———————ARTICLES———————

Problems and Solutions in the Development of Hard and Soft Versions of a Chemical Reaction Hazards Database

Peter E. Barker,[†] Leslie Bretherick,[*,‡] and Frank A. A. Kay[§]

Desktop Display, Brighton, Sussex BN1 7JA, U.K., Chemical Safety Matters, Bridport, Dorset DT6 6AE, U.K., and AAF Partnership, Chorleywood, Hertfordshire WD3 5AP, U.K.

The Bretherick Chemical Reaction Hazards Database, originally in encyclopedic book format, was compiled during 3 decades to meet the need of chemists and chemical engineers for a comprehensive and classified set of documented information on the real or potential hazards arising in chemical reaction systems. The methods of compilation and presentation of the data have steadily evolved alongside the development of general information technology, and a soft version of the database for selective data retrieval using a microcomputer has now been released. The successive phases in this overall development are described in terms of problems encountered and solutions adopted.

INTRODUCTION

The deeper origins of this long-term and continuing project lay in two unrelated personal experiences of the senior author (L.B.) of this paper in the distant past. In 1946 as an inexperienced chemistry graduate, he (and his right knee) had a close encounter of the worst kind with oleum (fuming sulfuric acid), which left him with a permanent scar and an enhanced respect for reactive chemicals and personal safety. Then in 1948, during a carefully supervized process development program involving oxidation with chromium trioxide/acetic anhydride mixtures, it emerged that under some circumstances the mixture could decompose spontaneously with extreme rapidity, accompanied by ignition. On this occasion, had he not fortuitously left the laboratory 5 s previously, his personal oxidation would have ensued, and this article would not have been written. His respect for reactive chemicals was further enhanced, and a warning to other chemists about the unpredictable nature of the oxidation mixture was published.[1] A published response[2] to this warning revealed that two similar incidents, one on a larger scale, had occurred 4 years before but had not then been thought to be significant or worthy of publication.

The immediate origin of the project was when, some 16 years later, another chemist published a letter in the same journal[3] describing the same hazard as a novel finding. It then became obvious that literature resources, as they then existed, were largely inadequate to inform practicing chemists of such reactive chemical hazards. However, it was not until several months later, after the publication of his letter[4] drawing attention to the need for a reliable safety data source for reaction hazards had been followed by a deathly hush, that Bretherick realized that if he wanted to see such a data resource, he would have to generate it himself. The fact that three publishers then offered to publish it when completed clinched the need for a comprehensive and well-ordered set of data which would be user-friendly to chemists.

The reasons for this lack of reaction safety data were not clear, but perhaps involved such factors as the reluctance to reveal professional errors, or commercial secrecy, or even a complacency in regarding fire and explosion as the normal occupational hazards of chemists. However, a directly identifiable cause was that such letters to editors of journals as were published and details of unexpectedly violent reactions not forming a main theme of technical papers were not subsequently indexed or abstracted, so they could not be retrieved later. A further direct factor identified some years later by Trevor Kletz is that unlike people, organizations have no intrinsic memory, and items of safety-relevant information not meriting an indexed written report are lost when the person concerned moves on.

THE NEXT STEP

Accepting the challenge to provide this data set was one thing, but implementing the decision was something very different. Initially, a scope including both reactive and toxic hazards of chemicals seemed desirable, but increasing contact with available literature resources soon led to the conclusion that toxic hazards already were well-covered in both specialized and encyclopedic compilations. Toxic hazards were thereafter excluded, and all efforts were concentrated on reaction hazards. The main aim of the project was then redefined as the provision of a comprehensive and logically arranged set of documented data which would allow experienced chemists to make an assessment of any reaction hazards likely to arise during the pursuit of their exploratory or development interests. A secondary aim was to present the information in a way that would allow less experienced chemists and students to understand the underlying cause(s) of the events described.

It was realized at this early stage that there would be two conflicting factors operating. First was that the book would need to be compact and produced as cheaply as possible to encourage wide circulation for maximum hazard reduction. Second, the wide scope and educational aspects mentioned above would need more space to achieve adequate coverage and clarity.

It was also clear from the outset that it would be impossible to scan the whole of the published chemical literature in an attempt to get complete coverage of every unstable substance

† Desktop Display.
‡ Chemical Safety Matters.
§ AAF Partnership.

CHEMICAL REACTION HAZARDS DATABASE

J. Chem. Inf. Comput. Sci., Vol. 33, No. 2, 1993 **185**

or violet reaction which had been described, and that secondary sources of information would have to be used to build up a collection of earlier data which would regrettably be incomplete. For more recent information, many of the multivolume descriptive texts on inorganic,[5,6] organic, or organometallic[7,8] chemistry served as rich sources, and the series devoted to strictly practical aspects of synthetic inorganic[9,10] and organic[11] chemistry were particularly valuable. In the area of industrial chemistry, two series covering accident case histories in the U.K. from 1930[12] and in the United States from 1950[13] gave many good examples of reaction hazards.

DATA CONTENT AND FORMAT

Work started in earnest at the end of 1964, and after a good deal of information had been assembled and reviewed in general terms for data content, a preliminary format was developed. It had been decided at a very early stage that each text entry should have a reference, preferably to the original literature, or minimally to the secondary source, to ensure that the collection would be regarded as authoritative, as well as providing, in most cases, access to more detail.

While most single chemical entities could readily be represented by a systematic name and an empirical or structural formula, it was also necessary to be able to include substances, many of natural origin, which could not be so represented. These were most readily accommodated, together with relevant topic titles not involving chemical names or formulas, in a smaller separate-alphabetically arranged section of the data collection.

The format adopted for single chemical entites began with the bold-title name, based on either the well-established IUPAC systems of nomenclature or on the preferred selection for chemicals used in industry set out in a recently published British Standard publication.[14] The empirical molecular formula (Hill system) was ranged right on the bold-title line. The literature reference and then the condensed text followed, and the entry was concluded wherever possible by assigning the title chemical to a characteristic descriptively named group (metal azides, nitroalkanes, nonmetal sulfides, etc.). For each of these group names there was a corresponding entry in the alphabetical section containing a list (index) of all the members of that group present in the main section text to correlate entries of similar type.

