UNITED STATES DEPARTMENT OF COMMERCE
National Bureau of Standards
Gaithersburg, Maryland 20899

May 23, 1985

Mr. Everett A. Wick
High Level Waste
Licensing Management Branch
Division of Waste Management
Mail Stop SS 965
U.S. Nuclear Regulatory Commission
Washington, D. C. 20555

Dear Mr. Wick:

The enclosed draft report on the HLW Database Requirements Analysis will require considerable input before it will be regarded as final, for reasons cited in the report. I would hope that within six weeks we may finalize it. An optimistic estimate of the time required for making a hardware recommendation is about three weeks.

We note that the information on the HLW package is a collection of documents that will include a database that is stored in computer memory to aid users in rapidly accessing and evaluating the information in the documents. Some might regard the database to be the collection of all documents on the subject, and in the attached draft report the database is only that contained in computer memory.

The draft report indicates that more precise definitions of the information to be stored must be developed, and that the combined efforts of NBS and NRC workers will be required to complete this task in the weeks ahead.

Sincerely,

C. G. Interrante

Charles G. Interrante
Corrosion Group

Distribution:

NMSS PM
WM Div. Director
WMEG Branch Chief
WM Document Control Center [4 copies]
Ofc of Research
Off. of Resource Mgmt, Div. of Automated Services (Attn: Director)

DRAFT REPORT

SUBJECT: HLW Database Requirements Analysis for FIN-A-4171

SUMMARY

A computer-assisted database is being developed for storage and retrieval of HLW data composed principally of reviews and assessments of pertinent reports, principally those generated by DOE. Initial suggestions by materials scientists at the NBS on the type and form of information to be stored in the database have been reviewed by NBS computer scientists, who have concluded that available software packages must be studied to find a database management system that is most suited for this application. That study is now in progress at the NBS. Details on software development were not discussed in the works furnished to NBS from a previous NRC-sponsored "HLW Data Base Requirements Analysis." After the most suitable software for this application has been selected, hardware that is compatible with existing NRC equipment will be suggested. Then, the format for storage and retrieval of information for this database will be established.

BACKGROUND

An NBS review of the available information on the high level waste (HLW) package has been initiated, and the requirements of a suitable HLW database system for automated storage and retrieval of NBS reviews of pertinent documents are now being developed. A computer database can be optimized only after understanding the application and the details of the data that are to be contained in the database. Therefore, this continuing review is most important to the development of the database.

An analysis of the requirements of the database was initiated at the NBS after an initial exposure to selected HLW documents. The documents included two memoranda dated September 20, 1984 and November 14, 1984 from T. Sullivan (BNL) to E. A. Wick (NRC) on the "Data Base File Structure" and "HLW Data Base Requirements Analysis."

The above memoranda were reviewed and assessed by materials scientists at the NBS, with the intention of determining whether or not the proposed structure for the database files and requirements for a HLW database were adequate for the intended purposes. This earlier study was not endorsed because the software had not been adequately addressed, and the NBS workers view the software and the structure of the files as a set of problems that must be addressed together in the development of an effective computer database. One way to proceed would have been to buy the hardware and then write software to meet the needs dictated by the hardware and the file structure chosen by the materials scientists. This approach is implied in the memoranda and is regarded by NBS workers as likely to be both inefficient and costly, both in the initial implementation and in the later maintenance of the system. Therefore, it would be taken only if existing software can not be found.

PLANNING A DATABASE

The establishment of a computer database for the HLW application is expected to be an iterative process that begins with a familiarization with the overall problem and the existing information. The most important part of this is to determine the types of information that will be stored and the types of queries that will be conducted on the stored information. Together with the form of reports to be prepared these are the basic requirements of the database system.

Next, the available software and hardware are studied to select packages likely meet the requirements. First, the software is chosen, and then the hardware can be specified. It is understood that the hardware must be compatible with existing NRC microcomputers.

