Development and Use of Numeric Databases for Properties of Metastable Chemical Species in Solution

W. P. HELMAN,* G. L. HUG,* IAN CARMICHAEL, and A. B. ROSS

Radiation Chemistry Data Center, Radiation Laboratory, University of Notre Dame, Notre Dame, Indiana 46556

Received December 20, 1985

A description is given of the design, structure, and use of numeric databases for property data derived from photochemistry, photophysics, and radiation chemistry. The databases are considered at various levels of abstraction, with most of the emphasis on the physical scheme. Also, various user views of the databases are given. Two such views are highlighted. They are the use of the numeric databases, first, for the on-line retrieval of data and, second, for the preparation of tables of evaluated data for publication. Many of the programs comprising the database management system are also described.

INTRODUCTION

There has long been interest in data on chemical species that are metastable. The data of interest concern both the properties of the metastable species itself and also the properties relating to the reactivity of the species. There are many ways of producing metastable chemical species, but ionizing radiation and light are so common that the study of these species has become an integral part of the fields of radiation chemistry[1,2] and photochemistry.[3,4] Critical evaluation of the data on the reactive species from radiation is complicated by the short time scales[5] that are often involved. Thus, there is a great need to carefully evaluate the data for fundamental research in these disciplines. In addition, there have been of late many applications of these fields that often require either fundamental parameters[6-8] or data that can be used in modeling schemes.[9,10] Although much work has been done in these two disciplines toward compiling and evaluating data on excited states[11] and radicals,[12] no work to date has been done to make this information available in computer-readable form. One purpose of the work described in this paper is to begin to remedy this situation.

In order to collect and evaluate data on reactive intermediates formed by ionizing radiation, the Radiation Chemistry Data Center (RCDC) was started in 1965.[13] The RCDC is associated with the Standard Reference Data Program of the National Bureau of Standards, Office of Standard Reference Data.[14] In 1977, the RCDC expanded its scope to include data on photochemical intermediates, including excited states.[15] Earlier data evaluation projects of the RCDC resulted in publications where tabulated data were entered into magnetically readable files suitable for computer-aided typesetting.[16,17] It was felt that the process would be more flexible if the data were contained in a database[18,19] that could be used either to make tables for publication[20,21] or to serve as a base for on-line searches of the numeric data. The design and use of these databases is the subject of this paper.

VIEWS OF THE RCDC DATABASE

On one level of abstraction, the databases at the RCDC can all be considered together as a single database. From the users' outlook, the RCDC database has the *views*[22] shown in Figure 1. How these views are implemented physically will be discussed below, with the exclusion of view 5.

At another level of abstraction, the conceptual scheme[22] of the RCDC database can be described by the entity-relationship model of databases. There are three different *entities* in the RCDC database, and the information associated with each entity is an *attribute* of that entity. The three entities in the RCDC database are *reported measurement*, *chemical species*, and *reference*. The three entities and some of their more important attributes are listed in Table I for future reference. To complete the entity-relationship model of databases, the *relationships* have to be specified. The relationships between the three entities are illustrated in Figure 2, with the notation that a double arrow is the symbol for a one-to-many relationship.

Table I. Entities and Some of Their Attributes in the RCDC Database

entity	attributes
chemical species	registry number
	name
	molecular formula
	synonyms
reference	serial number
	authors
	published source
	keywords
reported measurement	primary observable[a]
	name of species
	reference
	method
	name of solvent
	comments
	standards
	subsidiary observables[b]

[a] Examples of this attribute would be triplet–triplet absorption maxima, or maxima–extinction coefficient pairs, in the TTA database and second-order rate constants of radical reactions in the RATES database. [b] Examples of this attribute would be triplet lifetimes in the TTA database and activation energy in the RATES database.

Each of these three entities is physically housed in separate files. *Reported measurements* is in the numeric databases, *chemical species* is in the RCDC registry file (RCDCreg), and *references* is in the RCDC bibliographic database (RCDCbib). In addition, the numeric databases at the RCDC appear in two separate physical forms. The first physical form is used in association with RCDCreg and RCDCbib to produce tables for publication (view 3 of Figure 1). In the following, each of these databases, and the programs that use these databases, will be described. A schematic drawing (Figure 3) shows the interaction of the databases in the production of tables, indexes, and reference lists. This production scheme is discussed fully in a later section. A second physical form of the numeric databases is used for on-line searching (view 4 of Figure 1) and is described in the last section of this paper. Both forms of the numeric databases can be described by the same conceptual scheme of Table I, but they differ significantly in their physical scheme.