After much thought, it was decided to arrange the chemicals in the main section of the book in alphabetical order of their empirical formulae. This gave an immutable order for the text, unaffected by any subsequent changes in nomenclature systems. It also had the advantage that a chemical could be found if one were unsure of the name but could draw the structure and count the atoms. The upper part of Figure 1 illustrates all these points.

The main emphasis of the data collection was on reaction hazards of specific chemicals, either alone (instability) or in combination (reactivity). Secondary chemicals involved in reaction(s) with a main entry chemical were put in as secondary headings within the main entry with, of course, the literature reference to the particular effects of the reaction pair. A bold-titled main entry could then have several subsidiary segments, each dealing with the reaction of the title chemical with a (or several) secondary chemical(s).

Initially some difficulties arose in deciding which of the reaction pair of chemicals to use as the main entry. When it emerged that most hazardous reactions involve an oxidant (electron sink) as a reactant, this was taken to be the main chemical. In the few cases where an active reducing agent (electron source) was one of the pair, that was taken as the main chemical. Though this principle has been followed since, at a later stage cross-references were added subsequently as separate segments to main entries for the secondary chemicals to ensure that the reaction pair could be traced from either one of the pair.

FIRST PUBLICATION

As material was steadily accumulated, it was first suitably condensed, retaining only the salient points of reaction hazards, in manuscript onto 6-in. × 3-in. loose leaf sheets. This smallest size was chosen to encourage brevity (and eventually an economically priced book). The handwritten sheets were then typed up by the long-suffering Mrs. Bretherick (the senior author's writing was poor to say the least), and the typed copies were stored in binders kept for security in a different part of the house from the manuscript originals.

This went on for some years on a purely parttime basis, but by late 1971, the steadily increasing supply of information exceeded the capacity of the author to process it. It was then clear that alternative arrangements would be necessary to complete the task. At this point he managed to gain the support of the BP Research Centre management and was able to complete the work as a supporting research objective, devoting an extra 20 h a week to it. With this increased effort, the task was completed within 18 months, the binders now bulging with over 4000 sheets. After checking the entries and their order, a 1200-page working manuscript was generated in a week by photocopying 3–6 of the small sheets, overlapped if necessary, onto A4 paper. By this time, Butterworths had agreed to publish the work and had suggested that they could best meet the requirement for an economical method of production by having the text set on an IBM composer which gave camera-ready copy in a satisfactory proportionally spaced typeface, coupled with the ability to readily switch from roman to italic, bold, and symbol fonts using the interchangeable 'golfball' typeheads. This generally gave satisfactory results, though the composer operators found many of the long chemical names too much of a challenge, and the first proofs contained some 3500 errors in 976 pages (of which 808 pages comprised the main section). The subsequent correction and recorrection cycles became asymptotic in effect, with the introduction of new errors during each correction phase. Eventually this defeated the author, who gave in after three cycles with some 200 known errors in the text. Publication[15] was effected in April 1975, some 5 months behind schedule, as a result of the 3-day work week imposed on U.K. industry by coal and power strikes. Figure 1 (upper) shows the format, typeface, and unjustified text adopted for the first two editions.

Sales soon became sufficient to show that the book was meeting its anticipated need, and work began to update the text for a second edition, along much the same lines as before. The compilation task was greatly eased by the fact that *Chemical Abstracts* had started in 1974 to include letters to editors on hazard topics and also used the controlled vocabulary index-heading Safety to bring all such abstracts together in the Index Volume. The scope of the text was considerably extended by the recent availability of two modern encyclopedic inorganic chemistry texts,[16,17] among other sources.

The updating material was again handwritten (now abominably) on loose leaf sheets, but it was decided for family reasons to defer conversion to typescript to the end of the compilation phase. When this arrived, some 4 years later in September 1977, the author had a major decision to take. This was either to rewrite the manuscript sheets neatly so that

---

2,3-DINITRO-2-BUTENE $C_4H_6N_2O_4$

Bisgrove, D. E. et al., Org. Synth, 1963, Coll. Vol. 4, 374
Only one explosion has been recorded during vacuum distillation at 135°C/14 mbar.
See other   POLYNITROALKYL COMPOUNDS

DIACETATOPLATINUM(II) NITRATE $C_4H_6N_2O_{10}Pt$

Preparative hazard.
See   NITRIC ACID, HNO_3 : Acetic acid, etc.

393

Chlorine
See Chlorine: 3-Chloropropyne

See other HALOACETYLENE DERIVATIVES

†1027. Acryloyl chloride (2-Propenoyl chloride)
[814-68-6] C_3H_3ClO

$$H_2C=CHCO.Cl$$

1. Pyriadi, T. M., Chem. Eng. News, 1985, 63(44), 4
2. Griffith, T. E., Chem. Eng. News, 1985, 63(50), 2
3. Author's comment, 1986

A 500 ml bottle of the acid chloride stabilised with 0.05% of phenothiazine was shipped to a hot climate without refrigeration, and was stored on arrival in a fume cupboard for 2 days at temperatures approaching 50°C. The material polymerised, bursting the bottle and forming a solid foam [1]. The label recommended storage at 4°C. Another manufacturer proposed mislabeling or contamination by moisture to explain the incident [2], but the presence of polymeric foam appears to confirm that polymerisation at the extremely high ambient storage temperature was the primary cause, perhaps with some thermal decomposition and evolution of hydrogen chloride assisting in foam formation [3].

