consists of calculating the uncertainties in the smoothed and selected data stored in the file. In general, the property values and their uncertainties as well are functions of the independent variables of the system. Although this is a mathematical problem, no general solution considering the functional dependence among thermodynamic variables has been found. The final step is to calculate the uncertainty in the values of properties retrieved from the file. This is a trivial problem if the retrieved property is identical with a stored property, but if the retrieved property is calculated by combining several stored properties, especially if derivatives or integrals are involved, it can be a very difficult problem. This whole subject demands much additional work and thought.

REFERENCES AND NOTES

(1) F. V. Wetzler, "Data Banks R and D", *Res./Dev.*, **28**(6), 54-64 (1977).
(2) R. W. Counts, Y. S. Touloukian, and J. W. Phillips, "A New Dimension to Numerical Data Banks", *Res./Dev.*, **27**(1), 36-38 (1976).
(3) S. R. Heller, G. W. A. Milne, and R. J. Feldman, "A Computer-Based Chemical Information System", *Science*, **195**, 253-369 (1977).
(4) S. R. Heller and G. W. A. Milne, "The NIH/EPA Chemical Information System", a chapter in "Retrieval of Medicinal Chemical Information", M. Milne and J. Howe, Ed., ACS Symposium Series No. 84, American Chemical Society, Washington, D.C., 1978, Chapt 10.
(5) G. W. A. Milne, S. R. Heller, A. E. Fein, E. F. Free, R. G. Marquart, J. A. McGill, J. A. Miller, and D. S. Spiers, "The NIH-EPA Structure and Nomenclature Search System", *J. Chem. Inf. Comput. Sci.*, **18**, 181-186 (1978).
(6) L. M. Rose, "A Survey of Available Physical Property Data Banks", *ACHEMA Jahrb.*, I, 32 (1977/79). A description of 23 data banks collected by the Working Party for Information and Documentation of the European Federation of Chemical Engineers.
(7) J. N. Peterson, C.-C. Chen, and L. B. Evans, "Computer Programs for Chemical Engineers: 1978—Part 1", *Chem. Eng. N.Y.* 85, 145-154 (June 5, 1978); Part 2, *ibid.*, 69-82 (July 3, 1978); Part 3, *ibid.*, 79-86 (July 31, 1978); Part 4, *ibid.*, 107-115 (Aug 28, 1978).
(8) D. Garvin, V. B. Parker, D. D. Wagman, and W. H. Evans, "A Combined Least Sums and Least Squares Approach to the Evaluation of Thermodynamic Data Networks", Interim Report, Office of Standard Reference Data, NBSIR 76-1147, July 1976.
(9) W. H. Evans and D. Garvin, "The Evaluator Versus the Chemical Literature", *J. Chem. Doc.*, **10**, 147-150 (1970).
(10) "Critical Evaluation of Data in the Physical Sciences—A Status Report on the National Standard Reference Data Systems", *Natl. Bur. Stand. (U.S.), Tech. Note*, No. 947, 77-600016 (Jan 1977).
(11) N. Kurti, "Capture, Evaluation and Storage of Data—As Seen By CODATA", *CODATA Newslett.*, 8 (Sept 1977).
(12) R. C. Reed, J. M. Prausnitz, and T. K. Sherwood, "The Properties of Gases and Liquids", McGraw-Hill Book Co., New York, 1977.
(13) G. R. Somayajulu and B. J. Zwolinski, "Generalized Treatment of Alkanes—Part 2", *J. Chem. Soc., Faraday Trans.*, **68**, 1971-1987 (1972); "Part 4: Atomic Additivity Applications to Substituted Alkanes", *ibid.*, **70**, 973-993 (1974); "Part 5: Branching and Buttressing Effects", *ibid.*, **72**, 2213-2234 (1976).
(14) G. R. Somayajulu and B. J. Zwolinski, "Generalized Treatment of Aromatic Hydrocarbons. Part 1. Triatomic Additivity Applications to Parent Aromatic Hydrocarbons", *J. Chem. Soc. Faraday Trans. 2*, **70**, 1928-1941 (1974).
(15) S. W. Benson, "Thermochemical Kinetics", 2nd ed., Wiley, New York, 1976, Chapt 11.
(16) A. Fredenslund, G. Gmehling, and P. Rosmussen, "Vapor-Liquid Equilibria Using UNIFAC. A Group Contribution Method", Elsevier, New York, 1977.
(17) Y. Yoneda, "A Proposal of an Estimation and Retrieval System EROICA for Physical Properties of Organic Compounds by Chemo Inputs", in "Information Chemistry", S. Fujiwaya and H. B. Mark, Ed., University of Tokyo Press, Tokyo, 1975, pp 239-253.

