19 June 1981, Volume 212, Number 4501 SCIENCE

Critical Data for Critical Needs

David R. Lide, Jr.

Summary. Various types of scientific and technical data are required for the solution of key societal problems such as energy supply, environmental quality, and industrial productivity. Ensuring the quality of these data bases is essential. Modern computer and telecommunications technology offers opportunities for major improvements in the dissemination of data, but data management must be given a higher priority by the scientific community, industry, and government.

The "information explosion" in science and technology has been a popular theme for writers and speakers over the past 25 years. The proliferation of scientific literature and data files has taxed our ability to store, retrieve, and assimilate information. At the same time, society has become increasingly dependent on science and technology, and institutions are faced with more and more decisions that have a strong technological component. Some of these decisions are being made on the basis of grossly inadequate data; others are delayed by wrangles over conflicting or inconsistent data. At times the sheer mass of data tends to discourage a logical approach to decision-making. The effective matching of information resources to societal needs is a problem that must be overcome.

The growing power and availability of digital computers has prevented a complete swamping of our ability to handle scientific information. Impressive progress was made in the 1970's in developing computer-based bibliographic files that can be searched interactively from remote terminals (*1*). These files permit rapid searching of many thousands of documents on the basis of a profile of key words—a carefully selected set of terms which attempts to match the user's interest with the contents of the documents. The result of such a search is a list of potentially relevant documents. These on-line bibliographic search systems, although not fully perfected, have proved their utility in making a first pass through the information morass.

This article addresses a second stage of refinement in coupling information sources to needs, namely the provision of reliable factual information directly to the user. Can a user be given an actual number or other factual answer to his question, within the constraints imposed by a computer-based dissemination system? This is clearly a greater challenge, since it requires the exercise of some degree of quality control and a careful analysis of the way in which the user will apply the data. Quality assurance has not been a major issue in on-line bibliographic services; the vendors of such services simply rely on the traditional review processes of the primary journals and abstracting services. Likewise, once a user has identified a document through an on-line search, the burden is on him to extract the data and put it into the format required for his application. When attempting to provide data directly to the user, however, both quality assurance and format design become key factors.

The effect of the electronic revolution on the generation of scientific and technical data is one of the driving forces toward the use of these electronic advances for data storage and dissemination. Automation of the measurement process has led to vast increases in the amount of data generated in every scientific discipline. From nuclear physics to molecular biology, the instrumentation advances of the past two decades have resulted in the production of far more data than can possibly be handled through traditional mechanisms of the scientific literature. There is no choice but to utilize computers and modern telecommunication systems to preserve these data and make them accessible to present and future users.

Classes of Data

It is somewhat ironic that the scientific and technical information community has failed to develop a precise vocabulary for many key concepts. The term data base is frequently used for bibliographic as well as numerical data files, although certain groups have attempted to differentiate the latter by calling them data banks. In this article, data will be understood to mean information, usually expressed in numerical form, that has been derived from some measurement, observation, or calculation. A data base is an organized collection of such information on a well-defined topic. We shall be concerned particularly with data bases expressed in some computer-readable medium.

A detailed scheme for classifying scientific and technical data has been presented by the Committee on Data for Science and Technology (CODATA) of the International Council of Scientific Unions through its task group on accessibility and dissemination of data (*2*). The details of the scheme need not concern us here, but it is helpful to recognize three broad classes of data:

1) Class A—repeatable measurements on well-defined systems. This includes traditional physical and chemical data resulting from measurements of well-understood properties of systems of known composition. In principle, such

David R. Lide, Jr., is chief, Office of Standard Reference Data, National Bureau of Standards, Washington, D.C. 20234

