(c) The results of (a) and (b) will be reported in a Directory to be publicly available and to be revised at intervals.

(d) It is expected that the acts of carrying out steps (a) and (b) will result in the establishment in each country and each union of committees to provide support to the national and union representatives, respectively. (Such committees have already been established in the U. S. National Academy of Sciences and the U. K. Royal Society.)

(e) When appropriate, informal links will be established with governmental and intergovernmental groups concerned with critical tables or reference data projects.

(f) Subcommittees will be established to provide international communications on special relevant topics. (For example, a Task Group is in process of formation for exchange of information of computer applications to numerical data handling.)

(g) Plans have been initiated for a "Gordon-type conference" to be held in Western Europe in 1968 to be cosponsored by the ICSU Committee and a suitable European organization.

The above list is not exhaustive but indicates the variety of ways in which the CODATA can be helpful.

The title of this paper, "A World System of Evaluated Numerical Data for Science and Technology," expresses a hope and a need—not a neat blueprint. The system, if it can be created, will be the sum of numerous publications produced in many countries in many different ways. The funding of the efforts leading to the publications may be by governments, societies, trade associations and other industrial groups, and by private publishers. A leading and probably dominant role in supplying material for the integrated world system will be played by organized national efforts such as the National Standard Reference Data Program in the U. S. If communications can be established among compilers of the world, then standards of excellence can be agreed on and adhered to, undesirable overlap can be avoided, and a planned coverage of all areas of science can be maintained. Achievement of such international cooperation will require a willingness of all participants in the program to work together for the benefit of science and industry in all countries. The ICSU Committee on Data for Science and Technology will do all it can to bring about this needed world-wide cooperation.

LITERATURE CITED

(1) "International Critical Tables of Numerical Data: Physics, Chemistry and Technology," McGraw-Hill Book Co. for the National Research Council, 1926-1933, New York, N. Y.
(2) "Tables de Constantes et Données Numériques," fondées par Ch. Marie (1909): "Constantes Sélectionnées," Paris, Pergamon Press; current series (1947-).
(3) Landolt-Börnstein, "Zahlenwerte und Funktionen aus Physik-Chemie-Astronomie-Geophysik und Technik," 6th ed. (1950-).

Organizing Physical Molecular Data for Qualitative Chemical Analysis*

L. E. KUENTZEL
Wyandotte Chemicals Corporation, Wyandotte, Michigan 48193

Received October 20, 1966

The American Society for Testing and Materials Committee E-13 on Molecular Spectroscopy has been active for 15 years in the area of indexing and searching infrared, ultraviolet, and visible absorption data. Committee E-14 has made significant contributions in its field and, although the data concerned do not fall in the category of molecular data, the contributions of Committee E-19 on Gas Chromatography and the Joint Committee on X-ray Powder Diffraction deal with the same kinds of problems for the same end results. Of recent years, the National Standard Reference Data Program of the National Bureau of Standards has provided added impetus and support in some of these areas. General descriptions of these activities are given, including recent advances made through ASTM and the Coblentz Society via NSRDP support.

In a simple interpretation of the process of using physical molecular data for qualitative chemical analysis, such data have been likened to fingerprints of compounds, and the analogy of making identifications by comparison of unknown data with files of data obtained from known materials is made. Further, it would appear obvious that the larger the number of different known fingerprints in the files, the better the prospects of identifying any given unknown. Also, as with human fingerprints, most molecular data are rather complex and require mechanized indexing and searching procedures for maximum efficiency in utilization. However, the analogy soon begins to break down, for although there are 10 different fingerprints per

* Presented before the Division of Chemical Literature, Symposium on Compilations of Data on Chemical and Physical Properties of Substances, 152nd National Meeting of the American Chemical Society, New York, N. Y., Sept. 12, 1966.

individual, all yield the same kind of information; whereas each different type of molecular data gives us different information about the molecule involved. This makes the use of more than one type of data much more effective in the identification of a given unknown, so that it has become common practice to make use of as many different types of data as are needed.

This adds a new dimension to the problem of organizing such data: Now we must not only develop systems for mechanizing the searching of each type of data, but the several systems must be compatible so that searches may be made consecutively, if not simultaneously, on different files, with the results being compared by machine for a most probable answer. For example, it is now within the realm of possibility to search, by computer, all of the published infrared spectral data, x-ray diffraction data, ultraviolet spectral data, and mass spectral data to make fingerprint matches with a set of similar data obtained from an unknown compound and to come up with the name of the known compound that most nearly matches these specified properties of the unknown.