Problems in Physical Property Data Retrieval†

DONALD T. HAWKINS

Bell Laboratories, Murray Hill, New Jersey 07974

Received February 11, 1980

Problems in the retrieval of numerical data on physical properties are reviewed. Data compilations are valuable, but often do not exist for the properties of interest. Those that are available are usually arranged by substance, which causes problems when the substance has a variety of names. Finding a substance which fits a given set of properties is not possible with today's handbooks; an "online handbook", containing numerical data on properties of substances and searchable online, would allow such a question to be answered. On another level, two projects which have been undertaken at Bell Laboratories to help staff members obtain needed numerical data are an index to a list of compilations held by the National Bureau of Standards and a pathfinder to numerical values of the properties of silicon.

INTRODUCTION

The scientist or engineer who wishes to obtain and use numerical data faces a monumental challenge. Numeric data are widely scattered in the literature; data compilations and Information Analysis Centers (IAC's) exist only for some fields. This paper discusses some of the problems faced by an information retrieval service in searching for and retrieving numeric data. Although the discussion here is limited to data on physical properties, the general situation is the same for many fields. Experiences related here are those gained at Bell Laboratories, a large and diversified information community; they are by no means atypical.

Bell Laboratories, the research and development unit of the Bell System, has more than 11 000 technical personnel, many of whom have a potential need for numeric physical property data. The numeric data needs of Bell Laboratories are very wide, and many technical staff members are able to satisfy their own needs. One way a scientist or engineer can satisfy this need is to collect the relevant handbooks and data compilations, which is no small task. One metallurgist at Bell Labs, who serves as a consultant in his field, has such a collection. He has acquired handbooks of data on metals and alloys and related materials. His collection comprises 135 volumes–enough to fill several bookshelves! Often, many of these volumes must be consulted when a particular numeric value is wanted. Each volume is organized differently; frequently the values are stated in different units, at varying temperatures; for a wide range of experimental conditions, and so on. The location and comparison of the data are therefore difficult and time-consuming tasks. When data on a particular substance are located, one may find that the parameters under which the value was measured are too far removed from those

† Presented before the Division of Chemical Information, Symposium on "Techniques and Problems in Retrieval of Numerical Data", 178th National Meeting of the American Chemical Society, Washington D.C., Sept 12, 1979.

0095-2338/80/1620-0143$01.00/0

Table I. Some Typical Numeric Data Requests Received by the Bell Laboratories Information Retrieval Service

C_v of NH_3
tensile strength of Be–Cu alloy
dielectric constant of KOH
vapor pressure of Pb, Sn, and some Pb–Sn alloys
lattice parameter and structure of CuCl

Table II. Some Well-Known Handbooks

Handbook of Chemistry and Physics
JANAF Tables
Lange's Handbook of Chemistry
Metals Handbook
CRC Handbook of . . .

desired. For example, a table of heat capacity values for a material may extend only to a relatively low temperature, whereas high-temperature data are needed. An attempted extrapolation may be uncertain at best and at the worst, totally erroneous.

DATA COMPILATIONS

Many technical people do not attempt to maintain their own comprehensive collection of handbooks because of space, economic, or other considerations. Instead, they turn to their technical library or an IAC for assistance. Table I lists five requests for physical property data recently received by the Bell Laboratories Information Retrieval Service. These are typical requests; similar examples could be provided by many libraries and IAC's. Much of the needed data in these cases can be quickly found in handbooks and data compilations *if* the searcher knows where to look and *if* a data compilation exists for the property in question. Table II lists some well-known handbooks and data compilations which may have representative values useful in answering requests like those in Table I. Handbooks often provide citations to more detailed compilations, or to the original literature.

