

Evaluating Experimental Data on Heats of Combustion*

EUGENE S. DOMALSKI
National Bureau of Standards
Washington, D. C. 20234

Received July 16, 1971

The approach used by data evaluators at the Chemical Thermodynamic Data Center of the National Bureau of Standards for the interpretation of published articles reporting values of the heats of combustion is discussed. An appropriate set of corrections is applied to the values when required to bring all data under study to a common base. Corrections which must be considered are: changes in calibration, reduction of the sample mass to vacuum, changes in atomic mass, reduction to standard states, expression of the data in proper energy units, and bringing the reference temperature for the combustion process to 298.15 K. Some corrections must be estimated because insufficient information is given in certain articles for a standard reanalysis. Guidelines are offered to authors of papers which will provide a maximum amount of data for the reexamination of experimental results, but yet maintain conciseness in the over-all presentation.

As part of the task of the Chemical Thermodynamics Data Center at the National Bureau of Standards,¹³ thermochemical and thermodynamic information on inorganic, organic, and organometallic substances is collected, abstracted, evaluated, and stored for future retrieval. A goal of this work is the publication of authoritative compilations and reviews of thermochemical properties and the capability of updating them from time to time.

This paper discusses the techniques used by data evaluators as they interpret published articles reporting values on the heats of combustion of organic substances and apply corrections to the values when necessary to bring all data under consideration to a common base. In addition, guidelines are suggested for authors in the presentation of their experimental combustion data.

Meaningful information on the heats of combustion of organic compounds has been part of the chemical literature for about 100 years and has appeared there in a variety of forms. Basically, two main experimental techniques are employed to obtain the heats of combustion, bomb calorimetry (combustion process at constant volume), and flame or flow calorimetry (combustion process at constant pressure). Flame calorimetry is more applicable to the combustion of gaseous or highly volatile substances, whereas bomb calorimetry is more suited to studying liquids and solids. Bomb calorimetry has been the method used by most of the investigators and, therefore, has delivered most of the data on heats of combustion.

The number of organic compounds for which experimental heats of combustion have been determined has grown from about 600 prior to 1900, to about 1500 by 1930, and to about 4000 at the present time (1970). The need for evaluated data on the heats of combustion of organic compounds is widespread. Applications of combustion data in industry relate to problems such as combustion engine performance, fuel evaluation and rating, and incinerator design. Physical chemists require combustion data along with other thermochemical data to estimate whether certain reactions will proceed, to calculate strain energies in complex molecules, and to establish energy correlations for compounds in a specific class. With respect to a particular property, the determination of the heat of combustion of a substance provides a means of deriving its heat

of formation. Other types of reactions (such as heats of hydrogenation or heats of reaction in solution) are possible sources of obtaining this important thermochemical property; however, combustion measurements account for the vast majority of reliable data on heats of formation.

EVALUATION PROCESS

The evaluation process which has been applied to the experimental combustion data consists of three essential parts: examination and appraisal of the data, reanalysis and recalculation of the data, and selection of a "best" value.

During the examination and appraisal of the combustion data, the evaluator seeks answers to questions of the following type: Did the sample have sufficient purity? Had the sample been characterized well analytically? Were the authors aware of the proper procedures, corrections, and constants? Were enough experiments performed to allow a statistical analysis? How was the amount of reaction determined? Was the sample likely to have undergone premature reaction prior to the desired time of ignition? How are the energies and atomic weights used related to the current units?

At this point the evaluator goes not to the reported final values, but to the raw experimental data to begin reanalysis and recalculation. These raw data should be as close as possible to the experimentally observed parameters and should be independent of any theoretical assumptions. If necessary the evaluator applies an appropriate set of corrections to the experimental data to reduce them to a common base. This base is a reaction in which reactants and products are in their thermodynamic standard states at 298.15 K. Numerous experiments have been reported at other conditions. Numerical quantities are recalculated on the basis of the current values of the physical constants and units of measurement.

