XC MEMO
OBAR
FISHER
GIGRICH
GOTTSMAN
DATA CORP

JEROME S. RUBIN
606 Indian Field Road
Greenwich, CT 06830 -7239

copy to Tunick

chpt 8
10

February 4, 1998

Mr. Charles Bourne
1619 Santa Cruz Avenue
Menlo Park, CA 94025

Dear Charles:

I enclose some comments on the parts of your manuscript of which I have personal knowledge.

In addition to the specific comments, I have two general observations:

1. You appear to have received disproportionate input from the Data Central and OBAR people and inadequate input from the *LEXIS* creators and veterans. This may have been a factor in your not appreciating the sharp break from the past that *LEXIS* represented.

2. In your discussion of *LEXIS* you pay little attention to the successful business model that accompanied the brand-new system.

If you want any explanation or clarification, please let me know.

Very truly yours,

Jerome S. Rubin

Chapter 8

p. 3, etc. It is inaccurate to suggest that *LEXIS* was based on OBAR. It was a wholly new system, designed by me and my colleagues without any participation by OBAR or any Ohio lawyers and, except for some trivial numeric code (that was soon replaced), without a single line of code from the OBAR system.

p. 36. The inadequacy of the OBAR/Data Central system is grossly understated. With more than one user online or a search of any complexity, a search could run 5 or 6 hours or even more.

p. 38. I would not regard OBAR's use as large-scale, and its user interface was abominable.

p. 41. Carl Fisher's comments are correct. When I took over as president of MDC at the beginning of September 1971, I fired Giering and some other underperformers we had inherited from Data Corp. and hired Ed Gottsman (a member of the ADL team) to direct the development of an entirely new system (which I later christened *LEXIS*). Bob Bennett and I designed the new system, and Bennett, with my kibitzing and amending, wrote the functional specs. Gottsman imposed some feasibility criteria on the specs and took charge of building the system. The new system was easy to use, it was reliable, and it had acceptable response times. Incidentally, it incorporated many features that Giering had said were technically impossible.

p. 45, etc. We always used MDC, not MDCI. And we did not refer to the system as the Mead system. Almost everyone who developed and launched *LEXIS* came from outside Mead, and everyone regarded Mead as out of sync with our efforts.

p. 46-48. It was I who renegotiated the Ohio Bar contract and negotiated all the other sponsorship agreements, including the NCAIR agreement.

p. 48. I personally coined the name *LEXIS* (and later *NEXIS*), and it was in fact derived from LEX and IS for information service. Whoever gave you the consultant story got the facts mixed up. When I coined the name *LEXIS*, however, I was aware of the memorability and design advantages of names containing an x or two; the firm that coined the names Xerox and Exxon had been a client of mine when I was practicing law. That firm was **not** the firm that produced our logo (Helvetica bold italics, with a split x). The logo and all our promotional and training materials were done by George Nelson's industrial design firm, with Dan Lewis as the principal designer for everything but the logo (which Nelson personally sketched out in my presence late one Friday afternoon). Lewis also designed the award-winning UBIQ terminal.

p.60, etc. I was not added to the ADL study by Wilson. The strategic planning group at Mead asked me to participate. More important, the study went beyond saying that there was a need for "extensive redevelopment of the software." In the oral presentation to Mead (if I remember correctly, in February 1970), ADL told Mead that it had nothing but an idea, that the OBAR/Data Central software was essentially worthless but that appropriate software could be developed, and that there was the potential for a significant business.

p.63. I took the helm at the beginning of September 1971, and spent the next month or two cleaning house, bringing Ed Gottsman on board, and establishing some independence from the Mead Corporation. Bennett and I designed the system during the winter. The functional and performance specs went through several iterations as Bennett and I worked closely with Gottsman, Dana, Byruck and Thomson on the technical side, and by late summer of 1972 I approved the definitive specs and turned Gottsman et al. loose to do the coding. During this same period, I also defined our business plan and our pricing and marketing strategy and started sponsorship negotiations with state bar associations and NCAIR. At the same time, Bennett and I defined the contents and structure of the data bases (or, as we called them, the libraries), began the massive job of accurate data conversion of New York and Federal libraries (unlike the OBAR data, which were unacceptably dirty), and designed the training program (including writing the material and making a video).

