run—effect upon society and civilization

Sharing evaluation and organization by specialists through publication or limited dissemination

The cost and price of providing adequate services that supply organized data conveniently should not prove to be prohibitive in view of successful similar commercial services for $9000 or more per year. Publication by professional societies can be used to encourage cooperation by specialists for honoraria or gratis. Cost of specialists searching the present literature is probably far greater than the cost of successful data publications.

SHARING EVALUATIONS AND ORGANIZATION

Most evaluations and organizations of data are unshared. Selection or rejection of data is a form of evaluation. That which is selected is valued more highly than that rejected. Use of data from the literature in calculations, experiments, and papers nearly always represents approval by the user. Data selected may be organized, categorized, compared, re-evaluated, recalculated, and converted into graphic form. Usually this effort is unshared, except *via* citations to documents supplying the data. Evaluation and organization can be shared by sending evaluated and organized data to a central place for categorization and dissemination.[9] If this sharing were extensive, then current handbooks of evaluated, organized data (e.g., that on drugs) could be less costly and much more prompt than at present.[10] Such sharing can be encouraged by professional and scientific societies, governmental agencies, handbook publishers, and computerized data systems. Sharing can always be within security. Sharing within an organization can preserve security and save repeated evaluation, categorization, graphing, etc. Within an organization, cooperative sharing of evaluation and organization of data seems more easily achieved than for published data. However, in view of the extensive and valuable cooperation given to vast enterprises, such as *Chemical Abstracts*, by abstractors and section editors for honoraria or gratis, the relatively minor effort involved in sharing data is feasible. The re-evaluated or reformatted data shared can be a by-product of necessary reading and indicated for extraction by colored marks on pages read or on Xerox copies of the data extracted. References supplied can be to the author of the paper read as well as to re-evaluator and reformatter. Citation of data read, generally indicates approval and represents a judgement in favor of the data for the purpose used. Thus, number of citations of data can be used as a statistical measure of approval. Number of citations can be included in the datum signature. Correlation of number of citations with validity of the data should be studied later for possible use of the number as a rough measure of validity.

DISSEMINATION

Rules for dissemination (sharing) of data include:

1. Observe security, if any.
2. Work cooperatively with as many other organizations as possible, security permitting.
3. Categorize centrally for uniformity and efficiency.
4. Use standard classifications.
5. Use an existing organization for dissemination, where possible. It is to the advantage of professional and scientific societies, industrial organizations, and governmental agencies to provide promptly to their professional members data in order to save time and prevent obsolescence of these very expensive people. Charters of some scientific societies include dissemination of information as an objective.
6. Publish loose-leaf handbooks to provide the bulk of the service.
7. Store data later than the last updating in a computer.
8. Update the handbooks with colored sheets for interfiling and for eventual replacement with white sheets.
9. Make the handbooks of the scope that is a balance between being most economical and being most convenient to use and update.
10. Provide computer tape for those organizations wishing to make their own searches. Disseminate tape periodically.
11. Subsidize the service until it catches on—then price it at cost including overhead for professional and scientific societies; for commercial ventures, price for survival.
12. Give samples at the start to test the product and to accustom users.

After successful services are operating, investigate re-evaluation of data gratis by authors of special handbooks, re-evaluation with honoraria (e.g., copies of the product) by authorities, re-evaluation by computer-aided systems, and re-evaluation by peers or panels of authorities. Then decide whether or not to initiate re-evaluation programs utilizing the data already organized and disseminated.