0036-8075/81/0619-1343$01.75/0

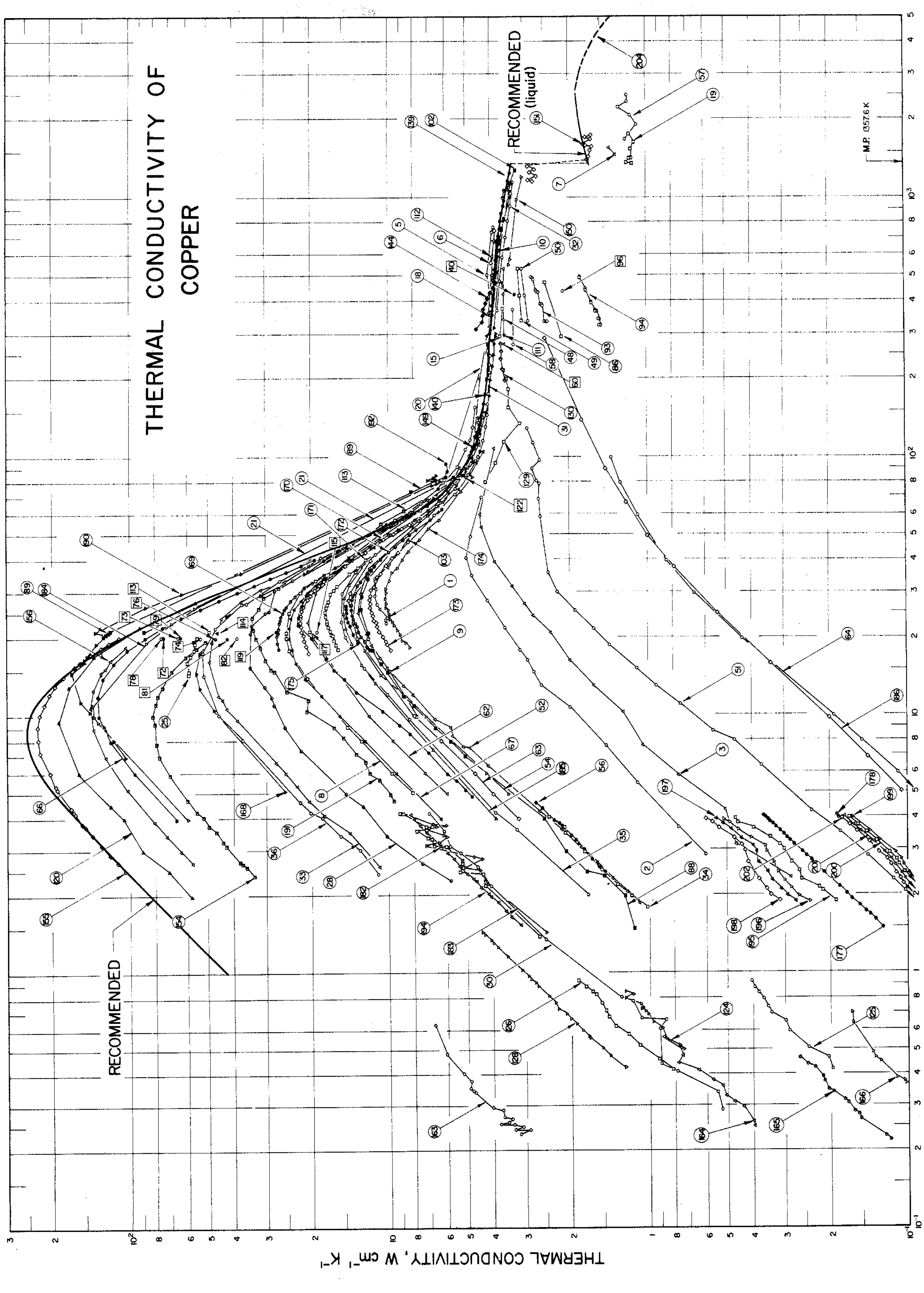
THERMAL CONDUCTIVITY OF COPPER
RECOMMENDED
RECOMMENDED (liquid)
M.P. 1357.6 K
TEMPERATURE, K
THERMAL CONDUCTIVITY, W cm⁻¹ K⁻¹

Fig. 1 (facing page). Display of all reported measurements of the thermal conductivity of copper as a function of temperature. Each set of connected points corresponds to the data given in a single literature source. The recommended curve is based on the evaluation described in (7).

data are subject to verification by repeating the measurements in different laboratories at different times.

2) Class B—observational data. Here we include the results of measurements that are dependent on time or space and cannot, in general, be checked by remeasurement. This category includes much of the data from the geosciences and environmental monitoring data.

3) Class C—statistical data. This class includes data not always thought of as "scientific" but important in many technological problems: demographic data, chemical production records, energy consumption figures, health statistics, and so on.

The three classes have sometimes been referred to as hard, semihard, and soft data, respectively. The boundaries between the classes are not sharp. For example, most biological data are time- or space-dependent, but a carefully designed experiment with an adequate sample can yield results with enough statistical significance to be placed in class A.

The Growing Need for Good Data

In more placid times, the typical scientist or engineer encountered no major problem in locating all the existing data that were pertinent to his needs. A small number of journals and handbooks, plus the informal peer group information chain, were generally sufficient. Judgments on the validity of the data so located could often be made on the basis of personal knowledge of the laboratory that generated the data. This process is becoming less and less tenable. Equally important, the most crucial and challenging problems that we deal with today usually require data that cut across traditional disciplinary lines. The peer group channels and personalized value judgments are no longer as effective as they once were.

As an example, consider the debates on depletion of the ozone layer by supersonic transport exhaust and release of chlorofluoromethanes. Elaborate mathematical models have been developed in an effort to predict the consequences of such atmospheric alterations. These models require various types of input data. Meteorological data (class B) enter into the prediction of transport of the pollutants into the stratosphere. Rate constants for several hundred chemical reactions and cross sections for photodissociation by solar ultraviolet radiation (class A) determine the complex chemistry occurring in the stratosphere. Long-term data on the ozone concentration as a function of latitude (class B) are necessary to provide a baseline against which perturbations caused by man can be compared. Epidemiological data (class C) are needed to predict the risk of diseases such as skin cancer induced by higher ultraviolet radiation levels. Numerous studies of ozone depletion (3) have rested heavily on these and other types of data.