After all measurements for a given compound have been so treated, the evaluator must make a decision as to which calculated heat of combustion is best. The selection is the evaluator's personal and professional opinion of what the most probable value is, based upon his examination and reanalysis of the data. A limited presentation of the combustion data, or lack of adequate discussion often leaves persistent doubts as to the quality of the study. A well-documented presentation of an investigation answers ques-

*Presented at the 6th ACS Middle Atlantic Regional Meeting, ACS, Baltimore, Md., February 3, 1971.

tions about the procedures or about the treatment of the results which most thermochemists would raise. The final value is placed upon a firmer footing because sufficient detail has been supplied. The consistency of an experimental value with those of a similar class of compounds for which the values or the trend in the values has been established is also an influential factor. A poorly-documented study makes evaluation difficult or impossible and renders the data reported of dubious value.

A more detailed look at the extent and degree to which the evaluator must reanalyze and recalculate the experimental values under study seems appropriate here and can be accomplished by elaborating upon the corrections which are applied to the combustion data when required.

CORRECTIONS TO THE COMBUSTION DATA

In the process of bringing the combustion data to a common base, the evaluator must apply the following corrections as appropriate.

Calibration Correction. Applying a calibration correction to bomb calorimetric data consists of examining the value obtained or used for the energy of combustion of benzoic acid by an investigator and comparing it to the current value. If the investigator's value differs significantly from the current value, it is multiplied by a ratio of the appropriate value to that used by the investigator.

At the National Bureau of Standards samples of benzoic acid are purified, the heat of combustion determined, and the combustion value certified for calibrating bomb calorimeters¹⁵ under the following conditions: the combustion reaction is referred to 25°C; the number of grams of sample burned is equal to three times the volume of the bomb in liters; the number of grams of water placed in the bomb before combustion is equal to three times the volume of the bomb in liters. No unnecessary departure should be made from the above certified conditions. The current value for the energy of combustion of benzoic acid is: $-\Delta U_R = 26,434 \pm 3 \text{ Jg}^{-1}$ (weight in vacuum).

If the calibration was performed electrically, the evaluator examines the paper to see that the proper standards of resistance, electromotive force, and time were used. If the calibration was based on a well-established reaction, the evaluator examines the reaction heat to see that it is still current. In the case of flame calorimeters, calibration is usually performed by electrical measurement or by a well-established gaseous reaction such as $\text{H}_2(\text{g}) + 0.5 \text{ O}_2(\text{g}) = \text{H}_2\text{O}(\text{l})$ ^{11, 19}.

Needless to say, if only cursory information is given regarding the calibration methods, the evaluator is limited in the amount of reappraisal that can be performed.

Vacuum Correction. In the early literature, such as that of Berthelot,¹ Stohmann,²⁹ and even some more recent papers, the combustion data presented are for the samples weighed in air. From estimates of the densities of the sample, of air (assuming a reasonable temperature, barometric pressure, and relative humidity), and the weights, a buoyancy correction is applied to reduce the weight of the sample to that in vacuum.

Molecular Weight Correction. The amount of the combustion reaction is often expressed on the basis of the number of moles of substance burned. An appreciable correction may be required because of differences in the presently accepted values of atomic masses and those reported with the combustion data. Occasionally the amount of reaction is based on a combustion product, such as the CO_2 ²¹ or H_2O ²⁴ formed in the reaction which is collected in an absorbent after the calorimetric experiments. In this

case the ratio of the weight of sample to weight of CO_2 or H_2O depends on the atomic weights. Current values of the atomic masses are based on the isotope $^{12}\text{C} = 12$.²

Washburn Reduction. Washburn³¹ suggested a need for a correction dealing with the reduction of combustion data from bomb conditions (certain mass of sample, pressure of oxygen, bomb volume, and temperature) to conditions in which the reactants and products of the combustion reaction are in the pure state at 1 atm. pressure. Subsequent developments and improvements to final conditions of thermodynamic standard states (CO_2 as an ideal gas, H_2O as the pure liquid) were made by Hubbard, Scott, and Waddington,^{6, 7} and Prosen.²⁰

Conversion to Desired Thermal Quantities. The conversion of bomb combustion data from the process at constant volume to that at constant pressure is made by means of the thermodynamic equation: $\Delta H^\circ = \Delta U^\circ + \Delta nRT$, in which ΔH° is the standard (the superscript ° denotes the standard state) molar enthalpy change at constant pressure, ΔU° is the standard change in internal energy at constant volume, Δn is the difference between the total number of moles of gaseous products and reactants involved in the combustion reaction, R is the gas constant, and T is the Kelvin temperature of the reaction process.