This situation did not develop overnight. The first embryonic idea for doing these jobs by machine was presented in a paper given at the 1951 Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy and published later that year in *Analytical Chemistry*, page 1413. Even then, systems for infrared spectral data and x-ray powder diffraction data were in practical operation. Expansion to other types of data followed and continues to this day. Refinement of computer search techniques and the generation of book indexes and lists via computer output make possible the efficient handling of large numbers of new data, with corresponding increases in effectiveness of the indexing systems. Thus assured of efficient handling of large masses of such data, one can press for faster rates of publication and the upgrading in quality of such molecular data. These latter aspects of the program are currently receiving the support of the National Standard Reference Data Program of the National Bureau of Standards.

From the beginning, committees of the American Society for Testing and Materials have been associated with the development and maintenance of indexes of data used as reference material in the qualitative analysis of substances. Some activities have grown to sizeable operations, and newer ones give promise of growth. It is the purpose of this paper to describe some of these activities briefly and to report on the latest developments.

ASTM Joint Committee on Chemical Analysis by Powder Diffraction Methods began assembling and publishing powder diffraction patterns to serve as references for the identification of crystalline substances in the early 1940's. Indexing was by means of the multiple listing of the three strongest lines of each pattern which became known as the Hanawalt Index. The patterns are published on 3 × 5 cards as lists of d-values and corresponding intensities for the lines. In the late 1940's an index system which made use of IBM cards and sorters was introduced. This permitted machine searching based on the presence or absence of all important lines, the presence or absence of specific elements, and other structural features of the compounds involved. Because of the relatively small number of patterns involved and the efficiency with which the Hanawalt Index worked, the IBM card system did not gain wide acceptance but it was maintained for special applications and it formed a foundation for the recent expansion into computer systems.

The Joint Committee is continuously concerned with the quality of its x-ray powder data and gives constant attention to methods of improving the indexing and use of such data. This has resulted in the setting up of numerous fellowship grants for responsible laboratory investigations of questionable data and the development of a variety of indexing schemes. The latter include (1) edge-notched or Keysort indexes for sorting with needles, (2) an optical coincidence or "peek-a-boo" index which makes use of perforated cards stacked on a light box, (3) an expansion of the Hanawalt Index to include multiple listing of the six strongest lines along with a listing of the eight strongest lines at each entry, (4) computer searches based on the presence or absence of selected lines plus their intensity values together with the presence or absence of elements, (5) a computer-generated bar graph representation of each pattern so that quick visual comparisons may be made, and (6) a book index which lists each pattern but once but is so organized that rapid scans to different sections of the book are possible.

The collecting and indexing of mass spectral data are a project of Committee E-14. Both book and IBM card indexes are currently available. The former, generated by computer from the cards, lists the six strongest peaks, their relative intensities, the molecular weights and names of compounds involved, together with references to the source of the data. There are five sections in the book; one is a numerical listing by molecular weight and the others are arranged in numerical order of the four strongest peaks. A new edition of the index, in course of preparation, will bring the total number of spectra indexed to approximately 8000. The same data are available in IBM cards, and programs for computer manipulation of the data are under development. An optical coincidence-type index is also being prepared for Committee E-14 by Jonkers Corporation.

Gas chromatographic data are hardly molecular data in the sense of revealing molecular structure, but their use in qualitative chemical analysis make them logical companions of other types of data used for this purpose. Committee E-19 on Gas Chromatography has undertaken to organize such data into indexes which relate sets of data to individual compounds for qualitative analytical purposes. The data are first punched into IBM cards which are then used to prepare the book indexes. Data used for this include relative retention, specific retention volume, column temperature, concentration of liquid phase, solid support, reference material, etc., together with the names of the compounds involved and references to the sources of the published material. Over 2000 compounds are included in the current edition of the book and considerable expansion and revision are in progress. The IBM cards are also available for hand files and machine manipulation.