Other examples are easily found. The national commitment to increase our dependence on coal as an energy source raises many questions that can only be answered through the use of adequate data bases. The needs include better data on the physical, chemical, and toxicological properties of compounds derived from coal; baseline data on pollutants likely to be released by expanded coal utilization; and a long list of data relevant to the effects of increasing atmospheric carbon dioxide. Even data on fatalities from grade-crossing accidents involving coal-carrying trains were used in one recent study of energy options (4). A comparable variety of data needs enters into the consideration of other energy sources.

It is safe to predict that examples such as these will multiply. Bitter debates on the risks from allegedly carcinogenic food additives and contaminants have already occurred. The implementation of the Toxic Substances Control Act of 1976 creates many needs for data—not only data on the toxic effects of chemical compounds, but also data on physical and chemical properties, chemical production and use, and environmental monitoring. These needs were explicitly recognized by Congress in section 25(b) of the act, which required the Council on Environmental Quality to coordinate a study of "the feasibility of establishing a standard classification system for chemical substances and related substances and a standard means for storing and for obtaining rapid access to information respecting such substances" (5).

Even aside from standards of health and safety which are subject to government regulation, most industries depend on good data bases for process selection, design, and control. Sudden shifts in the cost or availability of raw materials dictate changes in production conditions. The high cost of energy encourages fine tuning of processes to minimize energy consumption and optimize product yield. Long-term reliability of manufactured products becomes a factor in meeting foreign competition. Engineering design is done more systematically, with growing use of computer-based models for simulating actual processes. Of course, the answers obtained from these models are no better than the data put in.

In 1978 a National Academy of Sciences report (6) outlined the needs for data in various areas of national importance. The study group preparing this report, which was led by Walter H. Stockmayer, identified an extensive list of needs in programs such as energy, defense, space, and environmental protection. They concluded that current U.S. spending on data evaluation programs is insufficient, by a factor of 2 to 3, to meet these needs. An adequate program to evaluate and organize such data would cost only a fraction of 1 percent of the cost of generating the data; in the view of the study group, the benefits would greatly exceed this cost.

Quality Control of Data Bases

Scientific literature contains vast amounts of data collected for a specific purpose and presented by authors to support their conclusions. For example, a chemist claiming to have synthesized a new compound may present the boiling point, infrared spectrum, and other pertinent data on the compound. These data have great potential value, often for purposes that the author could never have anticipated. The history of science is filled with examples of old data being retrieved to test a new theory, provide a new insight, or apply to some technological problem.

Unfortunately, the quality of the data preserved in the literature leaves much to be desired. This becomes apparent when data on a much-studied subject are systematically retrieved. Figure 1 displays about 200 reported measurements of the thermal conductivity of copper as a function of temperature. The measurements were analyzed by the Center for Information and Numerical Data Analysis and Synthesis at Purdue University (7). The scatter of these data illustrates the pitfalls of relying on a single value retrieved from the literature. While this is perhaps an extreme case, aggravated by the sensitivity of the low-temperature thermal conductivity to impurities in the sample, many similar examples have been given (*6, 8*). Furthermore, many authors fail to specify auxiliary information (such as sample purity) in their papers, so an unsuspecting user may be badly misled.

Recognition of these problems led, some 60 years ago, to the preparation of the *International Critical Tables* (*9*). This broad-based effort required more than 10 years and produced a seven-volume set of carefully selected physical and chemical data. The project introduced the concept of critical evaluation of data by experienced scientists who could judge the accuracy of published data and resolve the numerous conflicts and inconsistencies that they found. The objective was to provide a comprehensive set of recommended data which could be used with confidence by the entire scientific and engineering community. The *International Critical Tables* became a standard reference work in all technical libraries; the original publisher continued sales until the stocks were exhausted a few years ago.

The philosophy established in the preparation of the *International Critical Tables* has been maintained in the National Standard Reference Data System (NSRDS), which was created in the mid-1960's (*10, 11*). This program is coordinated by the National Bureau of Standards and includes activities in academic, industrial, and government establishments. It is intended to provide critically evaluated data bases in important areas of the physical sciences on a continuing basis. In the 15 years that the program has been fully operational, its published data compilations in selected areas of physics and chemistry occupy 16 feet of bookshelf space. By contrast, the *International Critical Tables* volumes published in the 1920–1930 period, which attempted to cover all disciplines for which data existed, amounted to 9 inches.

A key feature of the NSRDS program is the maintenance of continuing data centers, each of which has cognizance over a well-defined disciplinary area (*11, 12*). These centers are staffed by scientists experienced in the relevant fields of research. A typical data center collects all pertinent literature on a continuing basis and evaluates subsets of data extracted from this literature. The data centers have developed methodologies suitable for evaluating each type of data (*11*). In thermodynamics, for example, the values of properties such as enthalpy, heat capacity, and equilibrium constants are linked through well-established laws. By systematically collecting all reported data for a given class of compounds, a data center can apply the laws of thermodynamics to detect any discrepant values and arrive at a preferred set of properties providing optimal agreement with all of the experimental data (*13*). Sophisticated computer programs have been developed to help carry out this procedure (*14*). It would be very difficult (and certainly inefficient) for individual scientists to go through this process; evaluation by a data center has obvious benefits.