The value of data compilations has long been recognized. Although the compilation and critical evaluation of data are recognized as worthwhile scientific activities, the number of persons engaged in them is small, and funding for such efforts can be difficult to obtain. Much has already been written on this subject; an editorial by Branscomb[1] is noteworthy. Data compilation is a slow and costly effort, and the flow of new data is great. Updating data compilations is therefore infrequently done.

Fulfilling even simple requests such as those shown in Table I often poses problems. Perhaps the foremost problem is that many fields are not covered by compilations. In these cases the searcher has no choice but to turn to the primary literature. Numerical data are widely scattered and frequently of dubious qualtity. Online bibliographic retrieval systems have helped considerably in the location of physical property data, but the existence of the data may not be mentioned in the abstract of a paper, and so the data are not indexed and cannot be retrieved. Data flagging projects[2] can do much to alleviate this problem.

A second problem common to virtually all handbooks and compilations is that the data are arranged by substance. One must know the name of the substance to be able to locate its properties. Finding the name of a substance is not trivial, especially for complex chemicals. IUPAC names, trade names, and generic or trivial names all exist, and a substance may be referred to and indexed under any of these. Table III shows a few of the equivalent names for common cartridge brass. Obviously, if one is looking in a handbook for "cartridge brass", and the compiler of the handbook has listed the properties under "Olin 260" without any cross-reference, the information will not be found. The reverse question, finding

Table III. Equivalent Names for Cartridge Brass

AMS 4505D	Mil-C-50 Shim Stock
ASTM B129, Copper Alloy No. 260	Olin Alloy 260
Cartridge Brass	SAE CA260
CDA No. 260	SAE 70B
Cu30Zn70	70 30 Brass
Mil-C-10375, Cartridge Case	

Table IV. Request for Material with Specific Properties

Find a chemical with the following properties:
high heat of fusion
high density
unreactive with common container materials
relatively cheap
melting point near 120 °F

Table V. Some Numeric Online Retrieval Systems of Interest to Chemists

NIH-EPA Chemical Information System
Sadtler Infrared Spectral System
PBM/STIRS
ManLabs-NPL Materials Data Bank

a substance with specified properties, cannot be answered with today's handbooks. Table IV is a example of such a question. Unfortunately, there are no sources available which contain this information, or even some portion of it, in an easily usable form. An "online handbook" (discussed below) is a solution to this problem and is urgently needed.

ONLINE SYSTEMS

Fortunately, the picture is not all dark. Some numeric online databases are beginning to appear, although the situation is not as advanced as with bibliographic databases. Table V lists some online systems of interest to chemists. The NIH–EPA Chemical Information System[3] is becoming more widely known, and its usage is growing. It contains textual and numeric data on several properties, such as mass spectra, ^{13}C NMR, X-ray powder diffraction, and toxicology. Searches can be made both for properties (i.e., spectra) of known substances or for unknown substances to fit a known spectrum. A Structure and Nomenclature Search System is also available. The components (databases) of the system are linked by the CAS Registry Number. The Sadtler system is similar for infrared spectra. PBM/STIRS contains over 41 000 spectra from the "Registry of Mass Spectral Data".[4] The ManLabs system[5] contains thermochemical data on about 1800 inorganic substances. Users can obtain data for chemical reactions involving these substances. A noteworthy and useful feature of the system is that the user's own data may be used, either alone or in combination with the system. Binary and ternary phase diagrams are also available.

BELL LABS IN-HOUSE PROJECTS

The remainder of this paper describes two in-house projects undertaken at Bell Laboratories to help the Information Retrieval Service in retrieval of physical property data. The first is an index to the Annotated List of Data Compilations[6] published by the National Bureau of Standards. This publication is a list of the data compilations in the library of the Office of Standard Reference Data. The list, though valuable, is difficult to use because it is organized only according to broad subject areas and contains no index. We obtained a machine-readable copy of this publication and produced a KWIC index of the titles of the listed compilations. The usefulness of NBS's publication was thus greatly enhanced; it has proven valuable in solving some of our data requests. (The publication is now 9 years old; an update or reissue, with index, would be most welcome.)