CONCLUSIONS

1. Data should be extracted by specialists who know the language and jargon to ensure authoritativeness.
2. These specialists probably cannot all be assembled and still be kept productive and happy; hence, extraction must be decentralized, or partly so.
3. Centralized organization, categorization, indexing, and publication are probably necessary for uniformity and efficiency. Centralization may best be by subject area, such as chemistry.
4. Cooperation of authors, editors, reviewers, referees, colleagues, peers, and users can probably be secured for extracting data just as has such cooperation been secured for over a century for abstracting.
5. Re-evaluation may turn out to be unnecessary and should be postponed at least until comprehensive, prompt compilations of disseminated data are operational.
6. Such compilations will aid in re-evaluation through comparison—perhaps computer-aided—and by taking the load of searching, acquiring, reading, organizing, categorizing, and compiling off the shoulders of the re-evaluators.
7. Output can be in the form of looseleaf handbooks of data updated by colored pages that are eventually replaced by white pages. Data more current than the last handbook updating can be carried in computer stores.
8. It is in the best interests of professional, scientific, and technical societies, and industrial and governmental organizations to promote prompt dissemination of organized data to their members and employees to save expense and to prevent obsolescence.

It is time that we take steps to make processed data much more convenient. Starting small is possible.

LITERATURE CITED

(1) Brady, E. L., and Wallenstein, M. B., "The National Standard Reference Data System [NSRDS]," *Science* **156,** 754–62 (1967).
(2) Schlessinger, B. S., "Physicochemical Property Data Treatment," *J Chem. Doc.* **9,** 20–4 (1969).
(3) Speight, F. Y., "Numerical Data Activities of Engineering Societies," *Ibid.,* **7,** 26–30 (1967).
(4) Cohan, L., "Directory of Computerized Information [and Data] in Science and Technology," Science Associates/International, Inc., New York, 1968.
(5) Way, K., "Free Enterprise in Data Compilation," *Science,* **159,** 280–2 (1968).
(6) Overhage, C. F. J., "Science Libraries: Prospects and Problems," *Science* **155,** 802–6 (1967).
(7) *CAS Today,* American Chemical Society, Washington, D.C., 1967.
(8) Brown, H., "International Cooperation: The New ICSU Program on Critical Data," *Science* **156,** 751–4 (1967).
(9) Rossmassler, S. A., "Modification of Dissemination Channels for Scientific Information," *J. Chem. Doc.* **9,** 17–19 (1969).
(10) de Haen, P., "Modernizing the Structure of Reports on Drugs," *Clin. Pharmacol. Ther.* **9,** 547–9 (1968).