See other ACYL HALIDES. POLYMERISATION INCIDENTS

1028. 1,3-Dithiolium perchlorate
[3706-77-2]

$$SCH=S^+CH=CHClO_4^-$$ $C_3H_3ClO_4S_2$

1. Ferraris, J. P. et al., Chem. Eng. News, 1974, 52(37), 3
2. Klingsberg, E., J. Amer. Chem. Soc., 1964, 86, 5292
3. Leaver, D. et al., J. Chem. Soc., 1962, 5109

345

Figure 1. Page formats of first (inset) and fourth editions of the Hardbook of Reactive Chemical Hazards.

his wife could read them or to learn to type himself. This latter seemed overall most useful, and a self-taught crash touch-typing course began. The updating material was converted into the 600-page typescript for a supplementary volume, as had been agreed with the publishers' editorial department. However, when this completed manuscript was delivered to the publishers, their Managing Editor told the author that this would not do at all. He was adamant that what was required was a new work containing the existing text and updating additions merged into a completely new volume. The author's first response was to consign the publishers and their minions to everlasting fire (with frequent but unpredictable explosions), but, of course, this was not likely to produce a second edition. Sober reflection, coupled with a

CHEMICAL REACTION HAZARDS DATABASE

J. Chem. Inf. Comput. Sci., Vol. 33, No. 2, 1993 **187**

firm belief in the sentiment of the dog-Latin pseudo-adage *Nil illegitemi carborundum,* led to the decision to swallow pride and recast the text into a fragmented mergeable format, and this was completed within 3 months. The subeditor must have had a tedious time in organizing and checking the merging process. Apart from this, the overall 'pasting up' and production process remained unchanged for the new edition, and the error rate and correction pattern in typesetting the new text (some 45% of the original volume) were also closely similar to those previously endured. This had very significant implications for future action.

Some reviews of the first edition had drawn attention to the pressing need for a full index of the chemicals in the text to supplement the numerous partial indexes represented by the compound lists in the general part of the text. In these lists, the location of chemicals in the text was on the basis of their empirical formulas, not the page number, for obvious reasons. Since these title chemical names had been the major cause of typesetting problems, a novel approach to ensure freedom from further keying errors in the index was adopted. Page numbers were added manually to the title names in a spare set of final proofs, and the title lines were guillotined out to give some 4000 strips of paper 4 mm by 11 cm. These were manually sorted by the author and his family, initially into 24 alphabetically random piles and then each pile into index order. After a final check by the author for missorts and ignorable Greek or italic prefixes, the strips were gummed down onto separated sheets of 132-column line printer paper to produce a printed manuscript of the index of chemical names. There was great anxiety while this was in progress, lest the draught from a hastily opened door might cause disarray in the numerous piles of paper strips which covered every available horizontal surface in the house at Ascot. Fortunately, it was needless anxiety. Other improvements incorporated into the revised version included a glossary of technical terms used in the text.

Production of the second edition proceeded uneventfully apart from the predictably boring proofreading and correction cycles, which surely are the nadir in a chemistry author's life. Although the memory of these usually soon fades, on this occasion it was kept fresh with an, as yet, unspecified decision to do something different for the next edition. The contrasting zenith of pleasure at holding the first copy of the updated 1281-page larger format volume was savored in January 1979. Updating work for a third edition had already begun in October 1977, following a 4-week break by the author after closing the updating files in September for the second edition.

NEW TECHNOLOGY AND A RETHINK

The advent of word processing as a general business procedure and of automated typesetting directly from floppy disks in the print industry together provided the opportunity for a radical change in production method for the book. As the major problem previously had been the poor rendition of many chemical names by typesetters, this new technology offered the theoretical possibility of ensuring that only the results of the author's keystrokes would finally appear on the printed page of succeeding editions. The publishers were themselves interested in evaluating this new possibility and provided appropriate encouragement and material backing for the first of their authors prepared to try it out. However, at that time (though not any more), there remained one final hurdle for the publisher—that of finding a typesetter who would be prepared to put the vital National Graphical Association (NGA) union imprint onto camera-ready matter

which had not been keyed by an NGA member. Suffice it to say that such a typesetter was eventually found in 1983.

Apart from preventing printing errors and generally easing his keyboard work, the author realized that the new technology offered other attractive possibilities for improving the book and its contents. The most significant of these was that the completed text could be searched electronically to generate automatically the group assignment chemical lists and the main index of chemicals, both error-free and complete with the text location (page or serial number) for each chemical, rather than by the manual drudgery as previously employed.

This new freedom encouraged a considerable extension of the group assignment and cross-reference systems throughout the text, with a corresponding increase in its coherence. Other improvements incorporated into the third edition text included a considerable body of measured values for individual energies of decomposition[18] and of calculated values for composition and maximum possible exotherms for several thousand binary mixtures.[19] The attempt to include CAS Registry Numbers for all of the 4000+ title chemicals in the revised edition was not an unqualified success. Several hundred entries remained devoid of numbers after many months of combat with the CA Index Volumes in the Royal Holloway College Library, now the nearest access to CA since the author's early retirement in 1982. The lack of CAS Registry Numbers was partly owing to the ancient vintage of some of the literature describing the compounds, many so unstable as to be incapable of full characterization, though of reasonably probable identity. The other reason was the aging author's failure to adapt to the sometimes capricious changes in the CA nomenclature system seen across the span of indexing years.

THE NITTY GRITTY

An initial demonstration by the typesetters of the capabilities of their Comart CP200 CP/M micros and Linotron 202 typesetting computer and phototypesetter was followed by discussion on the required formats, font, and symbol repertoires and on the methods for representing chemical structures. A set of typesetting codes and format calls was then developed in collaboration with the typesetters, and the author was equipped with an identical Comart micro with WordStar to ensure compatibility. After he had become accustomed to the new WP jargon and floppy disk regimes, thanks largely to Mr. Frank Kay who had installed the micro and was also well-versed in computer techniques, a test file was keyed and coded by the author to give a file containing examples of all the desired typographical effects. Conversion of the file to bromide output revealed a few errors and misunderstandings, but minor changes to the coding system and Linotron translation table eliminated these problems, and a further trial produced near-perfect output. While all this was going on, compilation work to deal with the steadily increasing flow of safety information proceeded apace.