Sometimes, the terms gross and net heats of combustion are encountered. The gross heat of combustion refers to the reaction in which the water formed as a product is in the liquid state while net heat of combustion implies the water is in the gaseous state. Our practice is to report heats of combustion to form water in the liquid state.

Temperature Correction. All combustion data are reduced to the reference temperature of 298.15 K (25°C). The appropriate change in heat capacity, ΔC_p , for the combustion reaction is calculated from existing data, or estimated for the temperature interval under consideration by some of the available methods.^{10, 22, 23, 27}

Uncertainty of the Combustion Data. One of the most valuable guides that the investigator can provide for the evaluator and general reader is a statement of the uncertainty in the measured combustion data. Although this has been recognized for some time, the practice is only recently becoming widespread. Rossini^{25, 26} has discussed the assignment of uncertainties to thermochemical data and uses twice the over-all standard deviation of the mean as a convenient estimate of the uncertainty interval. This gives approximately a 95% confidence limit. This is a very useful guide and has been adopted in our evaluation processing as much as possible. Systematic errors, such as applying an incorrect energy equivalent, or choosing an improper parameter to decide the amount of combustion, are also possible. If the systematic error can be unambiguously estimated, the data are corrected and no increase in the uncertainty is made. On the other hand, if the systematic error is present but not easily defined, the over-all uncertainty is increased as deemed necessary.

Calculation of the Heat of Formation. In calculating the standard heat of formation from its heat of combustion, the values and sources of the auxiliary data should be specified. If any ambiguity is possible with regard to the reference states of the elements, an explanation should be provided. Unless obviously more accurate data are available, "Selected Values of Chemical Thermodynamic Properties" as compiled and tabulated by Wagman *et al.*³⁰ is recommended as the source for auxiliary data.

Expression of Energy Units. The NBS supports and promotes the use of the International System of Units (designated SI units or Le Système International d'Unités)^{16, 17} as the international language of science and technology and urges calorimetrists, and scientists in general, to express their data in these units. The United States

of America as a signatory to the Treaty of the Meter is under obligation to support the Actions of the General Conference on Weights and Measures, which includes promotion of SI units. The unit of energy in SI units, and hence, in modern calorimetry, is the joule. The SI unit for the amount of substance is the mole.

NBS policy^{17, 18} on the usage of SI units makes allowances for the effective availability of information to scientists in disciplines in which the adoption and usage of SI units is apt to be slow or difficult. The NBS supports the use of SI units, but does not wish to impair scientific communication or reduce its usefulness. Parallel citation of numerical data in SI units and those more familiar to a particular audience is encouraged.

Early combustion data were expressed in calories which were based on and defined in terms of the heat capacity of water. Starting about 1930, the artificial thermochemical calorie was used for reporting thermochemical data, and was defined as: 1 calorie = 4.1833 international joules. Beginning January 1, 1948, the calorie was redefined in terms of absolute joules as: 1 calorie = 4.1840 absolute joules. With the passage of time, the adjective absolute has become unnecessary and has been dropped from usage; and joules are understood to be absolute joules. The engineer is perhaps more acquainted with the IT calorie (International Steam Tables Calorie) which is defined as 4.1868 joules.

In many respects the evaluator of experimental combustion data must be an expert in the historical development of combustion calorimetry. Being able to identify the state-of-the-art for a particular period is a valuable guide in understanding the limits of procedures and the calculations applied. For example, a significant breakthrough in calibrating calorimeters occurred in the 1900's^{8, 9} when the old method of summing the products of the masses and specific heats of the corresponding parts of the calorimetric system was replaced with electrical measurements of energy input to a calorimeter. Another important event took place in the 1920's when international agreement was reached on a chemical standard for calibrating bomb calorimeters, benzoic acid.