Chapter 10

p.4. It is a misleading overstatement to say that the Data Corp. system spawned *LEXIS*.

p. 77. There was little or no "upgrading and expanding the OBAR system" in 1971 and 1972. Bennett and I (and some consultants, including Gottsman of ADL and Prof. Tony Oettinger of Harvard) were unable to persuade Don Wilson to scrap the OBAR system and start over from scratch. In the late summer of 1971, Mead insisted that I take over as president and undertake the effort Bennett and I had been urging. Because of my partnership with Wilson, I was reluctant to do so, but Mead said that the alternative was to shut MDC down and Wilson encouraged me to accept. Once I became president (Sept. 1971), we were not "upgrading" OBAR; we were building an entirely new system and service. *See* comments on Ch.8, p. 63.

On the origins of the name *LEXIS*, see comments on Ch. 8, p. 48.

p. 78. You cite Giering for the assertion that some of the "behind-the-scenes software" of *LEXIS* was identical to the Data Central system. Basically untrue. See comments on Ch. 8, p. 3.

Hardly any Ohio firms were subscribers to *LEXIS* at the outset. Those who "had hung in through thick and thin" had paid almost nothing for OBAR. We imposed a significant minimum subscription fee ($36,000 a year) for *LEXIS*, and most Ohio firms, soured by the OBAR experience, were not prepared to pay. Large New York firms, who knew me personally and had not suffered through OBAR, were willing to roll the dice on the new service. Until word got out that major New York firms were making good use of *LEXIS*, the bad odor of OBAR was our principal marketing obstacle.

p. 79. While the computer center, software programming and back-office operations were in Dayton, the corporate headquarters (plus marketing, sales, communications, training, planning, etc.) were in New York from the day I became president until some time after my team and I left.

p. 80. Courts issue opinions with syllabi in only a handful of states. Headnotes are written by West, are covered by copyright, and were not included in *LEXIS*.

To the best of my recollection, *LEXIS* was launched on April 2, 1973, but there was a partial waiver of charges for the first month.

p. 81. Fisher is correct on the death of color.

The only reason for the special terminal was ease of use. Moreover, there were very few computer terminals in law offices in 1973. The suggestion that we wanted to exclude the use of other services is ridiculous.

p. 82. Fisher is correct.

p.83. Don't confuse Data Central and *LEXIS*.

p. 84. As I pointed out in my IIA Hall of Fame acceptance speech, Bennett and Gottsman played central roles. Fisher also played an important role, but Heilesen did not. If J. Sperling Martin played a role, he must have used an alias; I have never heard of him.

p. 172-173. The *LEXIS*-JURIS story is more complicated than your version. *Inter alia*, it was an element in the antitrust litigation against West.

cc IBM-PRIME
MAGNINO
IBM-ITIRC
IBM-CIS
IBM-STAIRS
GIERING
SEVE
DAMERAU

1

4/27/98 Comments from Sam Kaufman

Chapter 3 Page 124

Your comment that "PRIME-2" was specifically mentioned in my 1966 paper is not correct. I have carefully reread my reprint of my 1966 paper and found only two references to PRIME, (not PRIME-2), both in column 2 on page 506 of the Proceedings. I quote "In an experimental normal text system called PRIME, each document was built....etc." and "It was recognized that PRIME would be too slow for extremely large data bases". I found no other references to PRIME in the paper. The paper only referred to a 7090/1401 system. I've enclosed a copy of that page.

I used " PRIME-2" in my communications to emphasize that it was different from the original PRIME. I found an internal document, dated 10/63 , of operating formats for input and output, including tape formats and other information, whose cover page was entitled PRIME-2. I believe that the name PRIME-2 vanished with the establishment of ITIRC.

PRIME was an experimental working test system and we knew that it needed to be improved. As you may note from the 10/63 date above that we were well along with the "PRIME-2" version for retrospective searching then. The system described in my 1966 paper which also included CIS, was not released until it had been thoroughly tested not only by our group, but also at our European satellite in LaGoude, France. It also was tested by an interested corporation.

Chapter 3 page 123, line 3 of PRIME

Joe Magnino was not director of IBM's ITIRC in 1961 because ITIRC was not created until 1964. Joe was Manager of Technical Liaison at the time.

I must correct an impression that you seem to have. When ITIRC was established, "PRIME-2" was fully operational as a data processing and retrospective searching system. After ITIRC was established, I designed an additional set of programs for the CIS requirements that would use much of the operating system and add the CIS capability. All of the data from the merging units had to be converted to the existing data input specifications in a manner similar to the purchased data bases. The retrospective search capability was not altered, and CIS and Retrospective searching operated independently. Somewhere there was a comment that PRIME-2 was folded into CIS. Not so.

Chapter 5, Pg 78

Kalenich and Esposito were not involved in the design or development of TEXT-PAC!