The activities of Committee E-13 on Molecular Spectroscopy embrace the largest indexes of the most widely used data. An infrared absorption spectrogram is one of the most complicated sets of information one can obtain about a molecule, because its nature depends upon the

vibrational behavior of every atom involved. Consequently, such a spectrogram is the most revealing physical data a molecule can yield when properly interpreted. For this reason and the relative ease with which modern instruments produce such data, infrared spectroscopy has become the most widely and effectively used tool for qualitative analysis and structure delineation. The complexity of infrared data has made it difficult to index and expensive to publish but much has been done. Also, problems of instrumentation have made it difficult to get uniformly good data, but recent progress in instrumentation and the development of recommended practices for obtaining and evaluating good spectral data should improve the quality of the data going into the indexing systems.

From its inception in 1950, Committee E-13 has been concerned with the quality and indexing of infrared spectral data. It is somewhat surprising that the indexing system first adopted in those early days is still in use today and forms the foundation for recent advances into computer searching, optical coincidence indexing, and computer-generated book indexes and lists. Very briefly, the system provides for coding a description of an infrared spectrogram in terms of absorption bands present, bands absent, and very strong bands, into an IBM card together with a detailed code of the chemical structure of the compound involved, the elements present, physical and optical characteristics, melting or boiling point, and a partial molecular formula for the compound. The name of the compound and reference to the source of the spectral data complete the information in each card. This permits one to make use of machines or computers to determine whether any one or combination of these factors is present, conditionally present, or absent in each of some 100,000 sets of spectral and structural data that represent essentially all published infrared spectra. The combination of factors used for the search is determined by the spectral data obtained from the unknown substance together with what other reliable information is available. The answer to the search is delivered in the form of a short computer listing or a small deck of IBM cards that give references to the names of the compounds and published spectral data for those substances with corresponding factors that matched, or most nearly matched, those of the unknown.

On a computer this operation takes but a few minutes. On a sorter the operation can take from a few minutes to the better part of a day depending upon the preciseness and nature of the factors available and upon the organization of the card files. Both tapes and cards are available to interested laboratories. A spectral data search service via computer is being considered by Eastman Kodak Company. This will be available to anyone desiring it.

Having these spectral data and chemical structure details in IBM cards make possible useful by-products. Computer-generated book indexes include a listing in numerical order of the molecular formulas, compound names, and references to the original publications for 100,000 compounds and spectra. This list permits one to determine rapidly whether the infrared spectrum of a given compound has been published and, if so, where. A second listing is arranged in numerical order of the serial numbers so that one may quickly ascertain the name of the compound and a reference to the original publication for any serial number that results from the application of any of several search systems.

A book is in the course of preparation which will list the eight strongest bands in each infrared absorption spectrogram *in the order of their relative intensity*. This will be the first attempt to make use not only of wave length or frequency data, but also of band intensity information for book index comparisons. Data in the cards have also been put to work in the preparation of specialized optical coincidence or Termatrex indexes. These indexes have great browsing flexibility, are relatively inexpensive, and require very little time.

Present indexes must be limited to 10,000 or fewer spectra in one set of optical coincidence cards; corrections and updating still pose some problems, but as an index to a relatively fixed set of spectra numbering less than 10,000, there is nothing better. The first such index issued by ASTM included the spectra of all hydrocarbons so that, if one knew an unknown was a hydrocarbon, it would take but a few minutes to determine whether a matching infrared spectrum existed in the published literature. Other such special indexes are being planned.

Finally, because of the rapid growth in the total number of infrared spectra indexed which creates problems of economics, storage space, and sorting time for some laboratories, ASTM now offers specialized search services and the preparation of specialized decks of cards to individual customer specifications. It should be possible for any infrared laboratory to have decks of cards indexing compounds pertinent to its interests selected from the complete master deck and thus represent the entire published literature of spectra in the particular categories. Such smaller decks give added economy and flexibility to laboratories using the IBM cards. For example, the small deck of cards indexing all of the published infrared spectra of hydrocarbons is available, as mentioned above. Moreover, the IBM cards offer unequaled flexibility in updating, correcting, and further processing at such a low cost that one could afford many such specialized decks for the cost of one special optical coincidence index.