All that then remained (!) was to key in and then code the contents of the 1200-page second edition while incorporating the contents of the 3000-odd updating manuscript loose leaf sheets accumulated during the 6 years since 1977. Keying in the plain language version of the merged texts took some 14 months of steady application, to give a 3+-MB set of disk files which could be proofread on screen. Then a further 5 months of coding insertions gave an almost incomprehensible final coded text of around 4 MB on 14 360K disks.

Had the Standard Generalized Markup Language[20] been available in 1982–1983 when the fonts, formats, and specification for the new edition were being established (and if so,

had we known about it), we doubtless would have used it. But as it was, and in ignorance, our Nonstandard Specialized Markup Language (NSML), a system of 75 codes and format cells each consisting of a control character and two digits (^nn, mostly prefixed to text characters, but a few suffixed) together with @, which served as a hard carriage return, seemed to work well enough and gave us what we wanted on the printed page (and later, on the screen). All in all, a bit like the bumblebee that knows no aerodynamics yet flies anyway.

TEXT SEARCHING FOR LIST GENERATION

When the disk files were read into the typesetting computer, the original three-character typo-codes were translated into the five-character (^nnnn) codes or more usually a string (in a specified order) of these codes needed to drive the phototypesetter.

Two text searching programs, LISTGEN and INDEX, had been developed (by Frank Kay), one to generate the lists of group-assigned chemicals with page numbers and the other for the main chemical index, the latter also with cross-referenced synonyms and both with page numbers. The search targets for both these programs were thus strings composed of multiples (3–7) of five-character groups in a specified order, together with the words '*See other*' or '*See related*' prefixing the small capital group name(s) for LISTGEN.

When the main (Section 1) text had been set, autopaginated, proofed twice, and fully corrected, a copy of that file was processed using LISTGEN and then INDEX. Many of the output lists were shorter than expected, and this was traced to the fact that during manual correction of the typeset text file from marked proofs, the keyboard operators had not been consistent, individually or collectively, in their order of correction or insertion of five-character codes within the coding sequence strings. The original text had gradually become corrupted with a random variety of such code sequence inconsistencies that the text processing programs did not recognize as search targets.

EFFECTS ON SEARCH PROGRAM OPERATION

As a first step to reduce text processing times, all parts of the 2.5-MB Section 1 paginated text not directly involved in the searching procedures were deleted from the working copy, reducing it by some 80%.

The operation of the INDEX program involved downfile search of the text for a title coding sequence, then writing that title, with coding modified for nonbold roman typeface, and the page number to an index list which was alphabetically sorted after completion. The presence of a synonym generated an additional cross-reference back to the main title name. The LISTGEN program searched downfile for the small caps (s.c.) group assignment string(s), then upfile via any secondary chemical (with its page number) to the bold roman title. These were written to the group list(s) in the format roman Title: Secondary, p. nnnn.

String searching targets were normally located either at the start of a main entry or at the end of the entry or of a segment thereof. Normally only code-sequence variations at those points caused problems. The usual effect was to omit the title chemical (and synonym if present) from the index list or the whole entry from the group list(s). To identify and rectify the 'near-misses', INDEX and LISTGEN were each modified so that after writing to a list, the five-character code sequences used for this purpose would be copied to a 'correct code file' if not already there and then deleted from the truncated text copy (a separate copy being used for each program). The separate partially code-deleted copies were then each processed to permute the disordered multiple ^nnnn code sequences until they matched an item in the 'correct sequence file', which would then be overwritten. Both corrected files were then reprocessed to give additional list and index entries, which were combined with existing lists and sorted into page number order. The few remaining problem entries were listed and later hand-processed and merged by the author.

An unexpected source of problems had also arisen from the automatic page makeup routine of the Linotron. The code sequence for a bold chemical title always began with a 2-line drop code, but at the top of a page, the subroutine deleted this drop to save space. This code corruption caused INDEX to miss the title entry while searching downfile and LISTGEN to write the next but one title name above while searching upfile for the relevant title. The latter had caused some group assignment errors, but the remedial measures adopted picked up the lack of one code in some of the title sequences.

Figure 1 (lower) shows the improved appearance of the larger, justified and phototypeset text as used in the third and fourth editions.

THE FOURTH EDITION

This hard-won experience above led to the decision that in this next edition all text searching operations would be done on the original unpaginated and simply coded text files. It was therefore necessary to prefix each title compound with a serial number, so that this rather than the page number could be used in the compound lists and index to locate compounds within the text. This was probably the most significant change between the third and fourth editions. The serial numbers were inserted into the updated text files by a serializing program written by Frank Kay.

The overall plan was to move through the normal typesetting and first and second proof correction sequences, but with the Section 1 disk files onscreen, so that the author could correct the original files as proof correction, marking for printers' corrections proceeded. Copies of these fully corrected files would then be text searched to produce the group/topic and index lists for Section 2, for later sorting with the Fraser–Williams CHEMLIST program. These hopes were largely realized, though there were some minor problems which could be directly attributed to the steadily failing sight of the author, but these were eventually resolved at the keyboard by word processing manipulation.

The general aspects of the preparation and production of the recently published fourth edition[21] were expected to follow a course similar to that outlined above, with further extensions to the range and currency of publications used as sources of relevant data and addition of two new appendices giving a classified Index of Group and Topic titles, and an Index of the CAS Registry Numbers.

Undoubtedly, the high spot of the whole project was in 1988 when the originator gained the annual Award of the ACS Chemical Health and Safety Division for his tome, as 'an outstanding contribution to chemical safety'. All the labor had not been in vain.