FURTHER READING

(1) Allen, S. I., Barnett, G. O., and Castleman, P. A., "Use of a Time-Shared General-Purpose File-Handling System in Hospital Research," *Proc. IEEE* **54,** 1641–8 (1966).
(2) "Industry Attacks the Chemical Data Problem," *Chem. Eng. News* **47**(16), 22–3 (1969).
(3) "Materials Data Retrieval at Your Finger Tips," *Metalworking Prod.* **110**(10), 72–3 (1966).
(4) Becker H., Breitenlohner, H., Lang, C., and Schwarz, F., "Computer Application in Pathology; Methods and Experience Following Evaluation of 27,000 Autopsy Protocols" (in German), *Method. Inform. Med.* **8**(2), 60–7 (1969).
(5) Blagodarnyi, L. A., "Centralized Information on Non-standard Equipment. A Remedy for Diversity," (in Russian), *Nauch. Tekh. Inform. Ser. 1,* **12,** 7 (1968).
(6) Buchanan, O. H., Ginsberg, H. F., Traub, P. G., Marshall, I. H., and Allen, A. E., Jr., "Screen and Clinic—Three Year's Operational Experience," in "Symposium on Drug Information," Division of Chemical Literature, 148th Meeting, ACS, Chicago, Ill., August 1964.
(7) Burger, J. B., and Wilson, W. J., "A Review of Selected Methods of Machine Manipulation of Chemical Structures," CFSTI, AD-451 100.
(8) Chemical Abstracts Service, "Progress Toward a Chemical Registry System at Chemical Abstracts Service," CFSTI, PB-184 811.
(9) Citro, C. F., "The Census Bureau as an Information System. Developments in Increasing Access to Census Data," *Spec. Libr.* **60**(1), 10–16 (1969).
(10) Craig, P. N., "Correlation of Chemical Structures with Biological Data," *J. Chem. Doc.* **2,** 106–9 (1961).
(11) Deutsch, W., "Knowledge in the Growth of Civilization: a Cybernetic Approach in the History of Human Thought," in "The Foundations of Access to Knowledge," Syracuse University, New York, 1968.
(12) Eckenbach, H., "The Many Faces of SDA. The Capturing of Data in Various Forms at Writing—Information Put Directly on System-Compatible Media by the People Who Originate It," *Bus. Automat.* **12**(12), 33–6, 41 (1965).
(13) Evans, W. H., and Garvin, D., "The Evaluator Versus the Chemical Literature," *J. Chem. Doc.* **10,** 147–50 (1970).
(14) Fava, J. A., "A Framework for Future Data Centers," in *Proc. Amer. Soc. Inform. Sci.,* 6, 32nd Annual Meeting, Oct. 1969.
(15) Feldman, A., "User Controlled File Organization," in *Proc. Amer. Soc. Inform. Sci.* 6, 32nd Annual Meeting, Oct. 1969.
(16) Gluck, D. J., "A Chemical Structure Storage and Search System Developed at Du Pont," *J. Chem. Doc.* **5,** 43–51 (1965).
(17) Gore, N. B., and Way, K., "The Data Compilation as Part of the Information Cycle," *Ibid.,* **2,** 179–81 (1962).
(18) Goudsmit, S. A., "Is the Literature Worth Retrieving?," *Physics Today* **19**(9), 52–5 (1966).
(19) Griesser, G., "Statistics of Symptoms and Signs," *MIMC-A,* **4**(2), 79–82 (1965).
(20) Haenszel, W., and Lourie, W. I., Jr., "Quality Control of Data in a Large-Scale Cancer Register Program," *Method. Inform. Med.* **5**(2), 67–74 (1966).
(21) Henderson, M. M., "Evaluation of Information Systems: a Selected Bibliography with Informative Abstracts," The National Bureau of Standards, **NBS-TN-297,** 1967.
(22) Horton, C., Minkler, T. M., and Cady, L. D., Jr., "MEDATA—A New Concept in Medical Records Management," in, *Amer. Fed. Inform. Process. Soc.* Fall, 1967.
(23) Jager, H. D. M., Maxwell, D. C., and Ridley, R. G., "Information Retrieval and Future Developments at the Mass Spectrometry Data Centre," *Inform. Stor. Retrieval* **4,** 133–7 (1968).