Esposito's primary responsibility was machine operations. He had written one or two 1401 programs for PRIME. He had no part in the design of development of any program. Esposito was assigned to write one program for TEXT-PAC after which he went back to machine operations. He spent most of his career as a computer operator and eventually as a computer room manager. The reason his name appears in the Type III release is that I listed the names of all of those who wrote any program for TEXT-PAC in alphabetic order regardless of degree of contribution to the project. In fact, I was the primary author and editor of the program release documentation. I also had responsibility for arranging an independent internal beta-test to determine if all of the programs could be run with the supplied test data and operating instructions without problems. IBM was not responsible for any problems with either the released 7090/1401 system or TEXT-PAC. It was my responsibility and up to me to resolve it. The TEXT-PAC release included the source code, which made it easy to send any correction necessary to the user.

Kalenish was an information retrieval specialist, (a former librarian) who was involved in the development of the first experimental system and in PRIME-2 in helping set bibliographic format standards for data input and output and with query logic. He was also an expert in creating queries for searching as well as in testing the system. He was primarily responsible for development of project file data and conducting searches on that data base.

Joe picked the name TEXT-PAC, as you suspected, from TEXT PACKAGE. I think that there is a nice rhythm to the word TEXT-PAC.

Chapter 5, Pg.79

PRIME-2 retrospective search program was not incorporated into CIS. The CIS capability was incorporated into the PRIME-2 system.

Questions 5 and 9

PAGE 81.

TEXT-PAC, Terminal TEXT-PAC and STAIRS did not have proximity searching with N words separating the search words. I don't know whether proximity searching was limited to a sentence. If not, it could lead to spurious answers. My personal feeling is that proximity within N text words would mean anticipating text content or word patterns.

We used WITH logic to find words in the same sentence regardless of order or position. We used ADJ to search for contiguous words (strings). I believe that ADJ logic was limited to a sentence. If it went beyond a sentence boundary, it could lead to a false hit. TEXT-PAC, Terminal TEXT-PAC and STAIRS. in addition had the capability of limiting the search to a specific text unit in a document, such as title or author or any other unit. I believe that this capability may not have existed in other systems.

Question 6. Neither Joe nor I had heard of Dick Giering or his work before you asked this question.

Question 8 Pg 86 I must decline the credit you gave me for the Aquarius/STAIRS tutorial/help capability.

I recently was able to discuss this question, via phone, with both Steve Skye and Stan Friedman who were key people with STAIRS. Each of them independently stated that it was part of the original system analysis and design before programming was started. The rational for an on-line tutorial and help ability was stated in Skye's section on STAIRS PG 2 "Given the time constraints of getting the entire litigation defense effort underway quickly, we realized that there would be little time for extensive user education. This prompted us to incorporate a combined tutorial/help facility from the very start".

I agree with your assessment that it was the first such capability.

Question 10 The ranking feature was optional and only activated by the user after the search was completed.

In our recent phone conversation, Stan Friedman told me that he joined the TEXT-PAC group in 1966 near to or at the its start, so your 1965 date is probably correct. I noted at the end of my 1966 paper that we were planning a 360 system. Work had already started on it by the time Stan joined the group. I found in a copy of 1967 accomplishments that programming on the TEXT-PAC was complete and operational and was also installed at our satellite operation at LaGoude.

Cathy Harlin was the third key programmer who worked on STAIRS.

Chapter 4 pg 78 line 13

Joe Magnino's recollections are a bit off the mark. I discussed this bit with him. I believe Joe was thinking about the original PRIME. I joined Magnino's group on April 15, 1961. There was a 650/705 prototype operating at the time which could only process 5 character words.

On the basis of modest success, it was decided to have Service Bureau do a systems analysis and write a program to run on a 704/705. The computers were in the IBM showroom at the old 590 Madison Ave. headquarters. It was using this system that the question of the number of text characters was optimum for searching was raised. There were console switches which made it possible to search on 6,12, or 18 characters (1,2 or 3 704 storage words). Not surprisingly the tests resulted in chosing 18 characters. From this beginning, Service Bureau, wrote what became the first 7090/1401 PRIME system. Much of the first 7090 search program was based on the 704 system. This was going on about the time I started with Magnino's group.

The required 1401 programs were also under way in 1961. As we tested and gained experience with this system we added more input data processing capabilities to the 1401. Beyond normal data checks, we incorporated a spelling check capability, probably the first in use. As Joe mentioned, it seemed pitifully slow in the first version. It used linear matching of text against the spelling list. We improved the process by formatting the list and using a binary search method. We must remember that it was a batch system not an interactive one, and that the results had to be printed in a format that listed the potentially misspelled words in a column next to each of the documents together the paragraph number, line and word numbers, to facilitate corrections. The original spelling list was compiled by the information retrieval specialists (Kalenich and others).