On the other hand, there was a major disappointment in the typesetting phase when a new company, using phototypesetting equipment (and our typo-coding) identical to that used so successfully for the previous edition, produced first proof output riddled with an average of several spurious errors per page

2325. Hexamethylenetetramine

[100-97-0] $C_6H_{12}N_4$

⚠ **Structure VII, p. S-7**

Finely powdered dry hexamine is a significant dust

explosion   ference 22)

See entry

Acetic acid, A   itrate, Nitric acid

See ⚠ Nitr

Hexameth

PRESS ESCAPE TO CONTINUE

1-Bromopentaborane(9)

See ⚠ 1-Bromopentaborane(9): Air, or

Hexamethylenetetramine

ESC F1 Help   F5 Notes
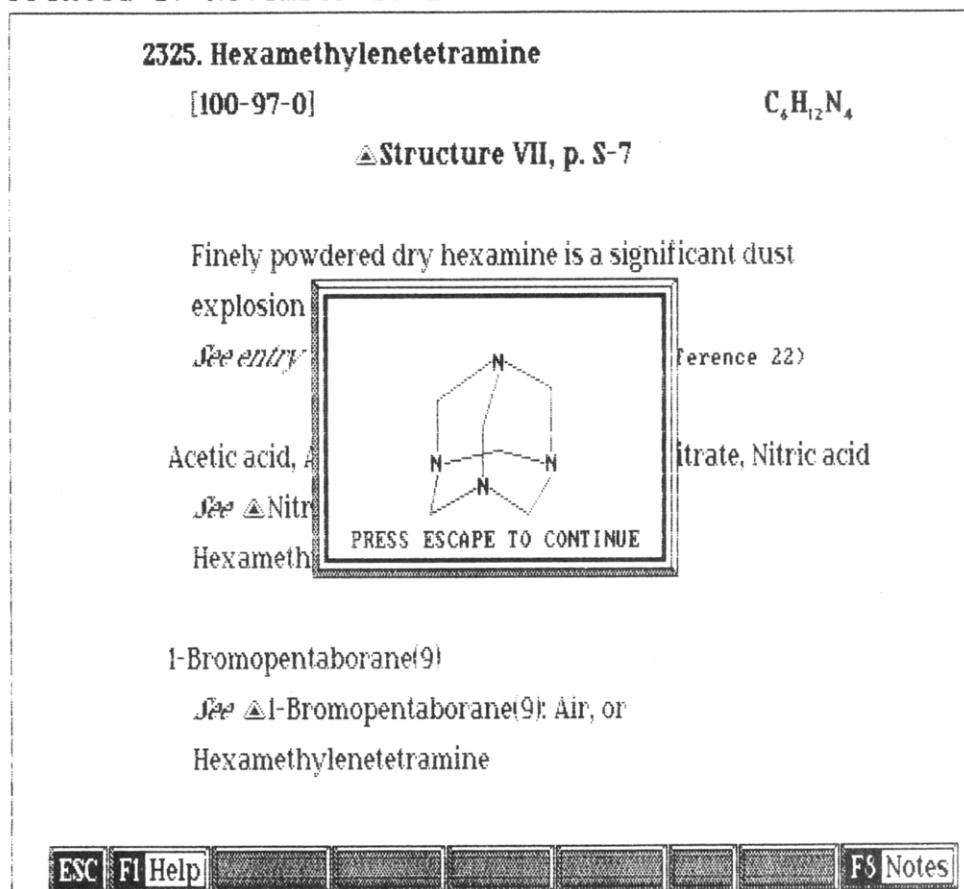
Figure 2. RCHD printout of ring compound text with structure displayed in window over main screen display.

caused by faulty translation tables. The specter of intensive proof correction still loomed over us, and we wondered whether SGML, the Specialized General Markup Language, might not have been better after all. Our efforts to get the translation tables fully corrected by the typesetters before proceeding were fruitless, and the second proofs were little better than the first had been, coming with a set of brand new errors in addition to uncorrected leftovers. Never in the field of typesetting conflict had so much aggravation been occasioned to so many by so few translation tables. We came to understand the true meaning of the phrase 'house of correction'.

Even when proofreading was far behind us, we were not yet out of the woods. The printers managed to mislay the very last page of the bromides and completed the print run and section stitching without it. As there was an insufficient amount of the special paper to reprint the last section, the single missing page was neatly fixed with adhesive into each of the 4000 final sections. Then, the binding machine broke down ...

When the book was finally published in September, 6 months behind schedule, it was discovered that the page between 1345 and 1347 was in fact an exact duplicate of 1046. One needs a well-developed sense of humor for such occasions!

MORE RECENT DEVELOPMENTS

In recent years, developments in other methods of publishing, distributing, and particularly in retrieving technical data have led us to the realization that the value of the existing database, hitherto available only in hardcopy form, could be greatly enhanced in various ways as a soft (onscreen searchable) version.

Some examples of online soft versions of long-established hardcopy data compilations which have become available within the last few years are the Kirk-Othmer Encyclopedia[22] and the Dictionaries of Organic[23] and of Organometallic[24] Chemistry. Of the two monumental Handbook series begun by Beilstein[25] and Gmelin[26] during the last century, and covering the whole of organic and of inorganic/organometallic/physical chemistry, respectively, the former was partly released online in 1988, and the latter at the end of 1991.

However, all of these compilations are at least some 2 orders of magnitude larger than the database under consideration, though all are highly structured in the book sense. A market survey by the Royal Society of Chemistry in 1986 indicated that this chemical hazards database would not be viable as a highly specialized online product. An alternative method of presentation was therefore sought, and that of the 'electronic book' seemed to be more appropriate, having the advantage that the existing magnetic text used to set the hardcopy version could be used with only minor modifications for the preparation of the onscreen version. An example of this type of presentation is concerned with the selection of gloves, etc. to prevent operator contact with corrosive or toxic chemicals.[27]