0095-2338/86/1626-0099$01.50/0

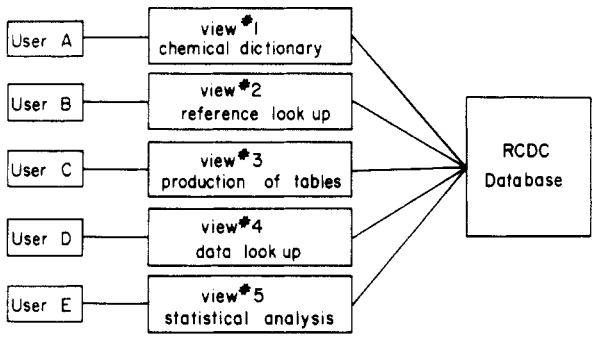

Figure 1. Various views for users of the RCDC database taken as a whole.

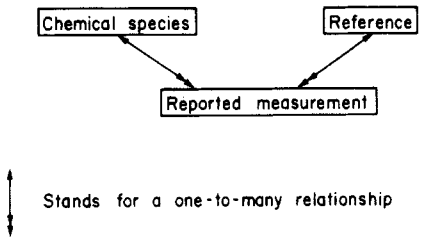

Figure 2. Conceptual scheme of RCDC database: relationship between the entities.

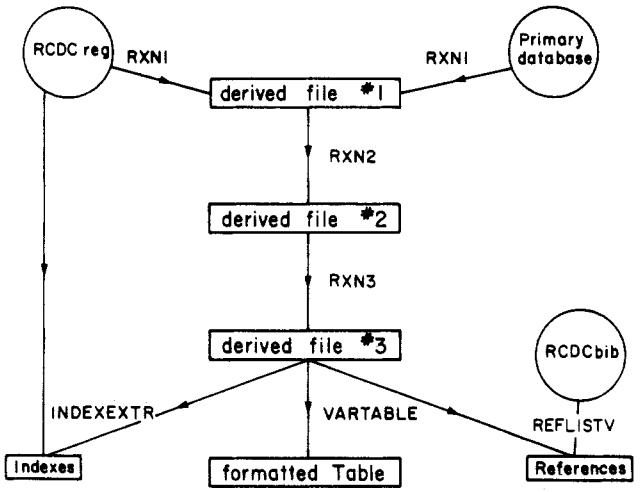

Figure 3. Overview of production of tables of data and their indexes and reference lists from databases. The three principal databases are represented by circles.

NUMERIC DATABASES

Usually, the RCDC numeric databases are initially produced as text files by entering the data items with a VAX EDT text editor. They have also been constructed from the files used previously for computer-aided typesetting of data tables. The text files used for these databases are ordinary sequential files. The lines in these files are to be considered as records of variable length.

The information in these text files is arranged as units to be called *database records*.[23] A database record consists of all of the data and information associated with a reported measurement for a particular chemical or physical quantity. Each database record contains a separate reported measurement, whether it is from a different paper, from a different technique in the same paper, or from a different reaction or compound. The term *database record* is used to distinguish the basic unit of information in these text files from the more usual idea of a physical *record* in the terminology of files.[24] Physically, the database records are a contiguous collection of lines in a sequential file. Each database record is delimited

```
SRN
REF     808090
NM_     Anthracene
NMS     Benzene
MET     LP/ET
EXM     431, 42000 $+ 4000
COM     $*Triplet ET from benzophenone;
        $e relative to benzophenone in benzene
        ($e @ 533^= 7200 L mol ^$-1 @ cm ^$-1 @)
KET     0.49 $v 10^9@
rnr     120-12-7
```

Figure 4. A *database record* from the triplet–triplet absorption (TTA) database. Special characters and fonts are encoded by ASCII character strings $, @, ^, \4, \8, \7, and \9; thus, $e = ϵ, $+ = ±, 10^9@ = 10^9.

by a line beginning with a three-character label, "SRN". Each piece of information within a database record is on a separate line (or a contiguous group of lines) and is tagged by its own three-character label. (Each distinct three-character label stands for an attribute of the entity, *reported measurement*.)

An example of a database record associated with a single measurement of a triplet–triplet absorption is shown in Figure 4. The primary data being compiled are extinction coefficients of triplet–triplet absorption, and a pair of numbers corresponding to a wavelength–extinction coefficient pair (λ, ϵ) is the content of the line starting with "EXM". Information about the system studied is given in the "NM_" (compound name) and "NMS" (solvent name) lines. Subsidiary information that may be useful in the evaluation of the data is also included in the same database record. For instance, rate constants for energy transfer (on the "KET" line) are not only interesting in their own right, but they are needed if an extinction coefficient that has been measured by an energy-transfer method is to be adequately evaluated. In Figure 4, the method (MET) is symbolized by LP/ET (laser photolysis/energy transfer). No physical units are put into the database but are taken to be implicit in the three-character labels.