In 1962 Fred J.Damerau made available to us, a word frequency distribution 7090 program which could use any 7090 compatible text data base and compile an alphabetized list of all words in the data base together with the number of occurrences of each word. (we named it the OMAHA program). We used this on all of our data bases to not only create a technical spelling list, but to find any misspelled words that might be imbedded in the data. Words that only appeared once or a few times certainly were candidates for checking and possible removal. We also used the program on purchased data bases to pick up new technical terms and possibly uncover misspellings. When we talked about data integrity, we did our best to achieve it. A version of this was included with the TEXT-PAC release.

The output from th OMAHA program had another significant use. It could be used to check the search results for accuracy. In testing search program logic, the occurrence of a particular word in documents found had to match the number of occurrences listed as output from he OMAHA run. In TEXT-PAC the results had to be read carefully, but in STAIRS, the word match counts would be shown on the screen,

Chapter 3 pg 126-127
Minor correction: Kaufman, as the person most responsible for the development of technical aspects of the PRIME and TEXT-PAC systems, continued....etc.

"History" of TEXT-PAC Your handling of the citation is a good compromise. Thanks.

TTAM

A TIME-SHARED TELEPROCESSING ACCESS METHOD

by

Richard Giering

October 1968

Data corporation

TTAM - A TIME-SHARED TELEPROCESSING ACCESS METHOD

At the present time there are few programs available to the general users of IBM 360 to take care of their teleprocessing needs. Generally the only modules available are those distributed by the equipment manufacturer-IBM. BTAM and QTAM have some disadvantages, one of which is the fact that the modules do not support most terminal equipment manufactured by companies other than IBM. A second restriction is the cost, primarily in the amount of memory that the modules require. For time-shared applications the distributed modules are not the most efficient since, under certain conditions, the error recovery that the module executes is not a recovery at all but merely a cancellation of the job executing the teleprocessing IO.

For the above reasons, (Data) Corporation, in support of the (Data) Central information handling system, designed and wrote its own teleprocessing access method called TTAM, for Time-shared Teleprocessing Access Method. TTAM currently operates under the standard (distributed by IBM) operating system-DOS. It is being considered for conversion to OS. This means that as newer versions of the supervisor are issued, as long as compatibility with previous versions is maintained, the TTAM will continue to operate without change. Additionally, for the advent of multiprogramming, the TTAM module resides in the partition needing it; this allows for a substantial saving in the allocation of core memory. When the program is not in memory, the TTAM is also absent, making more room available for other processing.

Logically, TTAM is made up of four parts:

1. THE CORE OF THE MODULE: This is represented by a MACRO, made available in the library, to define the teleprocessing line. A single entry in the program defines each line available. This definition includes a line number (1 to 255), a system number (sysxxx), and a definition of the terminal type. From these three pieces of information the assembler builds a block of data representing the line equipment specified. This block of data, itself, is comprised of two parts. First, all the necessary control information together with the channel programs are generated; secondly, a buffer is created (for the actual input-output of the messages) together with a translate table for the first of each device type. These tables allow for the translation of the internal character set, EBCDIC, to the character set of the terminal at the other end of the communication line. For each line defined, the block of defining data includes about 500 character positions of memory plus the size of the buffer (which is defined by the application programs' messages) plus, for the first of each device type, 768 character positions for the translate tables. Currently the module recognizes the following devices: IBM 1050 terminals, Teletype mod 33/35, and the CRT terminal manufactured by Computer Communications Incorporated. The amount of memory required is totally defined by the application program and not by whether the capability is generated into the system supervisor.

2. COMMO & ERROR RECOVERY: Probably the heart of the TTAM modules is the time-sharing rollout/rollin capability. During the assembly of the TTAM module, the applications programmer indicates how much memory is to be set aside for switches and temporary storage. This area also includes space for the definition of the location on secondary storage where files are to be maintained. The user also indicates how much of his program must be retained in its existing state during the actual I/O operation. Once the TTAM module starts a communication to and/or from a terminal, it causes the specified memory area to be placed temporarily on secondary storage until the communication is complete. Once one of the communication lines signals that it has completed its function, the TTAM performs the necessary error

recovery, including the restarting of that line if necessary. As soon as the communication is successfully completed (requiring no error procedures), TTAM then returns the material from secondary storage to main memory, thereby returning the problem program to the state at which it was at the time the communication was requested. Finally, it turns control over to either the conversational-mode message processor for input operations or to the calling program.