The available software packages must be evaluated for their capabilities in relation to the requirements dictated by the types of information to be stored, the nature of queries, and the form of the reports. Further, both flexibility and speed of the database system must be considered.

It is the NBS judgement that development of software (from scratch) is far too costly both for initial investment costs and for maintenance through the first two or more years of usage. Therefore, a commercially marketed database management system (DBMS) would be our first choice, provided that one is found that meets the requirements for this application. While a commercial DBMS generally can be expected to require computer professionals to write some software to initially adapt it to unique needs of the database system, the level of effort is minimized by the judicious selection of a DBMS for a given application. Further, changes in the software required to maintain the database system are much less demanding when a DBMS is used, and this maintenance software can be written frequently by personnel with lower levels of computer expertise.

After the hardware is selected and the software is tailored to the specific needs of the information, the data is entered, stored, and retrieved to complete the first cycle in the process of building a database. The data, taken from NBS reviews of selected documents, will have been formatted in the context required by the structure of the database.

This initial implementation will lead to additional requirements and realizations that will result in changes in the structure. Fields will be added; deleted; and changed in name, size, or type. Decisions for major restructuring of the database may even be in order, during the first few cycles. Thus, the software continues to be modified until its form has been tailored to meet all of the user needs, which themselves are not well understood until the system has been in operation for some time.

STRUCTURE OF THE DATABASE

Each NBS review of a document may be actually stored in one or more files, but for simplicity in this discussion, a file will be regarded as containing all stored information for one document. The file shall be constructed so as to facilitate computer storage and retrieval as reports prepared in an agreed upon format suitable for reports for use by the NRC.

More importantly, the stored information must be queried to satisfy a user's request for information. Normally, this query will be made to determine which of the available literature on HLW is suitable for use in decisions made by the NRC.

As the requirements have been evaluated by NBS in only a preliminary way at this time, it is expected that they will be firmly established only after the system has been in operation for some time.

After the appropriate documents have been identified by the normal query, the query furnishes a list of applicable documents. The user may opt to read the full review of each document or to use the review along with the detailed data contained in the original full document.

The preliminary conclusions of NBS materials scientists is that the HLW database should contain two parts. One is an index, which is used for making queries; the other will contain text. The text contains the abstract and the reviewers comments on the data, methods, etc.

Detailed numerical tables of data are believed to be unwarranted in the database itself. Other numerical data, such as temperature of test, temperature range over which the data may be applicable, pressure, various mechanical and physical properties of the material tested, etc., would be included in the index. Many of the index fields would not be numeric entries. These would include: Title, keywords, materials, enviornmental and other test conditions.

The index information are stored in various fields that are accessed by user options, in the normal query. The text fields will be available for "manual" queries, which may be conducted by either the user or the manager of the database system. If the DBMS will support large text fields in a way that permits an optional search of the text, that would be regarded as welcome enhancement, but not as a requirement of the system. This simplification of the demands made of the system may facilitate the use of microcomputers for this application. If minicomputers were used this limitation would not be considered, as the minis gererally have more capacity and speed.

One important conclusion is that within the index there should be three types of fields: One-string fields, repeating fields, and text-like fields. While some index data can be stored in fields containing only one word (or one string of characters that can contain #, $, or even spaces), numerous other fields will contain a number of words, perhaps more than 20. Both types of fields must be used to satisfy requests for information in queries of the database.

Examples of the former category include the Identification Number of the document (e.g., CAT001.DOE, or CAT034.MCC), Portion of the Waste Package (e.g., canister, or waste form), and Temperature (e.g., 25C, or 100C). This last item typifies a category that presents a special problem. There may have been measurements taken at numerous temperatures and this would require a different storage and searching system than that which is used for one word, or one string like "waste form."

Examples of fields that may have multiple entries follow:

Material (e.g., alloy steel, 1020 C-Mn steel, or any of a number of ferrous and non-ferrous materials that were tested);

Authors (e.g., Brown, A. W., Avery, R. R., Spike, G. T.)