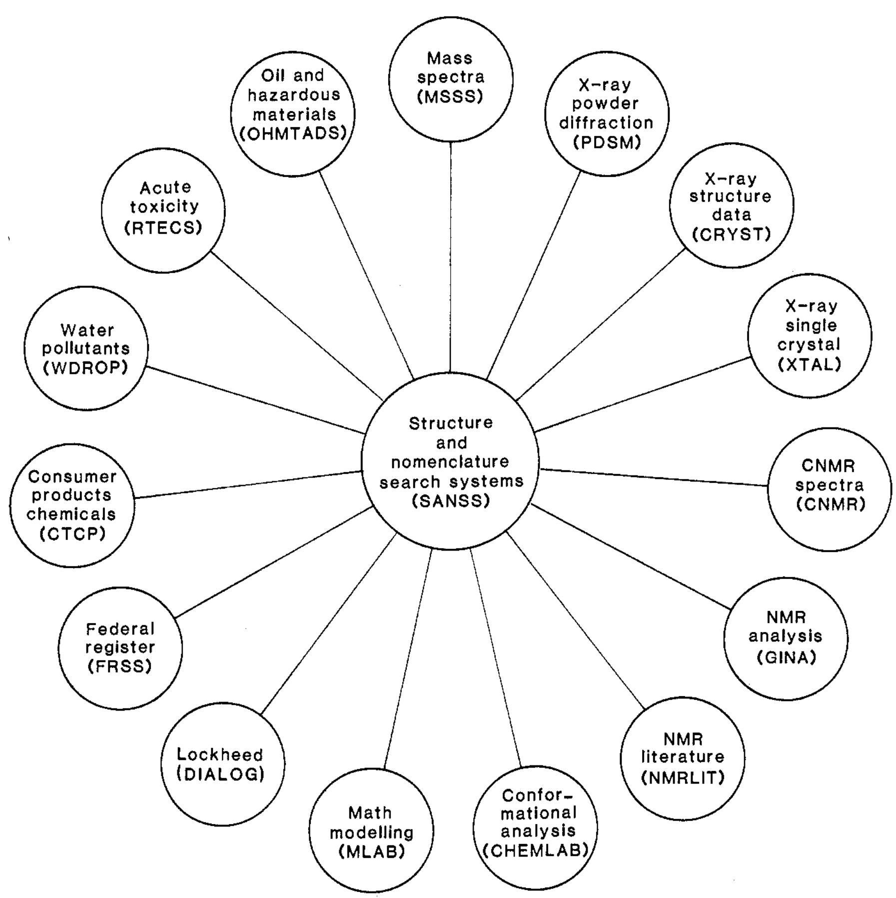

Fig. 2. Components of CIS in fall 1980.

The methodologies just described are most appropriate to class A data. In fact, the approach to a "best" or "recommended" value is a rather different process outside the physical sciences, as pointed out by Chayes (*15*). Nevertheless, certain efforts have been made, especially in the biosciences, to compile data selected by experts as representing the most typical values or ranges. A noteworthy example is the project that was conducted by the Federation of American Societies for Experimental Biology, which produced the *Biology Data Book* (*16*) and several other compilations.

The large quantities of class B data generated in the geosciences and astronomy are collected in various centers, such as the complex of World Data Centers (*17, 18*) first established during the International Geophysical Year. These centers organize the data into formats convenient for users and serve as distribution points for the interested community. In general, they do not carry out a selection or evaluation function, although they are very much concerned with smoothing large sets of observational data (*19*).

The most effective way to maintain the quality of class B data may be by careful control of the measuring instruments prior to acquisition of the data. It is also important to record and preserve all the auxiliary information required to use the data with confidence, such as calibration standards and ambient temperature and pressure. CODATA has published guides on the presentation of data in the geosciences and biosciences that give general rules of this nature (*20, 21*).

Quality control of class C data is an even more difficult task, especially when political or commercial considerations are involved. The questions raised about the accuracy of the 1980 census illustrate the problems in assessing the validity of this class of data. Consistent data on oil imports and consumption are surprisingly difficult to obtain, in spite of large-scale efforts by the Energy Information Administration of the Department of Energy and several other government agencies. This lack of reliable data appears to have contributed significantly to the 1979 gasoline crisis in the United States (*22*). One of the problems in handling class C data is the difficulty in achieving agreement on terminology and definitions.

The Electronic Revolution in Data Dissemination

Handbooks, journals, and other traditional publication formats still serve as the major source of data for most scientists and engineers. However, the increasing cost of composition and printing, as well as the difficulty in updating massive hard-copy volumes, are strong driving forces toward the use of modern computer and telecommunications technology for data dissemination. Computer-based formats offer several advantages:

1) It is easier to maintain currency of a data base through frequent updating.

2) More sophisticated search strategies are possible; for example, one may carry out multiparameter searches using Boolean operations, which is not practical with printed tables.