A database record in the numeric databases can be related to the other two types of databases through the labels REF and NM_. The contents of the line beginning with REF is a six-character code for a unique reference. The reference is tagged with this code in RCDCbib. Thus, the connection is present such that bibliographic data can be associated with the database records of reported measurements. In a similar manner, RCDCreg (see below) contains lines labeled NM_, which match the contents of such lines in the numeric databases. These links open up numerous possibilities for the association of descriptive information in the registry file with the numeric database record. These various interactions between the databases will be discussed further below.

The description has been up to now the description of a single database record in the numeric databases. The text version of the numeric database is a collection of such items. The order of the database records is arbitrary. They can be selected from the database and sorted in a variety of ways as is needed. One such sorting is done in the table programs as a prelude to publication, and this process is described below.

Also, the text versions of the numeric databases are used as the basic source of information for the numeric databases that are used in on-line searches at the RCDC. How the database records are modified and structured in these numeric databases is discussed in the last section.

The above description of the numeric database has mainly been of the physical scheme[22] of the text numeric databases at RCDC; the conceptual scheme[22] of these numeric databases has been described in Table I and Figure 2. If the text numeric database, the RCDCreg file, and the RCDCbib file are considered as one database, then relationships are defined between the entity of the numeric file and the entities of the other two files by the labels NM_ and REF, as described above. Thus,

the database is presently implemented as illustrated in Figure 2. This makes the logical model[22] of the database a simple network.[23]

REGISTRY FILE

Structure

In chemical databases, chemical species need to be identified in a unique way to satisfy the requirements of computer processing. This is to be contrasted to the requirement for useability of the database, which dictates that users must have some flexibility in specifying chemical species. There are several major chemical nomenclature systems,[25,26] not to mention the tremendous variety of common names and other ways that chemical species are represented. To satisfy the contrasting needs of uniqueness and flexibility, a registry file has been developed. The interaction of a registry file with numeric databases containing chemical information was pioneered by the NIH/EPA Chemical Information System (CIS),[27] where structure searching and referral were additional features designed to enhance the useability of the various databases in that system. The RCDC registry file acts as a dictionary to the various ways by which it is desired to represent chemical species in the RCDC numeric databases and derived publications. Chemical Abstracts Service (CAS)[28] Registry Numbers are used whenever available.

The RCDC registry file was initially constructed from magnetic tapes,[29] which were acquired containing CAS Registry Numbers, names, and formulas for a number of chemical compounds. Additional entries were added as needed, in particular Registry Numbers and associated information for radicals and ions. The present practice of some journals of including CAS Registry Numbers in published papers allows the data compiler to obtain those identification numbers at the time the data are extracted from the paper.

The RCDC registry file is in a database format. In the entity-relationship model of databases, the chemical species is the entity (see Table I). The attributes are name, synonyms, Registry Number, molecular formula, inverted name, and several miscellaneous items listed in the physical description of the database record. The choice of the chemical species as the entity of the registry file implies that the fundamental structure of the database will be such that there is only one database record for each Registry Number. (The Registry Number can then be used as a unique primary key in any indexing schemes for the registry database.)

The principle on which the registry file was set up also demanded that one other attribute be unique. It could be a name or a formula (NM_ or RX_). This allows a numeric database file to interact with RCDCreg and "find" the Registry Number with that unique attribute.

The basic information is physically contained in an ordinary sequential file that is organized in a similar manner to the text file of the numeric database described above. Again, a database record is a sequence of lines beginning with an "SRN" line. In this file, however, the content of the SRN line is the Registry Number. The various names and synonyms are listed separately on lines labeled NM_, and inverted names are the contents of "NI_" lines. Molecular formulas are the contents of "FOR" lines, and "RX_" lines are provided for line formulas, abbreviations, and symbols, which are often convenient and useful ways to designate chemical species.

Database Management System

All of the access to the registry file is made through a series of database management programs. These programs have been developed at the RCDC on the principle that they can gain their access to the information in the registry file through two separate index files. This type of access to information in an ordinary sequential file via an indexing technique is made possible because the Record Management Service (RMS)[30] of the VAX Virtual Memory operating System (VAX/VMS) allows random access to sequential disk files. In particular, RMS allows a user-written subroutine to obtain the record file address (RFA) of a record on the disk, and it also allows reading from a record of a sequential file at a particular RFA supplied by the user. This is implemented by specifying a USEROPEN parameter in a nonstandard FORTRAN 77 OPEN statement.[31] Thus, this DEC-supplied software allows a traditional database design of indexing with keys to be employed, even though a sequential disk file is used to hold the fundamental information.