RECOMMENDATIONS TO AUTHORS

Several attempts have already been made to acquaint the experimental thermochemist with what information is needed in the report of his investigations so that it will be amenable to reanalysis in the future if necessary. In 1953, the 8th Calorimetry Conference (U.S.A.) adopted a resolution outlining minimum publication standards for the guidance of authors, editors, and referees of calorimetric papers. The resolution had very beneficial effects upon the quality of such papers. The resolution was revised, expanded, and adopted by the 15th Calorimetry Conference.¹⁴ A similar document was published at the request of the Second All-Union Conference on Calorimetry (U.S.S.R.), and was approved by the Scientific Council on Chemical Thermodynamics of the U.S.S.R. Academy of Sciences.²⁸ Evans and Garvin⁵ describe the needs of the evaluator of quantitative chemical data and offer some suggestions for improving the exposition of results by authors.

We present here recommendations oriented specifically to researchers in combustion calorimetry not only to improve the presentation of their results but also to improve the lot of the data evaluator. Authors have a responsibility of making sure their results are amenable to recalculation if different values for various physical constants or units of measurement are adopted. By means of a satisfactory presentation of their experimental data, they can more readily ensure its long-term value.

The presentation of information can be divided into four parts: substance(s) under study, apparatus and procedures, experimental combustion data, and derived values.

Substances under Study. In this section the source or method of synthesis and purification should be given along with pertinent data which completely characterize the sample(s). Any other property data needed for calculations, such as density, heat capacity at the temperature of the combustion reaction, $(\partial E/\partial P)_T$, etc. should similarly be given.

Apparatus and Procedures. A complete description of the apparatus and procedures should be provided in this section including details of the method of calibration, type of combustion bomb or reaction vessel, the calorimeter proper and its environment, the temperature and time measuring equipment, and the sensitivities of auxiliary instruments such as power supplies, potentiometers, bridges, flow meters, and temperature regulators. If a chemical standard is used for calibration, it should be identified clearly with respect to source, purity, and thermal value. The results of the calibration experiments should be presented in this section citing the number of determinations performed, the average value of the energy equivalent and the standard deviation of the mean. Similarly, the results should be reported here on combustion experiments on fuse materials (cotton, chromel, tungsten), and substances used to dilute the sample (benzoic acid, Teflon powder, diethyl phthalate) in order to provide a more complete combustion. Procedures used to protect the samples prior to ignition in bomb calorimetry, such as placing the sample in a glass ampoule or in a plastic bag should be explained.

Experimental Combustion Data. This section should present the combustion data as closely as is practicable to the actual numerical information recorded during the experiment. The presentation should be as concise as possible but yet possess sequential logic and clarity. The description of the initial combustion conditions—such as, the volume of the bomb, pressure of the oxidant, quantity of distilled water added to the bomb prior to an experiment, and flow rates of gases—is appropriate here. Providing the reader with the chemical equation denoting the combustion process under study is most valuable. For example: $C_9H_{20}(l) + 19O_2(g) = 9CO_2(g) + 10H_2O(l)$; this leaves no doubt as to what process is being reported. The tests or methods used to examine the products of combustion and determine the amount of reaction should also be provided.

The following sequence of data is recommended for the exposition of a calorimetric measurement: (a) the mass of sample reduced to weight in vacuum undergoing combustion. If a correction for unburned sample was applied this should be explained. Alternately, if the mass of the combustion product was used in determining the amount of reaction, then that mass should be given; (b) the mass of auxiliary substances taking part in the combustion, such as fuse materials or sample diluents; (c) energy equivalent (heat capacity of the calorimeter) for the particular experiment; (d) the corrected temperature rise; (e) total observed energy change for the experiment; (f) energy contributions for various processes such as chemical and electrical fuse energy, energy of combustion of sample diluents, or of materials encapsulating the sample, energy changes owing to the flow of gases, energy of formation of aqueous nitric acid, energy of combustion of impurities in the sample; (g) final experimental energy of combustion under bomb conditions; (h) application of the Washburn correction reducing the energy of combustion to standard conditions; (i) conversion of the energy of combustion from a mass to a molar value. If the process under study was performed in a bomb at constant volume, a correction (ΔnRT) should

be applied to ΔU° to convert it to constant pressure conditions, ΔH° .