3) The data resulting from a search can be put into a computational program for further manipulation, without the need for human transcription.

4) Current projections indicate a decrease in storage and telecommunication costs, while all costs associated with printed matter are likely to continue to rise.

Computerized data dissemination appears to be developing along two parallel paths: distribution of data bases in some tangible storage medium (such as magnetic tapes) and distribution through on-line interactive networks. The former is widely used for certain types of data bases. The World Data Centers (*17*) provide many items in magnetic tape format, such as oceanographic data and geophysical data obtained from satellite observations. The National Technical Information Service handles magnetic tape distribution for a number of government agencies (*23*). The incorporation of data bases in the internal data processing systems of instruments such as mass spectrometers is a growing practice. This permits the results of a measurement of an unknown substance to be matched in real time against a library of reliable data, greatly facilitating the identification of unknowns (*24*).

Use of magnetic tapes and related media is most attractive when the user expects a continuing, reasonably heavy need for particular data. This justifies the effort required to install the data base and necessary software on his own computer. When the need is more sporadic, or when the user is unable to make the investment required to handle the data base in his own institution, access through an on-line network is appealing. This requires only a modest capital investment in a suitable terminal and perhaps a small subscription or entry fee. Beyond that, the user pays only for the data that he actually retrieves from the network.

```
TYPE PEAK,MIN INT,MAX INT
CR TO EXIT, 1 FOR REGN,QI,MW,MF AND NAME

USER: 237,100,100

  FILE #   # REFS                M/Z PEAKS

    1        26                     237

NEXT REQUEST: 293,30,40

  FILE #   # REFS                M/Z PEAKS

    2        1                    237 293

NEXT REQUEST: 1

  REGN    QI   MW   MF                        NAME
78-00-2  704  324  C8H20Pb
                         Plumbane, tetraethyl- (8CI9CI)
                         Lead tetraethyl
                         Lead, tetraethyl-
                         Tetraethyllead
                         Tetraethylplumbane
```

Fig. 3. Use of the MSSS for chemical identification. The number entered under *QI* gives a relative measure of the quality of the spectrum that was retrieved.

On-line information retrieval has become familiar to many scientists and engineers through their use of the bibliographic files available through services such as DIALOG, ORBIT, BRS, MEDLINE, TOXLINE, and many others (*1*). Over 2 million individual searches are estimated to have been made in 1980 through services such as these, which provide access to several hundred bibliographic files in science and technology. Interest is growing in extending on-line services from bibliographic files to numerical and factual data bases. There are, in fact, at least 150 such data bases already available to the public through on-line services (*25*). Although the majority are in the business and economics area, a significant amount of hard scientific data can be accessed on-line (*26*). The existence of these services demonstrates that there are no serious technological barriers to the growth of on-line scientific and technical data bases.

One on-line data service that has become well established is the Chemical Information System (CIS), which was initiated by the National Institutes of Health and the Environmental Protection Agency and now includes participation by several other agencies (*27*). CIS now has over 500 subscribers, who make a total of about 25,000 searches per month. The structure of CIS is illustrated in Fig. 2. Some of the currently active components, such as MSSS and PDSM, contain numerical data bases in the sense defined in this article. Others contain descriptive information that includes some numerical data; for example, OHMTADS gives a concise narrative statement of the degree of hazard to public health of each compound in the file plus quantitative data, when available, on its toxicity. The NMRLIT and FRSS components are bibliographic. Each component includes appropriate software for retrieving the data or other information contained therein.

The components of CIS are linked to a central hub known as the Structure and Nomenclature Search System (SANSS). This is a file of nearly 200,000 chemical compounds containing the formula, name, synonyms, and complete structure record for each compound. The file can be searched on the basis of structure, substructure, full or partial name, or formula. Once a compound is identified through SANSS, the user is informed which CIS components contain information on the compound, and he can be transferred to the component of interest by means of a simple command. The Chemical Abstracts Service registry number serves as the identifier for a chemical compound which permits the linking of the individual data bases to SANSS.

Several components of CIS are designed for chemical identification. Figure 3 shows how a simple search is conducted on the MSSS component. The user inserts each peak (mass-to-charge ratio) in the mass spectrum of the unknown material in turn. In response to the first peak, the system replies that the data base contains 26 compounds that might match the unknown. Insertion of a second peak narrows the choice to a single compound. The user is then presented with the molecular formula, proper name, and other common names of the compound. Another command will produce a listing of the complete mass spectrum to confirm the identification. The entire process requires about 2 minutes and costs about $1.50. With a suitable graphics terminal, the user can obtain a bar-graph presentation of any spectrum in the data base (Fig. 4).