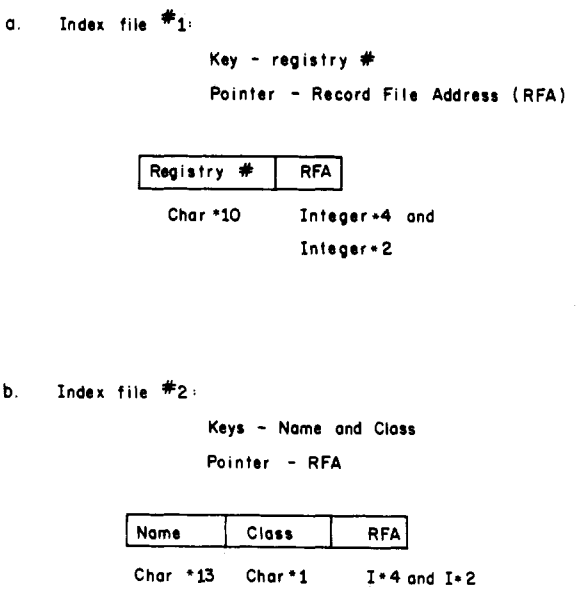

Figure 5. Form of index files associated with registry file.

Specifically, the design of the registry database involves two index files in addition to the sequential disk file containing the information. These index files (see Figure 5) are indexed sequential files. The simplest file (Figure 5a) contains two fixed-length fields, a *key* field containing the Registry Number and a pointer field containing the record file address (RFA) of the SRN line, which, in turn, contains the Registry Number. The Registry Number is unique; so this file allows a straightforward key indexing access to the actual registry file. The other index file (Figure 5b) contains three fields. Two of the fields are key fields. One of the key fields is a 13-character field containing a compressed chemical name, molecular formula, line formula, or inverted name. The other key field is a one-character field denoting which of the classes listed in the last sentence is in the other key field. The last field again holds the pointer that is the RFA of the database record on the disk. Both of the index files have fields of fixed length.

This structure of the registry file allows for the traditional database operations to be carried out with database management programs written almost exclusively in FORTRAN. The database operations that can be performed with the registry database are addition, subtraction, modification, and selection of individual lines or whole database records. There is a set of two programs that modify or add and subtract information. The first program, called REGUPD, logically deletes a database record and saves the deleted information in a text file. If the user wants to modify this information, the saved file can be edited and subsequently added back to the registry file with a program called ADDREGX. ADDREGX can also add totally new database records to the registry file. At the heart of both of these programs is the nonstandard FORTRAN 77 OPEN statement and two short assembly language subroutines, GETRFA, which contains the RFA, and

SETRFA, which resets the next record read. A separate program exists that can be run at convenient intervals to clean up logically deleted database records. Finally, CMPDCHECK is the look-up program that can be used to look for the presence of a chemical species by name, formula, or Registry Number.

The registry database and its management system can be used as a self-contained, on-line system (view 1 of Figure 1). However, since the main purpose of the registry file at the RCDC is to eliminate the inclusion of the same descriptive chemical information repeatedly in the numeric database, RCDCreg must intereact with the numeric databases themselves to achieve this purpose. The chief areas where this interaction takes place is in the construction of the numeric databases, in the execution of on-line searches of the numeric databases, and in the preparation of data for publication (through the usefulness of RCDCreg in the production of indexes for the tables). The last two uses will be discussed later. In helping to develop the numeric databases, the registry file serves to standardize the nomenclature. To assist in the implementation of this task, there exists a program, called CMPDMISSING, which interacts with both the text form of the numeric databases and the registry database. CMPDMISSING can be run directly against the numeric databases to see which compounds are not in the registry file or which may be entered incorrectly in the numeric databases.

BIBLIOGRAPHIC DATABASE

Just as the registry database saves space in the numeric database by containing chemical information, the RCDC bibliographic database saves space by containing the complete references. RCDCbib was developed much before[32] the RCDC numeric databases, and it is searchable on-line[33] (view 2 of Figure 1). It has been used in the past to produce the reference lists for published data compilations and evaluations,[16,17,20,21] as well as bibliographies, and other printed versions.