At this state no one is in a better position to assess the imprecision and inaccuracy of the combustion process studied than the investigator himself. As a minimum, authors should provide the number of experiments performed and the corresponding combustion values, the average of these values, and the standard deviation of the mean. A more desirable statistic is the over-all standard deviation of the mean which has, combined with the standard deviation of the mean of the combustion experiments, the standard deviations of the means of the calibration experiments and, when applicable, the standard deviations of the means of combustion experiments on sample diluents. The combination of these random errors can be made by summing the squares of the individual sets of random error, and extracting the square root. A convenient estimate of the uncertainty interval is then obtained by taking twice the over-all standard deviation of the mean (over-all precision). If the number of measurements is small (six or seven), then the appropriate factor in the Student *t* distribution is suggested as a better estimate of the random error limits. In either case, authors should clearly state the statistical path they have chosen.

An assessment of systematic errors should be provided, especially if the combustion process is a novel or complex one (see reference 3). We encourage authors to give as much detail as possible showing how the limits for each identifiable error are arrived at and what evidence they have to support their estimate. Any consistency checks to indicate the absence of certain possible errors would also be useful. After examining the over-all precision and the magnitude and signs of the systematic errors, an estimate of the inaccuracy should be provided.

Eisenhart⁴ furnishes some good insight into the expression of the uncertainties of final results. We urge authors to examine his summary prior to presenting an error analysis of their results. According to Eisenhart,⁴ four distinct cases need to be recognized in order to state the credible limits of the likely inaccuracy of a reported value. The cases examine the presence or absence of the systematic error and imprecision in the reported result. Another summary by Ku¹² discusses expressions of imprecision, systematic error, and the uncertainty associated with a reported value, and is presented in tabular form. From an examination of these tables, authors can select the most suitable statements which are designed to convey the message they wish to make about errors in their experimental work.

Derived Values. The heat of formation should be calculated from the heat of combustion and the sum of the heats of formation of the combustion products. The calculation of related thermodynamic properties, such as the Gibbs energy of formation, the heat of formation in the gas state, or a particular average bond energy, by combining the appropriate data with the calculated heats of formation may be desirable and should be decided upon by the authors. The sources of all auxiliary values used in such calculations must be given.