Environment (e.g., brine, groundwater, and NACE standard H2S solution)

Title (e.g., The Near-Field Waste Package Environment in Basalt and Its Effect on Waste Form Releases)

For fields like those for Material, Authors, and Environment, the examples indicate that numerous entries, each separated by commas, may be required. These are examples of fields with repeating groups, i.e., the field may be repeated an indefinite number of times.

For a field like Title, the information is similar in structure to that which would be given in a text or comment field. To search this type of field requires "substring" search capabilities. A substring can be nearly anything, including one of the items listed in a repeating group.

Thus, for the types of data expected in an HLW database, fields may be one-string, repeating groups, or text. The DBMS must be able to easily conduct the query for each of these types of fields, so as to facilitate data entry, retrieval, and reporting. The expectation is that the report may have a format that differs significantly from that of the data entries.

One concern of NBS computer scientists who reviewed the requirements as set forth by the materials scientists is that DBMSs designed for microcomputer applications usually are designed to handle only the "one-string" type of data entry, and this could prove to be very inefficient for the HLW database application. This may be an oversimplification, and the problem of selection of software involves more than meets the eye of the novice.

SOFTWARE FOR A DBMS

Software developed for microcomputers is preferred for this application, as in general it is more user friendly than that developed for either the minicomputer or the mainframe computer. In the context of this database, user friendliness translates into lower costs for initially establishing and for later maintaining the database system, with the software costs being kept to a minimum if the microcomputer can be used to do all or most of the storing and processing of the data.

Microcomputers have traditionally been used to process the one-string type of data entry. Microcomputers can also be used for the two other types of data described above, but the DBMS must be structured in ways that satisfy these special needs, and traditionally this type of software has been developed for use on larger computers. Software for use on minicomputers and mainframe computers have traditionally been developed to process information stored in "hierarchical and network" structures, whereas the structure of the one-string entry is a "relational database."

Information structured hierachically can be regarded as being in the form of parent and child. For example the parent category materials can have more than one child, in the hierarchy scheme; a child may even have more than one parent, in the "network" scheme. Thus, the software or design of the database is determined by knowledge of how the database is to be used, and software for larger computers have better features.

With the relational approach, the data are stored as tables of rows and columns, and the software must be designed to access that data in a manner consistent with the intended usage. This may sacrifice efficiency (speed) in accessing the data, although the task may be accomplished at lower cost due to the use of the microcomputer. The best known relational software packages were written for microcomputers.

At the NBS computer specialists who have expertise on database systems are still working to find the most suitable DBMS for the HLW database. To date these workers have identified only one DBMS with good apparent potential for this task. It is named REVELATION and is currently being used at the NBS, on an IBM PC-XT, in the Occupational and Health Safety Division. Other DBMS candidates for the HLW database are named DAFLO, and SCI-MATE. All three of these systems will be evaluated in terms of the requirements set forth above.

HARDWARE FOR A DBMS

After a DBMS has been selected, details on the structure of the database can be established through the collaborative efforts of software specialists and materials specialists, as the first iteration of a working system is completed. Further input from NRC workers will be needed to complete this task.

One of the principal guidelines set forth for this HLW database is that the system must be compatible with the IBM PC, IBM PC-XT, and COMPAQ microcomputers, which are currently in use and being purchaced by the NRC. It is noted that while minicomputers are more powerful and while they may have significant advantages for this application in terms of available software, capacity, and speed, the problem of interfacing them to the microcomputers could prove to be costly. Further, the initial investment in microcomputers would be higher as well. With this in mind the discussion that follows on microcomputers was prepared.

COMPARISON of the IBM PC-AT and PC-XT

The IBM PC-AT and IBM PC-XT are prime candidates for this application because of their compatability with the current NRC hardware.