At the conceptual level, the entity of RCDCbib is a *reference*; some of its attributes are given in Table I. On the physical level, the structure of RCDCbib is analogous to that of the registry database. It has a primary file that contains all information in an ordinary sequential file with database records again beginning with, and limited by, SRN lines. Random access to the primary information is possible at the RCDC through index files for authors and keywords. These index files are again indexed sequential files with fixed-length fields containing various keys and a pointer to the disk RFA of various SRN's. References can then be extracted by choosing various keys (on-line searching can be done with the program SELECT).[32] In RCDCbib, each reference is uniquely identified by a serial number.

RCDCbib interacts with the numeric databases in two ways. First, codes for the references (serial numbers) can be copied from the REF lines of the numeric databases, and then these codes can be used to obtain the bibliographic information in RCDCbib. Programs to generate reference lists have been written and used at the RCDC (REFLISTV in Figure 3). The second use of RCDCbib is to generate author indexes to the data tables that are being prepared for publication.

USE OF DATABASES IN PUBLICATION OF TABLES

Table Formation

The manner in which the three major types of databases at the RCDC are used to produce tables of data with accompanying lists of references and indexes has been briefly indicated above. A series of programs (see Figure 3) is available to assist in producing files from the numeric databases suitable for table formation. The first program (RXN1 in Figure 3) can select database records from the database for a particular chemical species, or all database records in the database can be selected. The derived file 1 consists of the selected database records to which lines containing information from the registry file have been added, which may be used in subsequent programs.

RXN1 also produces an "ord" line, which may contain a string of characters that distinguishes inorganic, organic, and macromolecular chemical species and allows the items to be sorted in a desired order for producing a table. The character string may contain a compressed chemical name with other strings concatenated as desired. As an example, in the triplet–triplet absorption (TTA) compilation,[21] the items are ordered according to triplet name, solvent, and year. Thus, the ord line has three separate strings concatenated. Lines of information added from the registry file contain lower-case, three-character labels, e.g., a Registry Number may be added to the file under the label "rnr". Such lines produced in RXN1 (or RXN3 described below) can be removed by the program STRIPRXN.

The second program in the series (RXN2) sorts the database records in derived file 1 according to the ord lines. Preliminary to the actual sort, an auxiliary file of disk records is created with one field containing the RFA of each SRN along with another field containing the contents of the corresponding ord lines. These auxiliary disk records (each containing an "ord field" and a corresponding "RFA field") are then sorted, on the basis of the ord field. The sorting is done with DEC-supplied software.[34] The ordered output from the sort can be used to direct the ordering of another derived file (2) with the sequence of the sorted disk records. The database records to be put into the sequence are retrieved from derived file 1 with the RFA's that have been carried along in the RFA field of the sorted disk records.

The third program (generically called RXN3 in Figure 3) in the series has as one of its main purposes the composition of additional lines (containing lower-case, three-character labels) from the data lines present in the database record itself or from the registry file. The program allows information from several data lines to be combined to create the contents of a column in the table, while at the same time units implicit in the labels are appended and punctuation is added. RXN3 also inserts consecutive numbers on selected SRN lines, which can be used as table entry numbers; for example, in the TTA compilation,[21] the sorted database records were in alphabetical order by compound name, and a new number was assigned when a new rnr line (the Registry Number for the compound added by RXN1) was found.

The mechanism by which comments are composed by RXN3 is through the use of two subroutines that are an important part of the RCDC database management software for sequential files. The first subroutine is called LOAD77, which loads an entire *database record* into a buffer. The second subroutine, called PARS77, parses the records in the buffer by being able to pick apart a database record into its constituent parts. An example of how the two subroutines work together in composing comments is shown by the exercise of adding triplet energies to the end of each comment column in the TTA compilation.[21] The database record for each table entry is sequentially extracted from the input file (second-derived file in Figure 3) and put into a buffer by LOAD77. PARS77 looks at the database record to see if there is any "TSE" line. If there is such a line, it extracts the numeric value, which is the contents of the line, and passes this number to the main calling routine of RXN3. The program finally composes the comment, adding punctuation, spaces, and the units kJ/mol. The composition of the comments proceeds with standard FORTRAN 77.

In the final step of table production, tables are composed by a program named VARTABLE. The input to VARTABLE is

a file (usually derived file 3, see Figure 3) containing a collection of *database records* from a numeric database arranged in the desired order. The output file from VARTABLE contains all of the information that is to appear in the table along with text-formatting commands. The program uses a formatting file containing the following information: number of columns, column heading and label of the line providing the data for each column, width of each column, and columns for which data on the same line but separated by ";" will be forced to a new line in the table. The output file from this program is in a form suitable for conversion to some typesetting machines.[35] More specifically, the commands are appropriate for a word-processing program called VARPRINT, developed at the RCDC, that produces a printout with variable-width characters, including superscripts, subscripts, Greek letters, and various fonts such as italic, bold-face, etc. on a Versatec electrostatic printer.