PROPERTY	QUALIFICATIONS	SOURCE	PAGE
Magnetic Moment	atomic, Si^{+} to Si^{11+}	Gm-15B	18
Magnetic Susceptibility		Gm-15B	92-3
	-170 to 1000 °C	ICT-VI	355
	temperature, doping effect	EPIC162	178-82
Magnetoresistance	n-,p-type, anisotropy, temperature effect, saturation	EPIC162	170-7
Mask Opening	impurity profiles	W	404
	junction depth	W	394,398
Mask Edge	see contours		
Mass	see effective mass		
Mass Attenuation	see gamma; see x-ray		
Mean Free Path	As, B doped	EPIC120	na
Melting	see volume change		
		ICT	I-104
	EVALUATED DATA	HDHE	467
		CRC56	15
	with defined impurities	Gm-15B	67
Metal-Semiconductor Contact Potential	vs impurity conc, Al, Au; n-,p-type	W	306
Metal-Semiconductor Interface Fermi Level	vs energy gap, various semiconductors	W	312
Microwave Fluorescence	detection limits	CRC55	839
Microwave Conductivity	see dielectric constant; see resistivity		

Figure 1. Typical page from "Pathfinder to Numerical Values of the Properties of Silicon".

CRC27 = CRC Handbk of Tables for Applied Engineering Science, Chemical Rubber Co, 1970 (620.3/H23:2) 118500
CRC33 = CRC Handbk of Lasers, Chemical Rubber Co, 1971 (621.36602/C51) 151417
CRC55 = CRC Handbk of Spectroscopy, v. 1 Chemical Rubber Co, 1974 (535.84083/C11) 119708
CRC56 = CRC Handbk of Spectroscopy, v. 2 Chemical Rubber Co, 1974 (535.84083/C11) 126505
DD = Diffusion & Defect Data (Shelved as a Periodical, covered to vol. 14, 1977)
EPIC9 = Electronic Properties Information Center, DS-9, Epitaxial Films, 1968 (620.1129/E38e3) 144292
EPIC117 = Electronic Properties Information Center, DS-117, Optical Absorption, 1962 (620.1129/E38s4) 143017
EPIC118 = Electronic Properties Information Center, DS-118, Debye Temperature, 1963 (620.1129/E38s8) 143020
EPIC119 = Electronic Properties Information Center, DS-119, Dielectric Constant, 1963 (620.1129/E38s9) 143019
EPIC120 = Electronic Properties Information Center, DS-120, Mean Free Path, 1963 (620.1129/E38s2) 143021
EPIC126 = Electronic Properties Information Center, DS-126, Electrical Conductivity, 1963 (620.1129/E38s10) 143018
EPIC162 = Electronic Properties Information Center, DS-162, Silicon, 1969 (620.1129/E38s2) 141289
Gm-15B = Gmelin syst # 15, part B (1959) (R 546/G56-8)
HDHE = Hultgren RR,Desai PD,Hawkins DT.... Selected Values of Thermodyn..., Elements, 1973, (546.3/H91.2) 121621
HDHA = Hultgren RR,Desai PD,Hawkins DT.... Selected Values of Thermodyn..., Alloys, 1973, (546.3/H91.2) 121622
ICT-I thru VI = International Critical Tables of Numerical Data, v. I to VI 1926-33, (R 530.08/N27)
LBn3/1 = Landolt-Bornstein gr. 3, v. 1 (new series 1966) (R530.08/L25N)
LBn3/2 = Landolt-Bornstein gr. 3, v. 2 (new series 1966) (R530.08/L25N)
LBn3/8 = Landolt-Bornstein gr. 3, v. 8 (new series 1966) (R530.08/L25N)
LBn3/6 = Landolt-Bornstein gr. 3, v. 6 (new series 1966) (R530.08/L25N)
LBo1/1 = Landolt-Bornstein bd 1, t. 1 (old series 1950) (R530.08/L25:6)
LBo1/4 = Landolt-Bornstein bd 1, t. 4 (old series 1950) (R530.08/L25:6)
LBo2/1 = Landolt-Bornstein bd 2, t. 1 (old series 1950) (R530.08/L25:6)
LBo2/8 = Landolt-Bornstein bd 2, t. 8 (old series 1950) (R530.08/L25:6)

Figure 2. Partial list of sources from "Pathfinder to Numerical Values of the Properties of Silicon".