LITERATURE CITED

- Berthelot, M., and Vieille, P., "Sur les Chaleurs de Combustion et de Formation des Carbures d'Hydrogene Solides," *Ann. chim. et phys.* [6] 10, 433-55 (1887).
- Cameron, A. E., and Wichers, E., "Report of the International Commission on Atomic Weights (1961)," *J. Amer. Chem. Soc.* 84, 4175-97 (1962).
- Churney, K. L., and Armstrong, G. T., "The Heat of Combustion of Beryllium in Fluorine," *J. Research NBS* 73A, 281-97 (1969).
- Eisenhart, C., "Expression of the Uncertainties of Final Results," *Science* 160, 1201-4 (1968); also reprinted in NBS Special Publ. 300, Vol. 1, Chap. 1.6, pp. 69-72, "Precision Measurements and Calibration, Statistical Concepts and Procedures," H. H. Ku, Ed., U.S. Government Printing Office, Washington, D. C. 20402, Feb. 1969.
- Evans, W. H., and Garvin, D., "The Evaluator Versus the Chemical Literature," *J. Chem. Doc.* 10, 147-50 (1970).
- Hubbard, W. N., Scott, D. W., and Waddington, G., "Reduction to Standard States (at 25°) of Bomb Calorimetric Data for Compounds of Carbon, Hydrogen, Oxygen and Sulfur," *J. Phys. Chem.* 58, 152-62 (1954).
- Hubbard, W. N., Scott, D. W., and Waddington, G., Chap. 5, pp. 75-128, in "Experimental Thermochemistry," Vol. 1, F. D. Rossini, Ed., Interscience, New York, N. Y., 1956.
- Jaeger, W., and Steinwehr, H. von, "Bestimmung des Wasserwertes eines Berthelotschen Kalorimeters in elektrischen Einheiten," *Verhandl. Deut. Physik. Ges.* 5, 50-9 (1903).
- Jaeger, W., and Steinwehr, H. von, "Eichung eines Berthelotschen Verbrennungskalorimeters in elektrischen Einheiten mittels des Platinthermometers," *Ann. Physik* [4] 21, 23-63 (1906).
- Janz, G. J., "Estimation of Thermodynamic Properties of Organic Compounds," Revised Edition, Academic Press, Inc., New York, N. Y., 1967.
- King, R. C., and Armstrong, G. T., "Fluorine Flame Calorimetry. III. The Heat of Formation of Chlorine Trifluoride at 298.15 K," *J. Research NBS* 74A, 769-79 (1970).
- Ku, H. H., "Expression of Imprecision, Systematic Error, and Uncertainty Associated with a Reported Value," *Measurements & Data* 2, 72-7 (1968); also reprinted in NBS Special Publ. 300, Vol. 1, Chap. 1.7, pp. 73-78, "Precision Measurements and Calibration, Statistical Concepts and Procedures," H. H. Ku, Ed., U. S. Government Printing Office, Washington, D. C. 20402, Feb. 1969.
- Lide, D. R., Jr., "Critical Evaluation of Data in the Physical Sciences—A Status Report on the National Standard Reference Data System," NBS Tech. Note 553, Sept. 1970.
- McCullough, J. P., Westrum, E. F., Jr., and Evans, W. H., "Calorimetry Group Adopts Revised Resolution on Data Publication," *Science* 132, 1658-9 (1960).
- NBS Certificate of Analysis, Standard Reference Material 39i, Benzoic Acid, Calorimetric Standard (July 15, 1968, Washington, D. C. 20234).
- NBS Tech. News Bull., "NBS Adopts International System of Units," 48, No. 4, 61-2 (April 1964).
- NBS Tech. News Bull., "NBS Interprets Policy on SI Units," 52, No. 6, 121-4 (June 1968).
- NBS Tech. News Bull., "Policy for NBS Usage of SI Units," 55, No. 1, 18-20 (January 1971).
- Pilcher, G., and Chadwick, J. D. M., "Measurements of Heats of Combustion by Flame Calorimetry Part 4. n-Pentane, isoPentane, neoPentane," *Trans. Faraday Soc.* 63, 2357-61 (1967).
- Prosen, E. J., Chap. 6, pp. 129-48, in "Experimental Thermochemistry," Vol. 1, F. D. Rossini, Ed., Interscience, New York, N. Y., 1956.
- Prosen, E. J., and Rossini, F. D., "Heats of Combustion of Eight Normal Paraffin Hydrocarbons in the Liquid State," *J. Research NBS* 33, 255-72 (1944).
- Reid, R. C., and Sherwood, T. K., "The Properties of Gases and Liquids: Their Estimation and Correlation," Second Ed., McGraw-Hill, New York, N. Y., 1966.
- Rihiani, D. N., and Doraiswamy, L. K., "Estimation of Heat Capacity of Organic Compounds from Group Contributions," *Ind. Eng. Chem. Fundam.* 4, 17-21 (1965).
- Rossini, F. D., "Calorimetric Determination of the Heats of Combustion of Ethane, Propane, Normal Butane and Normal Pentane," *J. Research NBS* 12, 735-50 (1934).
- Rossini, F. D., Chap. 14, pp. 297-320 in "Experimental