Index Formation

The accompanying indexes to the tables of data are produced by the program INDEXEXTR (see Figure 3). This program must associate the data-table entry number of a compound or reaction (a number assigned by RXN3) with a chemical name, chemical formula, or author. This involves a complicated interaction between the numeric database, the registry file, and the bibliographic database. The program operates on the same file that was input to VARTABLE (usually derived file 3 in Figure 3). The data-table entry number is first associated with the Registry Number, listed in the rnr line, which provides access to the registry file where the various synonyms and also the molecular formula can be obtained. From the synonyms and the associated data-table entry number, a chemical name index can be formed; and from an association of the table number with the formula, a formula index can be formed. Authors must be found from the bibliographic file with the contents of the REF line (not shown in Figure 3). From the information derived from INDEXEXTR, files can be produced containing VARPRINT text-processing commands for printing indexes to the chemical names, to the molecular formula, and to the authors of the cited references. The files can also be converted to typesetting files.

Reference List Formation

Reference lists are produced by a sequence of two programs: REFERENCE extracts the contents of REF in one or more numeric files, duplicates are deleted, and the list is sorted. REFLISTV uses the list of references to obtain bibliographic data from RCDCbib; there are several options that can determine a variety of outputs. The files are formatted to be printed by VARPRINT or converted to typesetting files.

SEARCHABLE DATABASES

Tables of data can be produced very efficiently and conveniently with the computerized system described above. However, the physical form of the numeric databases used for that purpose is not very suitable for on-line searching. These numeric databases are in the form of sequential files, which can be searched by the VAX EDT text editor, but such files, by themseleves, do not allow for random access as do the registry and bibliographic databases.

In order to develop a searchable numeric database, it was decided to use again indexed sequential files. The most straightforward way to do this would be to take a database record in the text file and convert it into a physical record of an indexed sequential file. This would need to be done so that the contents of each line in a single database record would be mapped into a field of a single record in the indexed sequential file (see Figure 6). The problem is that many of the lines within a database record in the numeric database, such as the comments, are of variable length. Version 4.0 of the VAX/VMS operating system allows for indexed sequential files with variable-length fields. This feature of the operating system makes feasible the use of single, indexed sequential files for each of our numeric databases.

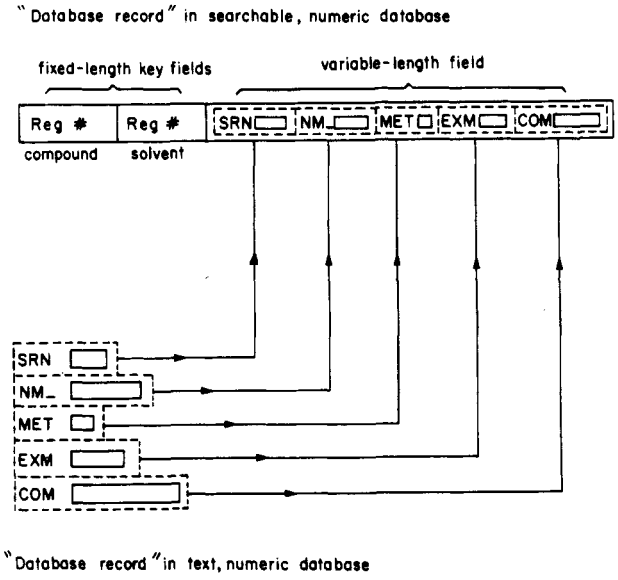

Figure 6. Mapping of a single database record from the text, numeric database onto a single record in the searchable, numeric database.

Currently there are two searchable, numeric databases at the RCDC. Both databases are used in keyed-access mode. The first one is the triplet–triplet absorption (TTA) database, which has two fixed-length key fields, one for the triplet state and one for the solvent. The primary key is the Registry Number of the triplet state, with the Registry Number of the solvent being the secondary key. (The terms primary and secondary are specific to RMS VAX/VMS, and their use is described in the following.) The Registry Numbers have the dashes stripped out and are left-justified in their fixed-length fields. The records are directly accessible through the primary key, which does not have to be unique in VAX/VMS RMS. If the value of the primary key is not unique and the secondary key is not specified, then RMS will give the first occurrence of the secondary key. The next access to the file will give the second occurrence of the primary key with some secondary key. The third field is of variable length and contains the information about the measurement.