In 1964, with the help of the National Standard Reference Data Program of the National Bureau of Standards, ASTM instituted a project for collecting, evaluating, processing, and disseminating infrared spectra. It was recognized at the outset that the process of evaluation would be most important, so the Coblentz Society was invited to participate. Through the efforts of the Coblentz Board of Management and NSRDP funds, a set of specifications for the evaluation of infrared spectra for reference purposes was prepared. As published in a recent issue of *Analytical Chemistry* (August 1966), these specifications recognize three types of spectra: Class I or "Standard Spectra," which are of sufficiently high quality as to be acceptable as physical constants of the substances and obtained under such precisely defined conditions that further refinements in spectrophotometric techniques would not be expected to change them significantly. Few, if any, such spectra exist today. Class II or "Reference Spectra" are those obtained on the best currently available commercial grating spectrophotometers operated with maximum efficiency. Purity of the compounds involved must be rigidly specified. Class III spectra are also reference spectra but are those obtained

from any good commercial spectrophotometer, properly operated so as to give spectra with sufficient accuracy to be useful in the identification of unknown materials. Most published infrared spectra now fall in this category, but it is hoped that, with detailed specifications now available for guidance, more Class II spectra will be produced and published. The development of detailed specifications for Class I or Standard Spectra has been left to international effort through the Commission on Molecular Structure and Spectroscopy of the International Union of Pure and Applied Chemistry.

The process of collecting spectra is under way, and some 2000 contributed spectra are being organized for the evaluation process. All infrared laboratories interested in the objectives of expanding and improving published infrared spectra for qualitative analytical purposes under the sponsorship of the National Standard Reference Data Program and the American Society for Testing and Materials are invited to participate. Arrangements have been made for microfilming, at no cost to the contributing laboratory, such collections of spectra as would be helpful in the evaluation program and which might well be published also. The program contemplates publication of only such spectra that are superior in quality to any that already have been published for each given compound or of cases where the published spectrogram is relatively inaccessible to laboratories in this country. Results of the evaluation process will generally be made available to the contributing laboratories.

Finally, recognizing that there are a number of organizations interested in the process of using physical molecular data for qualitative chemical analysis, ASTM proposed the formation of a Joint Committee on Atomic and Molecular Physical Data. This has been done with charter members from the Coblentz Society, the Society for Applied Spectroscopy, the Canadian Association for Applied Spectroscopy, the Manufacturing Chemists Association, the American Petroleum Institute, and the American Society for Testing and Materials. It is the purpose of this new committee to pool talents, influence, and effort for a common objective of generating, collecting, evaluating, editing and approving for publication, and encouraging the distribution of atomic and molecular data in suitable form to serve as reference standards for pure compounds and mixtures. Initially concerned primarily with infrared spectral data, exploratory examination of programs in mass spectrometry and nuclear magnetic resonance have begun.

A corollary activity of the infrared spectral data indexing project should be of interest to those concerned with the mechanized handling of large masses of data or information where a link between such stored material and its origin in the open literature must be maintained. The publishing of lists of tens of thousands of citations to published infrared spectra prompted an investigation into the use of CODEN for concise references. CODEN are five-letter mnemonic codes for the titles of periodicals which permit the reduction of a reference to five letters plus a few digits to express volume, page, and year. Use of these codes by others has now grown until they include many industrial, institutional, and governmental organizations in many countries of the world. From some 3000 codes in 1961, when ASTM assumed custody of the project from its originator, Dr. Charles Bishop, the number has grown to 40,000 CODEN. New ones are being added by request of users at a rate of nearly 1000 per month. A new edition of *CODEN for Periodical Titles* is now available and a special service for making new code assignments has been in operation for some time. CODEN and titles are also available in computer tape and in IBM cards with a monthly updating service. An example of the use of CODEN may be found in *Chemical Titles* issued by Chemical Abstracts Service. CODEN will also be included in the next edition of the *List of Periodicals*.

Numerical Data Activities of Engineering Societies*

FRANK Y. SPEIGHT
Engineers Joint Council, 345 East 47th Street, New York, New York 10017

Received October 7, 1966

This paper will consider only the area of physical properties of substances, further considering that mechanical properties—tensile and compressive strengths, hardness, shear strength, etc.—are physical properties. Engineers' data interests extend far beyond intrinsic properties of relatively pure materials that are well defined in composition and structure and include also those properties which are dependent upon an interaction of the material and the measuring system, as for example, hardness, breaking strength, and the like. Of course, engineers also need data on intrinsic properties of many materials too, such as the steam tables for the mechanical engineer designing high pressure power systems and the thermodynamic properties of chemicals used by the chemical engineer in designing process equipment. In general, an engineer's

* Presented before the Division of Chemical Literature, Symposium on Compilations of Data on Chemical and Physical Properties of Substances, 152nd National Meeting of the American Chemical Society, New York, N. Y., Sept. 12, 1966.