ELECTRONIC BOOK VERSION

The availability of machine-readable text and the character of the book, highly structured for knowledge content and cross-reference leads, suggested it was more suitable for electronic searching by title (chemical identity or group type, or topic), rather than by direct text searching. A block diagram of the book contents and the interrelationships thereof indicated that
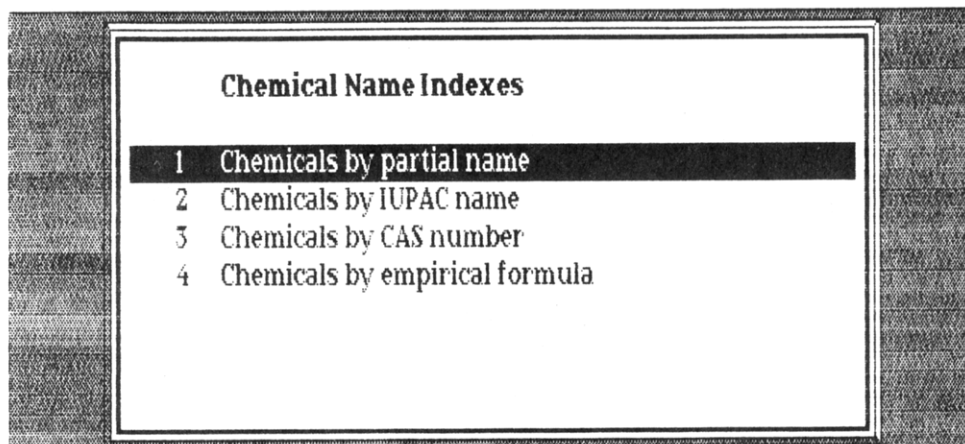
Figure 3. Main menu display (partial screen).

a menu structure with three primary (1y) menus could be developed to give rapid access to all the text and appendices via the existing or derived indexes. The menu structure and operation is further detailed below.

It was agreed with the publishers that screen formats and the user interface would be designed to simulate the appearance and use mode of the printed version as closely as possible, the latter of course at a much enhanced rate. This involved use of relatively large black on white multifont characters with sub- and superscripts and generous line spacing to ensure easy reading of screens and simple key control sequences for speed of use. In fact, only the 11 keys F-1–F-8, ENTER, ESC, and SPACEBAR are used to control the operation.

Preliminary processing of the text, containing ca. 1 MB of embedded tabulation and typographical control codes, allowed the latter to be simplified and reduced by 80%, using ASCII text encryption methods and tokenization of repetitive multiple coding sequences. A major enhancement of the text enabled users to follow up with keystrokes the numerous cross-references to related materials or incidents. In this context, the title line dagger symbol, indicating tabulated flammability data in an appendix, and the final *See other ...* phrase, pointing the relation of a chemical to its group and, conversely, of a group to its constituent chemicals, were all considered cross-references. This enhancement was effected by using the dagger, or an inserted triangle, as a 'hot-spot' symbol for a hypertext link at each such point. Key controls activate the 'jump' to the cross-referenced text. Insertion of some 23 500 'hot-spots' or symbols by the text processor was necessary.

Another feature of the book which needed conversion was the 20 pages of artwork necessary to show the structures of the 300-odd cyclic and multicyclic compounds which could not be reproduced linearly (with a super- or subscript ring closure rule) in the text. The artwork was input by hand, using a point and vector drawing package custom-built by Desktop Display. This and the associated electronic point and vector display routines proved to be extremely economical in space usage. The references in the text to the absent structures were 'triangle hot-spotted', with hypertext links to the electronic structure file, similar to other cross-references. Figure 2 shows a typical screen with a windowed structure displayed on the relevant text, after highlighting the triangle symbol.

The search and retrieval routines used for text-based searching were custom made by Desktop Display, based on buffered Boyer–Moore methods, with fast index-list searching using a binary chop procedure. Turbo Pascal was used throughout the software creation phase.
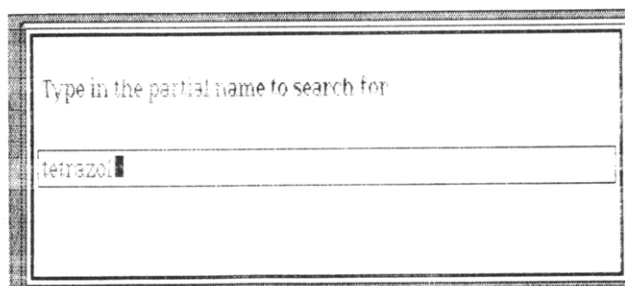

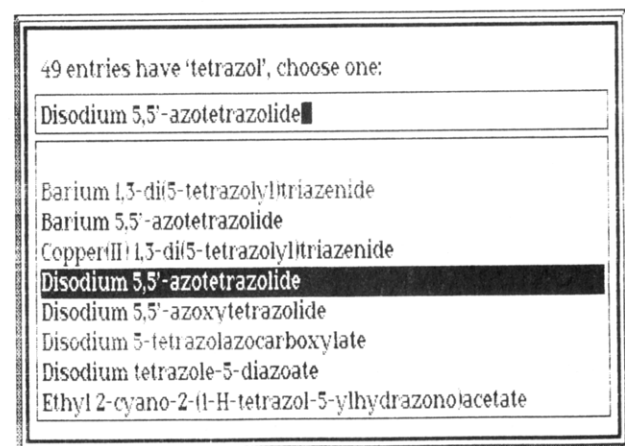**Figure 4.** Dialog box for partial name search (partial screen).

Figure 5. Compound list resulting from partial name search (partial screen).