The AT has the potential for much greater speed of proceesing and therefore shorter times may be required to make complex queries of the database. This speed might be available only when the DBMS has been written to take advantage of the higher-speed chip that is resident in the AT. Generally, speed is three to five times faster for the AT than for the XT. Further, code written for the XT can be used on the AT at a speed at least equal to that obtained with the XT, as the AT can be operated using the disk operating system (DOS) of the XT.

We note that the marketplace motivates those who develop software (DBMSs). Software developed for the XT can be modified (enhanced) to run faster on the AT, so even if an AT version of PC-XT compatible software were not available at the time a DBMS is first purchased, an upgraded package is likely to be made available after the system has been in operation. This may be an important consideration as the database grows and the number of records is increased to levels that require greater times for conducting a search.

With this in mind, Table 1 has been prepared to compare the relative costs and some of the limitations of each system. It indicates that the cost of the XT may be of the order of 1200 to 1900 dollars cheaper than the AT, depending on how the random access memory (RAM) is configfured. At the level of 540Kb RAM, the 1200 figure applies; and at the 640 Kb level and with the addition of faster processing in the AT, the 1900 figure applies. A RAM of 640 Kb is supported by some DBMSs, and this is the limit of the capacity of the XT configuration.

The AT can be expanded to 3 Mb with IBM furnished boards listed in the table and to 20 Mb with another RAM board already advertised by another manufacturer. Higher RAM may greatly increase speed of processing for a database. Thus, the expandability of RAM is an important consideration for the long term usage of the database.

There are other advantages to purchassaveing the AT over the XT. These include the fact that it can be upgraded to use a much more powerful operating system called ZENIX, which is not available for the XT. The AT is designed to be a multiuser machine, i.e., more than one user may operate the system at the same time, and this too may become an important feature should the need arise to use the hardware in this configuration. In addition, the AT contains a multiplicity of slots for peripheral devices that may become desirable for addition to the database system later.

Thus, the AT has numerous advantages over the XT for use with a database, but the question of software must be first decided before the hardware question can be addressed.

When the hardware is selected, and both the AT and the XT are attractive candidates at this juncture, the next task is that of structuring the software to meet the needs of the HLW database, and for that a detailed description of the file structure will be developed, as discussed above.

TABLE 1. Hardware Costs

Microcomputers	IBM-AT	IBM-XT
	(dollars)	
PC-AT version includes: 20 Mb fixed disk, 1.2 MB floppy disk, 512 Kb RAM	3941	
360 Kb floppy disk (required to write disks for a standard PC or a PC-XT)	289	
PC-XT version includes: 10 Mb fixed disk, 360 Kb floppy disk, 256 Kb RAM		2649
384 Kb RAM in quadram board (required to bring RAM total to 640 Kb, which is the capacity of the PC-XT)		384 *
CRT Monitor, high resolution	187	187
Monochrome display board	170	170
Printer: IBM Quietwriter- report quality, dot matrix	947	947
Cables, for printer	31	31
Totals	5535	4368

Optional Equiptment	IBM-AT	IBM-XT
	(dollars)	
Cable- RS 232 for peripherals like modem	44	44
1200 baud Modem, telephone communications:		
External to microcomputer- (est.)	420	420
Internal microcomputer card- (est.)	380	380
Additional Disk Drives (external to main cabinet):		
Option with 10 Mb	813	813
Option with 20 Mb hard disk plus another floppy disk drive	1085	

Further expansion of RAM is available only for the AT, as 640 Kb is the capacity of the XT. Alternatives available for the AT include:

3 Mb RAM (2 Mb board for $3196 plus piggyback board for $1596	4792	
1 Mb quadram board	1972	

ITEMS THAT MAY DECREASE PROCESSING TIME
(depending on the requirements of the DBMS)

128 Kb AST Advantage (to increast RAM of the AT to 640 Kb)	471	
Math coprocessor (80287 chip)	255	

---

* SPECIAL PURCHASE NOT ON GSA SCHEDULE.