More recently, a "Pathfinder to Numerical Values of the Properties of Silicon" has been prepared. A sample page from this document is shown in Figure 1. A search was made for properties of silicon in sources held by the Bell Laboratories Library Network (and hence readily available to Bell Labs personnel). The properties were listed alphabetically in tabular form, along with information about the data, such as temperature range. Numerous cross references to the properties were added to enhance the usefulness of the Pathfinder. Abbreviated information about the sources of the data is given in the table; detailed references, including the library call number, are listed at the end of the Pathfinder. Figure 2 shows a portion of the list of sources.

The Pathfinder was conceived and developed as an experiment because library users are generally unaware of all the possible resources. For example, the Landolt–Bornstein compilations are not well-known to engineers; Gmelin and Beilstein are relatively unknown outside the chemical world. Large compilations such as these are difficult and time-consuming to search for first-time or occasional users. As the number of sources grows, it becomes difficult, if not impossible, to remember them. Hence, the Pathfinder was developed to help lead data users to potential sources. Silicon was chosen as the first material to be covered because of its importance in integrated electronics; other materials will no doubt be added. The silicon Pathfinder was well received by the Bell Labs technical staff. We estimate that two person-weeks were required for the collection and organization of the data. The collection time will probably vary widely from one substance to another. For silicon, for example, a large number of the entries came from a single source.

ONLINE HANDBOOK

Many of our numerical data needs would be satisfied by an "online handbook". Such a handbook does not exist now, but the technology to support it does. One can envision the entries from many printed handbooks and data compilations being collected into an online file and made searchable using software available today. Since the data base would be an inverted file containing both data and substances, properties would be searchable. Standard Boolean operations would allow several to be specified. Questions such as the one shown in Table IV could then be easily answered. Some provision would have to be made for dealing with duplicate and conflicting data and also for maintenance of the data base. An online handbook would be immensely useful and would be heavily used, not only at Bell Labs, but at many other institutions as well. Pestel and Rubin[7] discuss some of the advantages of an online handbook. They also rightly point out that the cost of producing it is high, and substantial funding or subsidies will probably be needed in order to bring the online handbook into reality. A small pilot project in developing and constructing such a handbook is currently underway at INSPEC, the publishers of *Science Abstracts*. We welcome such efforts and look forward to their completion.

ACKNOWLEDGMENT

The Pathfinder on silicon was prepared by my colleague, Bernard A. Stevens.

REFERENCES AND NOTES

(1) Branscomb, L. M. "Support for Reviews and Data Evaluation", *Science* **1975**, *187*, 603.
(2) Murdock, J. W. "Numerical Data Indexing", *J. Chem. Info. Comput. Sci.* **1980**, accompanying paper in this issue.
(3) Heller, S. R.; Milne, G. W. A.; Feldman, R. J. "A Computer-Based Chemical Informaion System", *Science* **1977**, *195*, 253–9.
(4) Stenhagen, E., Abrahamsson, S., McLafferty, F. W., Eds. "Registry of Mass Spectral Data"; Wiley: New York, 1974.
(5) Hawkins, D. T. "The MansLabs-NPL Materials Data Bank", *Online* **1979**, *3*(2), 40–52.
(6) Weisman, H. M.; Sherwood, G. B. "Annoted Accession List of Data Compilations of the Office of Standard Reference Data", National Bureau of Standards Technical Note 554; National Bureau of Standards: Washington, D.C., 1970.
(7) Pestel, H. C.; Rubin, S. A. "Electronic Handbook-Fact or Fiction", *Online* **1977**, *1*(4), 84–7.