TEXT INTERROGATION

The menu types required for text interrogation were closely related to the type of information content of the three discrete sections of the database. The main menu for the major Section 1, containing information specific to the 4597 named chemicals, gives access via a choice from four identifiers, as indicated in Figure 3. Of these, the first gives the greatest advantage over the book, that of being able to search for fragments of chemical names. Keystroked selection of item 1 produces a dialog box wherein the required fragment may be typed (Figure 4), and 3 s after pressing ENTER the alphanumerically sorted result is displayed (Figure 5). Selection of an item with the scroll-bar and then ENTER, gives immediately (<0.5 s) the text entry. Other hypertext jumps are similarly fast. The slowest response seen was a name fragment search for 'thyl' which took 34 s to compile and sort a list of 1313 compounds (30/s). These response times were seen on a PC-AT clone with a 386 20 MHz processor and would be slower with a less powerful processor at a lower clock speed.
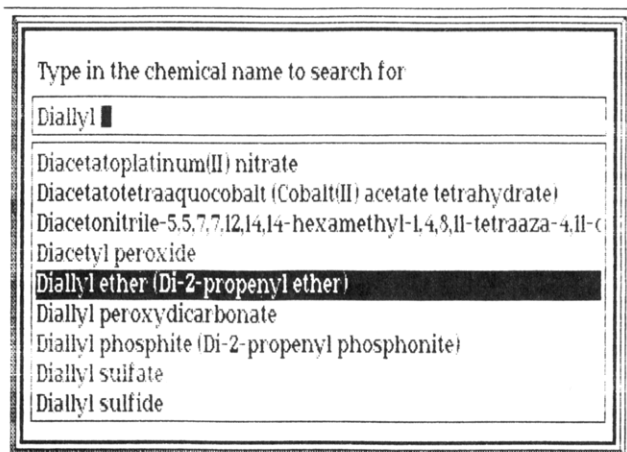
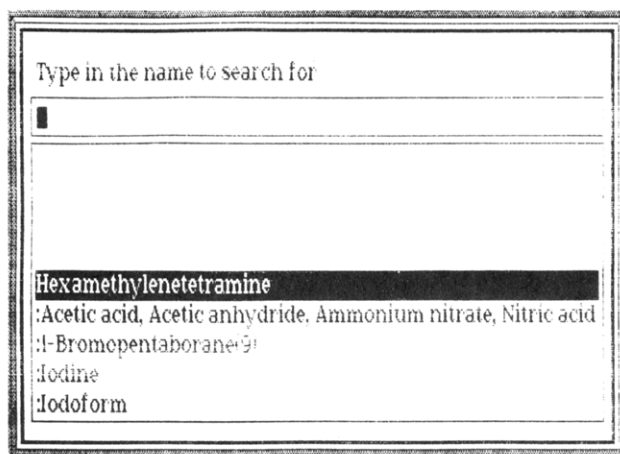**Figure 6.** Searching the IUPAC name index (partial screen).

Figure 7. Searching for a reaction pair segment (partial screen).

Selection of a later item on the main menu also gives a dialog box, but a window below displays the chosen index. As each character is input, the index entry scrolls to match the input characters, and the required identifier is often highlighted before input is completed. Figure 6 shows such a search for an IUPAC chemical name. The CAS Registry Number and empirical formula indexes are searched similarly. In all cases, ENTER produces the identified text entry onscreen.

When there are several secondary chemical segments within a main text entry, accessed as above via the IUPAC name index, ENTER produces a further window-displayed index of all the secondary chemicals, each prefixed by a colon, below the name of the title chemical, as in Figure 7. ENTER then shows directly the chosen segment of the main text. So much for the major (81%) Section 1 text.

In the minor (19%) Section 2 text, the 560 title-identified entries cover a wide variety of chemical groups and chemical-related topics, which required the use of some secondary menus as well as the primary one. The latter offers the five choices of classified chemical information: total class index by partial name, total class index, classification by group, classification by topic, or substances. The first choice produces a program-generated subset of the total index, while the last goes directly to the substance index. The three secondary menus offer a choice of several subdivisions, where ENTER leads to the appropriate scrollable index and then to the text.

Finally, the residual auxiliary text from the book preliminaries and Appendices 1–3 requires only a five choice menu of arrangement of text, overview of reactive hazards, journal abbreviations, tabulated fire-related data, and glossary of technical terms.

OTHER FEATURES

A back-track memory stores the last 100 screens viewed, and these may be reviewed backwards a screen at a time with the press of a key. When scrolling into a long entry, the current title details may be displayed in a pop-up box. At any point in a search, context-sensitive pull down Help screens are available, and relevant function key actions are permanently shown at the screenbase, as in Figure 2. All texts displayed are scrollable up or down, and screens may be printed if required, as in Figure 2. Additionally, a pop-up Notepad for recording comments or extracts of the text is available, and the file may be printed or exported to a word processor.

The 102 KB search program runs with 3.3-MB data files and 1-MB index files. Current machine requirements to run the Reactive Chemical Hazards Database (RCHD) are an IBM-compatible PC with MS-DOS 2.1 or later, 640K RAM, preferably a 286 or higher processor, 3.5-in. 1.44 MB or 5.25-in. 1.2 MB disk drive, 6 MB free on hard disk, EGA or Hercules graphics adaptor, and Epson or LaserJet printer (optional).

The electronic book version of the database has recently been reviewed[28] in this journal.

THE FUTURE

With advancing years, vision problems have overtaken the originator of this database, and he is now partially sighted with very impaired reading ability and so is unable to continue actively in the project. However, this will continue in the very capable hands of Dr. Peter G. Urben of Courtaulds Research Ltd., who has agreed to act as editor for future editions. He has played a significant part in the completion of the recent fourth edition by accepting without demur many of the thankless tasks of ferreting out missing CAS Registry Numbers and of checking the 'less than perfect' (to say the least) proofs. Compilation work for a fifth edition started in April 1989; Dr. Urben being assisted in this task by Mrs. Lindsay Battle and Dr. Martin J. Pitt. He already has in mind several ways to improve further the format and content of the hard version, and the recent soft version will doubtless also be improved in various ways, in the light of users' experience and of improvements in electronic search and retrieval possibilities. An update of the electronic book text to include information available to late 1992 is under way. A longer-term goal is to develop a database shell structure which will allow the editor to enter uncoded text data directly into precoded fields and to subsequently permit direct production of either hard or soft versions of the database as required.

We are indebted to Dr. W. A. Warr for some of the references and are grateful to Dr. L. W. Boyle (at that time, Fraser Williams Scientific Systems Ltd.) for index sorting, to Stella Cardus of Desktop Design for screen designs, and to the editorial and production staff at Butterworth-Heinemann Ltd. for the support and encouragement enjoyed during the past 2 decades of this project. We look forward to more progress in other media during the next 2 decades. Watch this space!

Both hard and soft versions of the database, and site licensing for the latter, are available from Butterworth-Heinemann Ltd, Linacre House, Jordan Hill, Oxford OX2 8DP, U.K.

REFERENCES AND NOTES

(1) Tuey, G. A. P. Chromium Trioxide Oxidation Risks. *Chem. Ind. (London)* **1948**, 766.