The CIS illustrates the power of an online search system to provide a variety of information in a quick and convenient manner. For example, a user might employ MSSS or XTAL (Fig. 2) to identify an unknown substance. He can then determine from other CIS components whether the substance has been detected in certain environmental monitoring records, its physical properties and toxicity, and whether it has been cited in recent government regulations. Although some

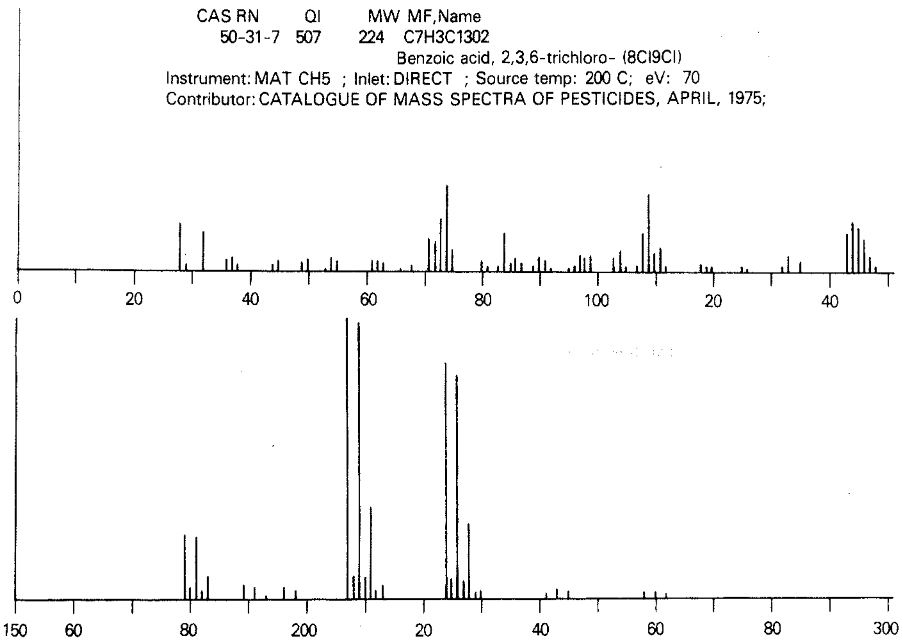

Fig. 4. Bar-graph display of a mass spectrum in CIS.

of the components are not yet completely developed, the advantages of the CIS concept are clear. The user gains rapid access to diverse classes of information which would be quite laborious to obtain by conventional methods.

Other on-line systems for disseminating numerical data are being developed here and abroad. The Metals Properties Council, in collaboration with several other organizations, is planning an on-line materials property system. The American Society for Metals and the National Bureau of Standards are cooperating on plans for on-line dissemination of alloy phase diagrams. Lawrence Livermore Laboratory has developed an extensive on-line system that emphasizes data related to energy storage and conservation (*28*). In France an on-line network called PLURIDATA is now in operation, and numerical data bases of scientific interest will be included in the EURONET-DIANE network (*29*). All observers foresee very rapid growth in the next 5 years.

There are, however, certain problems that must be overcome before the potential of these on-line systems is realized. Start-up costs are high and the economics of such systems are very complex (*30*). The most efficient and equitable basis for charging users is not always clear. Standardization is a major problem. Even with bibliographic services, a user must learn the language and protocols peculiar to each service. In dealing with numerical data, the more complex file structure and greater variety of formats exacerbate this problem. While much attention is being devoted to designing "user-friendly" software that will minimize the need to master a complex set of instructions, this goal is not likely to be reached quickly or easily.

Finally, there is the central problem of quality assurance. In conventional publications of reference data, it is possible to include expository text and footnotes pointing out the accuracy limits and explaining any caveats on the use of the data. Also, citations can be given so that an interested user can track each recommended value to its sources. The more rigid constraints imposed on a computerized data file make it difficult to convey such auxiliary information to the user. It is very important, therefore, that the contents of an on-line data base be carefully verified and that the user be assured that the values he retrieves are the best available. The National Bureau of Standards is following this policy in all its computerized data bases of physical and chemical properties; in general, the computer version will be backed up by publications that document the choices of data and explain the evaluation process. Unfortunately, such a policy has not been clearly established for other types of data. The practice of putting unverified data into publicly available systems is very risky.

The Need for Cooperation

The cost of developing on-line services and of creating the data bases that go into them is not trivial. Systematic planning and cooperation can help reduce these costs. While some countries are planning highly centralized on-line information networks, it seems unlikely that the United States will develop a comprehensive master plan. The private information industry, the nonprofit societies, and the federal government have their own distinct interests and responsibilities, so a pluralistic approach is inevitable. There are strong arguments, in fact, that no one can predict the optimum design for on-line data systems. A reasonable level of competition is desirable in order to provide users with a variety of options and allow the marketplace to decide which are most effective. Nevertheless, some central coordination is desirable, if only to encourage a degree of compatibility that will benefit all services.

While such arguments can be applied to the systems for delivery of the data, it is more difficult to justify duplication in the creation of the data bases. This is a very expensive process, especially if adequate quality control is exercised. Thus there are strong incentives for collaboration in building data bases and for reaching agreement among interested organizations so as to avoid duplication of effort. In the area of physical science data, the Office of Standard Reference Data of the National Bureau of Standards carries out this coordinating role on a national basis. It currently sponsors a number of groups engaged in the development of reliable physical and chemical data bases and encourages all parties to channel their efforts into supportive rather than competitive activities. The Numerical Data Advisory Board of the National Academy of Sciences provides a forum for groups concerned with broader classes of data bases. However, more formal coordinating mechanisms are needed for data base activities in fields outside the physical sciences. This is a subject that warrants government attention at a high level.