(24) Johnson, V. J., "Development and Operation of a Specialized Technical Information and Data Center," *J. Chem. Doc.* **8,** 219-24 (1968).
(25) Kiely, J. M., Juergens, J. L., Hisey, R. L., and Williams, P. E., "A Computer-Based Medical Record: Entry of Data from the History and Physical Examination by the Physician," *J. Amer. Med. Ass.*, **205**(8), 571-6 (1968).
(26) Kilgour, F. G., "Publication of Scientific Discovery: A Paradox," in *Amer. Doc. Inst. Proc.*, Annual Meeting, October 1966.
(27) Kunz, W., and Rittle, H., "On the Logic of Research and Documentation. Some Strategies in Designing Amicable Information Systems for Scientific Research," *Naturwissenschaften* **55,** 358-61 (1968).
(28) Legatt, T., Grandy, R. P., and Lorenzo, S. X. D., "A Biologically Oriented Data Retrieval System," *J. Chem. Doc.* **9,** 177-80 (1969).
(29) Levitt, D. G., "Drug Monitoring System," *Ibid.*, **6,** 34-7 (1966).
(30) Lewis, J. R., "Drug Information Handling by the American Medical Association," *Ibid.*, **5,** 258-65 (1965).
(31) Lindberg, D. A. B., "Symposium on Information Science. VII. Electronic Reporting, Processing, and Retrieval of Clinical Laboratory Data," *Bacteriol. Rev.* **29,** 554-9 (1965).
(32) Lindelof, K., "Ringdoc: Trade Documentation for the Pharmaceutical Industry," *Tidskr. Dok.* **21**(2), 17-20 (1965).
(33) Lowry, W. K., "Science Information Problems Needing Solution," *Amer. Doc.* **19,** 352-4 (1968).
(34) Oatfield, H., "The ARCS System: Ringdoc as Used with a Computer," *J. Chem. Doc.* **7,** 37-43 (1967).
(35) Oatfield, H., "Information Centers, Clearinghouses, and Referral Centers Which Offer Chemical Data," *J. Chem. Doc.* **5,** 131-4 (1965).
(36) Pergam, A., "Computerized Monitoring of Clinical Investigations," Rept. of ARIES Corp. McLean, Va., 1968.
(37) Rice, C. N., "Toward a National Systems Resource in Toxicology," *J. Chem. Doc.* **9,** 181-3 (1969).
(38) Rosenberg, L., "Information Retrieval as a Problem in Packaging," in *Amer. Doc. Inst. Proc.*, Annual Meeting, October 1966.
(39) Rossini, F. D., "Historical Background of Data Compiling Activities," *J. Chem. Doc.* **7,** 2-6 (1967).
(40) Röttger, P., Reul, H., Klein, L., and Sunkel, H., "The Automatic Handling and Statistical Evaluation of Pathologic-Anatomical Findings," *Method. Infor. Med.* 8(1), 19-26 (1969).
(41) Scheublein, M., Steidle, W., Germann, W., Nübling, W., and Wilhelmi, R., "Machine Documentation of Literature in the Pharmaceutical Industry," *Naturwissenschaften* **55**(8), 362-7 (1968).
(42) Skolnik, H., "A Correlative Notation System for NMR Data," *J. Chem. Doc.* **10,** 216-20 (1970).
(43) Soni, A. H., and Harrisberger, L., "Wanted: a Mechanism Information Research Center," *Mech. Eng.* **91,** 730-4 (1969).
(44) Sussenguth, E. H., Jr., "A Graph-Theoretic Algorithm for Matching Chemical Structures," *J. Chem. Doc.* **5,** 43-51 (1965).
(45) U. S. Department of Labor, Bureau of Labor Statistics, "The BLS Information System. Background and Principles," 25 pp., U.S.D.L., Washington, D.C., 1967.
(46) Verhulst, H. L., "Information to and from Poison Control Centers," *J. Chem. Doc.* **9,** 71-3 (1969).
(47) Wadington, G., "CODATA—Its Organization, Activities, and Goals," *Ibid.*, **9,** 174-7 (1969).
(48) Waldo, W. H., "Searching Two-Dimensional Structures by Computer," *J. Chem. Doc.* **2,** 1-6 (1961).
(49) Weber, R. W. "Associative Processing of Fragmentary Information," *IEEE Trans. Eng. Writing and Speech* 8(2), 71-80 (1965).
(50) Weisman, H. M., "Needs of American Chemical Society Members for Property Data," *J. Chem. Doc.* **7,** 9-14 (1967).

Journal Coverage by the Major Chemical Title and Abstract Publications

EBERHARD KIEHLMANN
Department of Chemistry
Simon Fraser University
Burnaby 2, British Columbia, Canada

Received May 8, 1972

The journal coverage provided by *Chemical Titles, Current Contents, Science Citation Index, Chemischer Informationsdienst,* and *Index Chemicus* is discussed and compared with the CASSI list of the thousand primary journals most frequently cited by *Chemical Abstracts.* On the basis of this comparison, frequency of journal citation in the *Journal of the American Chemical Society,* and frequency of journal citation in the major chemical title and abstract publications, two lists of 186 and 64 core journals of chemistry are presented in alphabetical and rank order, respectively.

The world's chemical literature continues to grow at an annual rate of 8 to 9% as measured by the number of abstracts of scientific papers from journals and other serial publications published by *Chemical Abstracts,* and the task for the active research chemist to keep up-to-date on recent developments in his chosen specialty becomes more and more difficult. Although it has been shown[1] that some 250 "core journals," out of a total number of 8500 abstracted in 1970, produce approximately 30% of all papers of chemical interest (excluding patent literature), even this relatively "small" number of primary information sources is too large to permit regular weekly scanning. Statistical studies[2] have revealed that on the average a chemist manages to check regularly the contents of