- Thermochemistry," Vol. 1, F. D. Rossini, Ed., Interscience, New York, N. Y., 1956.
- (26) Rossini, F. D., and Deming, W. E., "The Assignment of Uncertainties to the Data of Chemistry and Physics, with Specific Recommendations for Thermochemistry," *J. Wash. Acad. Sci.* **29**, 416-41 (1939).
- (27) Shaw, R., "Heat Capacity of Liquids. Estimation of Heat Capacity at Constant Pressure and 25°C using Additivity Rules," *J. Chem. Eng. Data* **14**, 461-5 (1969).
- (28) Sokolov, A. V., Kolesov, V. P., and Vorob'ev, A. F., "From the Scientific Council on Chemical Thermodynamics at the Institute of General and Inorganic Chemistry of the U.S.S.R. Academy of Sciences. Recommendations concerning the Publication of the Results of Calorimetric Measurements," *Russ. J. Phys. Chem.* **39**, 693-4 (1965).
- (29) Stohmann, F., Kleber, C., and Langbein, H., "Ueber den Wärmewerth von Carbonsäuren der aromatischen Reihe," *J. Prakt. Chem.* [2] **40**, 128-57 (1889).
- (30) Wagman, D. D., Evans, W. H., Parker, V. B., Halow, I., Bailey, S. M. and Schumm, R. H., "Selected Values of Chemical Thermodynamic Properties," *NBS Tech. Note* **270-3**, January 1968.
- (31) Washburn, E. W., "Standard States for Bomb Calorimetry," *J. Research NBS* **10**, 525-58 (1933).

Multi-Level Retrieval Systems

IV. Large Systems*

LEE N. STARKER,** K. C. OWEN, and J. W. MARTIN, JR.
Science Information Services Department,
Warner-Lambert Research Institute, Morris Plains, N. J.

Received June 28, 1971

The computer phase of a Multi-Level Retrieval System designed for handling large data collections is discussed. This is an outgrowth of the techniques that were described for earlier systems where the data load was not heavy and/or where the system was queried on an infrequent basis. In all instances, the same data base has been carried along from one phase to the next. The system has been implemented with a minimum of cost and effort, as it utilized Inquire, a general information retrieval package of programs that had been purchased for a different application, but has also been readily applicable in this case.

In earlier papers of this series,^{1, 2, 3} we described the first two steps in the development of a comprehensive information retrieval system. The design of this system was such that the mechanical aspects of the retrieval hardware could be applied to the problem of the moment, without fear that in a very short time the application would be outdated. When this occurred in the past, it often became necessary to redesign and restructure the system, and consequently, to modify and recreate the data base.

Designers are often accused of using an "atom bomb to kill a fly," when a flyswatter would have been more than adequate, because they tend to overdesign so as to avoid the need to rework the system in the immediate future. On the other hand, such overdesign may be justified if it is felt that there is a good chance that file size and/or usage of the system will grow with sufficient rapidity to warrant a more sophisticated approach. This is particularly so when it becomes difficult and expensive to reformat and reprocess the file for each new level of activity.

The Multi-Level approach that we have developed obviates many of the reasons for starting a simple system in a complex environment. This is possible because we have adopted, as a basis for this technique, keyword indices as the method for retrieval from our data bases. Keyword indices are of course readily adaptable to many manual, semi-manual, and automatic systems, and thus present a most

desirable index format, when maximum versatility is required. The same data bank can then be used as we progress from the simplest to the most complex hardware. It will also be relatively independent of the hardware aspects of the system.

Thus the variable-field visual collation (peek-a-boo)¹ card (Figure 1), based on a modified IBM card, served to initiate new systems that were either small in size or to which infrequent reference was made. Use of this card was predicated on a do-it-yourself basis, but information-center assistance was often desirable. As these peek-a-boo systems grow they can be extended by the use of additional decks of cards (one new deck for each 500 documents). Each search of the file must now be carried through all of the decks, and when frequent referral is necessary, this procedure can become quite tedious.

At this stage of growth then, consideration can be given to conversion to a Termatrix system.² Since the J-400 Termatrix drill can use IBM cards as input data, it was only necessary to convert our original peek-a-boo cards to term/document IBM card formats. This was done by computer (Figure 2), and the resultant deck was used to generate a Termatrix system that corresponded to the earlier peek-a-boo cards, but eliminated many of the objections that had arisen.

We have used both facets of this Multi-Level System in our Library as a means of maintaining control over collections of documents dealing with our Warner-Chilcott Division's products. A separate file is maintained for each product and several of them have moved through the steps that we have presented.

*Presented before the Division of Chemical Literature, 161st Meeting, ACS, Los Angeles, Calif., March 30, 1971.

**To whom correspondence should be addressed.