192 *J. Chem. Inf. Comput. Sci., Vol. 33, No. 2, 1993*

BARKER ET AL.

(2) Peter, R. C. Chromium Trioxide Oxidation Risks. *Chem. Ind. (London)* **1948**, 797–8.

(3) Dauber, J. G. Chromium Trioxide Hazard. *Chem. Ind. (London)* **1964**, 973.

(4) Bretherick, L. Chromium Trioxide Oxidation Risks. *Chem. Ind. (London)* **1964**, 1196.

(5) Mellor, J. W. *Comprehensive Treatise on Inorganic and Theoretical Chemistry*; Longmans Green: London, 1941–1971; 16 Vols. and some Supplements.

(6) Sidgwick, N. V. *The Chemical Elements and their Compounds*; Oxford University Press: Oxford, 1950.

(7) Rodd, E. H., Ed. In *The Chemistry of Carbon Compounds*; Elsevier: Amsterdam, 1951–1962; Vols. 1A–5 (2nd ed. begun in 1979, not yet complete).

(8) Muller, E., Ed. In Houben-Weyl's *Die Methoden der Organischen Chemie*, 4th ed.; Georg Thieme: Stuttgart, 1953 to date; in 16 occasional multi-part volumes.

(9) *Inorganic Syntheses*; New York: McGraw-Hill, 1939–1989; Vols. 1–25.

(10) Brauer, G. *Handbook of Preparative Inorganic Chemistry*, 2nd ed.; Riley, R. F., Transl.; Academic Press: London, 1961, 1965; Vols. 1, 2.

(11) *Organic Syntheses*; Wiley: New York, 1941–1988, Collect. Vols. 1–6.

(12) *ABCM Q. Safety Summ.* Association of British Chemical Manufacturers: London, 1930–1964; later *CISHC Chem. Safety Summ.*, Chemical Industries Association: London, ceased publication in 1986.

(13) *Case Histories of Accidents in the Chemical Industry*; Manufacturing Chemists Association: Washington DC, monthly publication 1950–1978 (publication ceased).

(14) BS 2474: 1965. *Names for Chemicals Used in Industry*, British Standards Institution: London, 1965 (amended 1970, revised 1983).

(15) Bretherick, L. *Handbook of Reactive Chemical Hazards*; Butterworths: Borough Green, 1975.

(16) *Comprehensive Inorganic Chemistry*; Bailar, J. C., Emeleus, H. J., Nyholm, R. S., Trotman-Dickenson, A. F., Eds.; Pergamon: Oxford, 1973; 5 Vols.

(17) *Nouveau Traite de Chimie Minerale*; Pascal, P., Ed.; Masson: Paris; and 1970; Vols. 1 and 15.5.

(18) Duch, E.; Grewer, T. Thermochemistry of Exothermic Decomposition Reactions. In *Proceedings of the 4th International Symposium on Loss Prevention and Safety Promotions in Process 8 Industries*; IChE: Rugby, 1983; Vol. 3, pp A1–A11.

(19) Yoshida, T. *Handbook of Hazardous Reactions with Chemicals*; Tokyo Fire Dept.: Tokyo, 1980. Though the brief text is in Japanese, most of the book (tables of binary mixtures of chemicals with calculated percentage composition for maximum exotherm and its size, with

assessment of the probability of ignition or explosion) also has English headings and arabic numerals, so is readily comprehensible.

(20) SGML BS 6868: 1987. British Standards Institute: London, 1987.

(21) Bretherick, L. *Bretherick's Handbook of Reactive Chemical Hazards*, 4th ed.; Butterworth-Heinemann Ltd.: Guildford, 1990.

(22) *Encyclopedia of Chemical Technology*, 3rd ed.; Mark, H. F., Ed.; John Wiley: New York, 1978–1983; 22 Vols. Its conversion to the online version is described by Johnson, E. W.; Kutz, M. P. Online Publishing: Producing a Full Text Encyclopedia. *Elect. Publ. Bull.* **1984**, *2* (1), 17–19.

(23) *Dictionary of Organic Chemistry*, 5th ed.; Buckingham, J., Ed.; Chapman and Hall: London, 1982; 5 Vols. This became available online in 1986 as the combined database HEILBRON, also including the text next below.

(24) *Dictionary of Organometallic Compounds*; Buckingham, J., Ed.; Chapman and Hall: London, 1984; 3 Vols. This became available online in 1986 as the combined database HEILBRON, also including the text next above.

(25) *Beilstein's Handbook of Organic Chemistry*; Springer-Verlag: Heidelberg, 1960–1979; Supplement V and Vols. 17–22. The full bibliographic details for this multi-part 27-volume series (comprising over 350 volumes) are too complex to include here. The online database BEILSTEIN as yet covers only part of the total data content, but will steadily be extended. The exhaustive validation and editorial review procedures used hitherto to prepare the text have caused a delay of several years between data publication and appearance in the printed text.

(26) *Gmelin's Handbook of Inorganic Chemistry*; Springer-Verlag: Heidelberg. The full bibliographic details of this multi-part 71-volume work (comprising over 600 volumes) is too complex to give here. So far the online version contains the Formula Index and catalog of the whole work for rapid location of where in the printed volumes the data may be found. The text is being added currently.

(27) Keith, L. H. *Instant Chemical Protective Material Selection*; Instant Reference Sources Inc.: Austin, TX, 1987. This is a combination of the contents of a reference book (Forsberg, K. *Chemical Permeation and Degradation Database and Selection Guide for Resistant Protective Materials*; Royal Institute of Technology: Stockholm, 1986) and an expert system GlovES. The latter is described by Keith, L. H.; Johnston, M. T.; Hudak, C. E.; Conoley, M.; Walters, D. B.; Prokopetz, A. T. An Expert System for Selecting Protective Glove Materials. *Polym. Mater. Sci. Eng.* **1988**, *58*, 16–20.

(28) Whitbeck, M. Bretherick's Reactive Chemical Hazards Database. Version 1.00. *J. Chem. Inf. Comput. Sci.* **1992**, *32* (4), 395–397.