The second searchable, numeric database (RATES) contains the rate constants for the reactions of H, OH, hydrated electrons, and HO_2/O_2^- with a variety of solutes in aqueous solution. The records in these files have three key fields. The first two key fields contain the Registry Numbers of the two reactants, and the third key field contains the rate constant of the reaction. In the third key field the rate constant is converted into an *I**2 integer binary field. The rate constant is converted to an integer by the transformation

$$\text{integer} = 1000 \log k \qquad (1)$$

where k is the second-order rate constant that is being compiled. The fourth field in the record is a nonkey field of variable length. This field contains information about the source of the rate constant.

Both of these databases can be searched on-line at the RCDC. The registry file serves as an intermediary to convert user-supplied names into Registry Numbers. The Registry Numbers then allow random access to the information in the numeric databases through the key fields. The use of a key field that contains numeric data, as with the third key in the RATES database, allows the file to be searched for reactions having values within a specified range of values of the rate constant.

The searchable, numeric databases are generated from the text numeric databases using the LOAD77 subroutine. When a database record is read into a buffer with LOAD77, it is then searched for strings containing Registry Numbers. The dashes are stripped out, and the numbers are saved for the first two fields in the record of the indexed sequential file (see Figure 6). Next, the character strings corresponding to the database record in the buffer are concatenated into a single character string. The entire database record is then written on the disk, with the concatenated character string comprising the last field of the record of the indexed sequential file. This record is written with no fixed length and no fixed position.

Random access to the records in the searchable, numeric databases is implemented through the use of keys and depends on the system software, RMS of VAX/VMS. The problem of keeping track of the record file address (RFA) on the disk is transparent to the user and programmer alike. This is not the case with the text versions of the other on-line databases (RCDCbib and RCDCreg). Even though there is keyed access to these text databases, it is keyed access to the RFA, which is used as a pointer to the data itself. Thus, the programmer has to keep track of the RFA explictly and needs assembly language programs to do so.

Finally, there are some contrasts, with respect to maintenance and size, that point up the relative merits of the two types of physical implementation of the on-line databases at the RCDC. The type of database exemplified by the searchable numeric databases is the easier to maintain. For instance, by use of a nonstandard FORTRAN DELETE statement,[36] a logical delete of a record can be done. In contrast, for the type represented by the registry and bibliographic databases, the RCDC's database management system must do this task, which involves not only marking the database record for deletion but also marking records in indexed files. On the other hand, with respect to size, the type of database design in the registry and bibliographic databases (currently occupying about 5.6 and 53.5 Mbytes of disk space, respectively) has a distinct advantage over databases based on indexed sequential files with only fixed-length fields. This is the case because they allow databases with information requiring a wide range of length to be fit into a small amount of disk space. However, as long as the information does not vary in length by too much, then the use of variable-length, indexed sequential files is probably the best choice for the types of database design discussed above.

ACKNOWLEDGMENT

The Radiation Chemistry Data Center is supported jointly by the National Bureau of Standards, Office of Standard Reference Data, and by the Office of Basic Energy Sciences of the Department of Energy. This is Radiation Laboratory Document No. NDRL-2810.