The evolution of on-line bibliographic services may provide a model for developing on-line data systems. Several services now coexist in the United States, and others are coming on-line elsewhere in the world. All these services offer the same bibliographic files in chemistry, physics, biology, and other major disciplines, but they differ significantly in their search software and other operating features. From the users' viewpoint, this diversity is a great advantage. However, no one would suggest duplication of the enormous effort required to create, for example, the Chemical Abstracts file. On-line data dissemination should follow a similar pattern, with the same numerical data bases available on several different systems, each perhaps emphasizing a different type of application or a different group of users. Services that combine numerical data bases with bibliographic files and other information may also prove attractive. There appear to be many opportunities for creative thinking in the design of on-line services; what is most needed at this stage is rapid development of a sufficient number of good data bases to go into these services.

Cooperation at an international level can also expedite the development of on-line services and help reduce the costs. The World Data Centers of the International Council of Scientific Unions provide a useful precedent for the exchange of geophysical data in an organized fashion between distribution points in different countries or regions. The International Atomic Energy Agency has performed a similar function for nuclear and other data relevant to reactor development. In 1966 the International Council of Scientific Unions established CODATA as a standing organization concerned with various aspects of data collection, evaluation, and dissemination. CODATA has served as a medium for international efforts to standardize data handling and presentation. Also, through its biennial conferences, it brings together scientists from all disciplines who are concerned with data problems. CODATA is expected to play an important role in promoting international cooperation in development of the computer-based dissemination systems of the future (*31*).

Conclusions

Predicting the changes in data dissemination and use over the next decade requires considerable courage. Nevertheless, certain trends are clear.

Needs for reliable data will become more pressing. Many political and economic decisions important to our society will rest on scientific and technical data. The accuracy of these data will frequently be challenged in adversary proceedings, placing a burden on the scientific community to concern itself more deliberately with the quality of the available data.

Computer-based data dissemination methods, especially on-line systems, will grow in use. This process will be accelerated by the entrance of younger scientists who have become comfortable with computer terminals during their education. The printed handbook is not likely to disappear but will gradually become a by-product of the computerized data bases.

Coordination in the development of computer-based systems will be essential. There is danger of wasteful duplication in the creation of machine-readable numerical data bases. Such duplication could greatly increase the overall costs of implementing on-line systems, which will in any case be very expensive. Furthermore, the linkage of related on-line systems will require that careful attention be given to questions of compatibility and format standardization. This coordination must be carried out at an international level.

We cannot take for granted that the data generated by the research establishment will automatically flow smoothly to those who need it. Changes in attitude are required by the scientific community, industry, and the federal government. The scientific community must place a higher priority on organizing the data it produces and presenting these data in a form suitable for technological applications. Private industry should put more resources into developing data bases to support long-term industrial needs (*32*). The federal government must recognize that its commitment to supporting basic research for the long-range benefit of the country also implies a commitment to make the results available in a form that maximizes their utility.