REFERENCES AND NOTES

(1) O'Donnell, J. H.; Sangster, D. F. *Principles of Radiation Chemistry*; Elsevier: New York, 1970.
(2) Swallow, A. J. *Radiation Chemistry: An Introduction*; Wiley: New York, 1973.
(3) Calvert, J. G.; Pitts, J. N. *Photochemistry*; Wiley: New York, 1966.
(4) Turro, N. J. *Modern Molecular Photochemistry*; Benjamin/Cummings: Menlo Park, CA, 1978.
(5) Matheson, M. S.; Dorfman, L. M. *Pulse Radiolysis*; MIT Press: Cambridge, MA, 1969.
(6) Gordon, S.; Schmidt, K. H.; Honekamp, J. R. "An Analysis of the Hydrogen Bubble Concerns in the Three-Mile Island Unit-2 Reactor Vessel". *Radiat. Phys. Chem.* **1983**, *21*, 247–258.
(7) Kalyanasundaram, K.; Grätzel, M. "Photochemical Conversion and Storage of Solar Energy". *Photochem. Photobiol.* **1984**, *40*, 807–821.
(8) Pavlopoulos, T. G. "Prediction of Laser Action Properties of Organic Dyes from Their Structure and the Polarization Characteristics of Their Electronic Transitions". *IEEE J. Quantum Electron.* **1973**, *QE-9*, 510–516.
(9) Chameides, W. L.; Davis, D. D. "The Free Radical Chemistry of Cloud Droplets and Its Impact upon the Composition of Rain". *J. Geophys. Res., C: Oceans Atmos.* **1982**, *87*, 4863–4877.
(10) Schwarz, H. A. "Applications of the Spur Diffusion Model to the Radiation Chemistry of Aqueous Solutions". *J. Phys. Chem.* **1969**, *73*, 1928–1937.
(11) Brummer, J. G.; Helman, W. P.; Ross, A. B. "A Catalog of Data Compilations on Photochemistry and Photophysical Processes in Solution". *NBS Spec. Publ.* (U.S.) **1980**, *No. 578*.
(12) See, for example, reference 16 and 17.
(13) Ross, A. B. "The Radiation Chemistry Data Center". *J. Chem. Doc.* **1967**, *7*, 34–36.
(14) Lide, D. R., Jr. "Critical Data for Critical Needs". *Science (Washington, D.C.)* **1981**, *212*, 1334–1340.
(15) Hug, G. L. "Photochemical Activities of the Radiation Chemistry Data Center, University of Notre Dame". *Eur. Photochem. Newsl.* **1983**, *June*, 5–6.
(16) Wilkinson, F.; Brummer, J. G. "Rate Constants for the Decay and Reactions of the Lowest Electronically Excited Singlet State of Molecular Oxygen in Solution". *J. Phys. Chem. Ref. Data* **1981**, *10*, 809–999.
(17) Latest example: Ross, A. B.; Neta, P. "Rate Constants for Reactions of Aliphatic Carbon-Centered Radicals in Aqueous Solution". *Natl. Stand. Ref. Data Ser. (U.S., Natl. Bur. Stand.)* **1982**, *NSRDS-NBS 70*.
(18) Rumble, J. R., Jr.; Hampel, V. E., Eds. *Database Management in Science and Technology*; North-Holland: Amsterdam, 1984.
(19) Ullman, J. D. *Principles of Database Systems*; Computer Science Press: Potomac, MD, 1980.
(20) Bielski, B. H. J.; Cabelli, D. E.; Arudi, R. L.; Ross, A. B. "Reactivity of HO_2/O_2^- Radicals in Aqueous Solution". *J. Phys. Chem. Ref. Data* **1985**, *14*, 1041–1100.
(21) Carmichael, I.; Hug, G. L. "Triplet-Triplet Absorption Spectra of Organic Molecules in Condensed Phases". *J. Phys. Chem. Ref. Data* **1986**, *15*, 1–250.
(22) Bestougeff, H. "Designing the Database Management System". In *Database Management in Science and Technology*; Rumble, J. R., Jr.; Hampel, V. E., Eds.; North-Holland: Amsterdam, 1984; Chapter 6, pp 119–166.
(23) Kroenke, D. *Database: A Professional Primer*; Science Research Associates: Chicago, 1978.
(24) Martin, D. *Database Design and Implementation on Maxi- and Mini-Computers*; Van Nostrand-Reinhold: New York, 1980.
(25) Rigaudy, J.; Klesney, S. P. *IUPAC Nomenclature of Organic Chemistry*; Pergamon: New York, 1979.
(26) *Chemical Abstracts Index Guide*; Chemical Abstracts Service: Columbus, OH, 1985.
(27) Heller, S. R. "The Chemical Information System and Spectral Databases". *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 224–231.
(28) *Chemical Abstracts Service Registry Handbook—Number Section*; Chemical Abstracts Service: Columbus, OH, 1971; 1971 ed. and annual supplements.
(29) Magnetic tapes of catalog information were very kindly furnished by Aldrich Chemical Co. and Eastman Kodak Co.
(30) *VAX Record Management Services Reference Manual*; Digital: Maynard, MA, 1984.
(31) *Guide to Programming on VAX/VMS (FORTRAN Edition)*. Digital: Maynard, MA, 1984; Sections 9.8.1 and 9.8.2.
(32) Helman, W. P.; Ross, A. B. "Radiation Chemistry Data Center. Information Services Produced from the Bibliographic Data Base". *Radiat. Phys. Chem.* **1980**, *16*, 425–430.
(33) RCDCbib is on-line at the Notre Dame Radiation Laboratory for in-house use on the NDRL VAX 11/780. It is available to other users within the U.S. through DOE/RECON, a facility of the Department of Energy, Office of Scientific and Technical Information, Oak Ridge, TN.
(34) *VAX/VMS Sort/Merge Utility Reference Manual*; Digital: Maynard, MA, 1984.
(35) Calabrese, J. T.; Seymour, C. L. *Bedford Mark-up Language*; National Bureau of Standards: Washington, D.C., 1985; private communication.
(36) *Programming in VAX FORTRAN*; Digital: Maynard, MA, 1984; Section 13.7.