References and Notes

1. T. E. Doszkocs, B. A. Rapp, H. M. Schoolman, *Science* **208**, 25 (1980); D. B. McCarn. *Annu. Rev. Inf. Sci. Technol.* **13**, 85 (1978); M. E. Williams, *Bull. Am. Soc. Inf. Sci.* **4**, 21 (1977); M. E. Williams, Ed., *Computer-Readable Data Bases: A Directory and Sourcebook* (American Society for Information Science, Washington, D.C., 1979).
2. M. Kotani, Ed., *CODATA Bull. No. 16* (1975).
3. *Environmental Impact of Stratospheric Flight* (National Academy of Sciences–National Research Council, Washington, D.C., 1975); *Stratospheric Ozone Depletion by Halocarbons: Chemistry and Transport* (National Academy of Sciences–National Research Council, Washington, D.C., 1979); *The Stratosphere: Present and Future* (National Aeronautics and Space Administration, Washington, D.C., 1979).
4. *Energy in Transition, 1985–2010* (National Academy of Sciences–National Research Council, Washington, D.C., 1980).
5. The results of this study are presented in a report by the Council on Environmental Quality, "The feasibility of a standard chemical classification system and a standard chemical substances information system" (Stock No. 041-011-00039-4, Government Printing Office, Washington, D.C., July 1978).
6. Committee on Data Needs, "National needs for critically evaluated physical and chemical data" (Numerical Data Advisory Board, National Academy of Sciences–National Research Council, Washington, D.C., 1978).
7. C. Y. Ho, R. W. Powell, P. E. Liley, *J. Phys. Chem. Ref. Data* **3** (Suppl. 1) (1974).
8. L. J. Kieffer and G. H. Dunn, *Rev. Mod. Phys.* **38**, 1 (1966); L. M. Branscomb, *Sci. Res.* **3**, 49 (1968); A. H. Rosenfeld and P. R. Stevens, in *Proceedings of the Fifth International CODATA Conference*, B. Dreyfus, Ed. (Pergamon, Oxford, 1977), p. 19; D. R. Lide and M. A. Paul, Eds., *Critical Evaluation of Chemical and Physical Structural Information* (National Academy of Sciences–National Research Council, Washington, D.C., 1974).
9. E. W. Washburn, Ed., *International Critical Tables of Numerical Data, Physics, Chemistry, and Technology* (McGraw-Hill, New York, 1926–1933).
10. E. L. Brady and M. B. Wallenstein, *Natl. Stand. Ref. Data Ser. Natl. Bur. Stand. No. 1* (1964). The major output of this program appears in the *Journal of Physical and Chemical Reference Data*, which is published jointly by the National Bureau of Standards, American Institute of Physics, and American Chemical Society.
11. D. R. Lide and S. A. Rossmassler, *Annu. Rev. Phys. Chem.* **24**, 135 (1973).
12. Further information on the NSRDS may be obtained from the Office of Standard Reference Data, National Bureau of Standards, Washington, D.C. 20234.
13. D. D. Wagman, D. Garvin, V. B. Parker, W. H. Evans, J. B. Pedley, in *Proceedings of the Seventh International CODATA Conference*, P. Glaeser, Ed. (Pergamon, Oxford, 1981), p. 361.
14. D. Garvin, V. B. Parker, D. D. Wagman, W. H. Evans, *Natl. Bur. Stand. Intern. Rep. 76-1147* (1976).
15. F. Chayes, in *Proceedings of the Seventh International CODATA Conference*, P. Glaeser, Ed. (Pergamon, Oxford, 1981), p. 116.
16. P. L. Altman and D. S. Dittmer, *Biology Data Book* (Federation of American Societies for Experimental Biology, Bethesda, Md., ed. 2, 1972–1974), vols. 1, 2, and 3.
17. Panel on World Data Centers, *Fourth Consolidated Guide to International Data Exchange Through the World Data Centers* (International Council of Scientific Unions, Washington, D.C., 1979). Further information may be obtained from the Geophysics Research Board, National Academy of Sciences, 2101 Constitution Avenue, NW, Washington, D.C. 20418.
18. A. H. Shapley, in *Proceedings of the Fifth International CODATA Conference*, B. Dreyfus, Ed. (Pergamon, Oxford, 1977), p. 277; C. Jaschek and F. Ochsenbein, in *ibid.*, p. 281.
19. A. H. Shapley and R. Tomlinson, in *Data Handling for Science and Technology*, S. A. Rossmassler and D. G. Watson, Eds. (North-Holland, Amsterdam, 1980), p. 41.
20. G. A. Wilkins, *CODATA Bull. No. 32* (1979).
21. P. L. Altman, Ed., *CODATA Bull. No. 25* (1977).
22. M. Waterhouse and S. Abbasi, *Energy Information and Data: Sources, Reliability, and Problems* (Issue Brief No. IB 79106, Congressional Research Service, Washington, D.C., 1980).
23. *Directory of Computer Software and Related Technical Reports* (PB80-110232, National Technical Information Service, Springfield, Va., 1980).
24. A data base containing mass spectra of over 30,000 compounds is available on magnetic tape through the Office of Standard Reference Data, National Bureau of Standards, Washington, D.C. 20234.
25. Cuadra Associates (Santa Monica, Calif.), private communication.
26. J. Hilsenrath, *Natl. Bur. Stand. (U.S.) Tech. Note 1122* (1980).
27. G. W. A. Milne and S. R. Heller, *J. Chem. Inf. Comput. Sci.* **20**, 204 (1980); S. R. Heller and G. W. A. Milne, *Database* **3**, 45 (1980); *Environ. Sci. Technol.* **13**, 798 (1979).
28. V. E. Hampel, in *Proceedings of the Seventh International CODATA Conference*, P. Glaeser, Ed. (Pergamon, Oxford, 1981), p. 472.
29. G. W. P. Davies, in *Proceedings of the Sixth International CODATA Conference*, B. Dreyfus, Ed. (Pergamon, Oxford, 1979), p. 201.
30. S. R. Heller, in *Proceedings of the Seventh International CODATA Conference*, P. Glaeser, Ed. (Pergamon, Oxford, 1981), p. 578.
31. The CODATA secretariat is located at 51 Boulevard de Montmorency, Paris. Information may also be obtained from the U.S. National Committee for CODATA, Numerical Data Advisory Board, National Academy of Sciences, 2101 Constitution Avenue, Washington, D.C. 20418.
32. An encouraging step was the recent establishment of the Design Institute for Physical Property Data by the chemical industry. This organization, which is managed by the American Institute of Chemical Engineers but which receives financial support from about 50 chemical companies, will compile and distribute chemical engineering data of interest to the members (*Chem. Eng. News*, 17 December 1979, p. 19).