3. CONVERSATIONAL-MODE MESSAGE PROCESSOR: This section of TTAM allows the calling program to issue a series of messages (PUTTP), receive a message (GETTP), or perform both operations (PUTGET). In any case, the module does all the translating necessary, does all processing necessary to set up the hardware instructions for the transmission, and executes the input/output.

4. PAGE-MODE OUTPUT PROCESSING: While the above (paragraph 3) section allows for the processing of messages, there is the requirement for the preparation of a series of messages that should be sent as a block, called a page. The TTAM allows for this operation. In this case, the program calls for the opening of the buffer; this makes the buffer available, and the user may enter the information into it. As each line of the page has been entered, the problem program notifies TTAM. This section of the module then translates the material into the terminal code structure, appends the group of data characters with the necessary controls (end of block, etc.), and checks to see if more data can be handled in the buffer. If not, it causes the actual communication to take place prior to returning to the calling program. If more data can be accommodated, control is immediately returned to the problem program.

The TTAM module was designed and written in a general purpose manner. It is NOT a part of the (Data) Central information handling system logically but acts as a separate supporting module for that system. It is believed that, as more devices are defined and as more features are implemented (such as the use of the CRT light-pen), TTAM will be a valuable asset to any programming application.

Presentation to the

Conference on Large Data Bases

May 22, 1974

SEARCH STRATEGIES AND USER INFORMATION

By
Richard H. Giering
Mead Technology Laboratories
Research Park, Dayton, Ohio 45432

NATIONAL ACADEMY OF SCIENCES
National Research Council
2101 Constitution Avenue
Washington, D. C.

Good morning, ladies and gentlemen.

It is indeed a pleasure to address this audience and to discuss what we in The Mead Corporation have found to be certain aspects of search strategy with respect to searching large data bases and, of course, the definition of certain problem areas that remain. Most of the presentation this morning will concern the definition of some problem areas, especially problem areas in the freeing-up of the interface language between the user and the search process. There will be four specific areas in the interface language defined, and a recommendation made concerning basic research into some linguistic meanings for the four areas.

I intend to be very provocative and to stir your imagination during this discussion. I hope that there will be not just a little disagreement over these matters, because as a result of disagreement there will come new ideas to solve the problems I hope to identify. This presentation, then, is in two parts; the second will be a call for some basic research.

Let me start, however, by raising a fundamental question with respect to this conference. This is a LARGE DATA BASE CONFERENCE. Question - what is your definition of a large data base? (Slide 1) I would venture a guess that although each one of us has our own definition of a large data base, there are some subtle differences between those definitions. Let me interject a comment at this point. We at The Mead Corporation are involved with large data bases in many classes, and I have found that the people involved with large data bases of the bibliographic class are far ahead of others in their thinking concerning the use of large data bases in general. But the problem areas for other types of large data bases (large for other reasons) also present some formidable difficulties.

Now, under what criteria can a data base be large? (Slide 2) First, a data base can be large because the number of records or entries in that data base is large. Five hundred thousand entries in a data base represent a large data base, independently of the size of each of the individual entries. A data base containing merely bibliographic information, including a limited number of manually assigned index terms (on the 500,000 entries), could well require storage of a mere 50 to 100 million characters of information. Not a very large amount of storage in anyone's evaluation, but a large number of entries.

The second criterion (Slide 3) for defining a large data base, then, could be the numbers of characters involved. Again, a bibliographic data base with only 10,000 entries in it, but with each entry being the full text of the document involved, represents a data base whose character count is on the order of a billion.

A third measure (Slide 4) against which a data base could be considered large is a measure very important to the search or selection process. If the number of selectable elements, possibly a controlled thesaurus, is on the order of a thousand, independently of the size of the data base in characters or entries, the data base could be said to be small. On the other hand, in a full-text operation, where every potential word and every potential value, including all variants, are search-selectable elements numbering in the hundred-thousand area, the data base could be said to be large, especially if the software involved required a serial search of those elements.

Now, finally, a fourth condition under which a data base could be said to be large, again independently of the previous three, appears primarily when we get away from bibliographic data bases — and I believe that this Conference should address this question — because as bibliographic experts, we have led the field. We have been at the "head of the pack", so to speak. In defining Information Science activities, designers of management systems, for example, are just beginning to think about "large data bases" in the way we in the bibliographic area have been thinking of them for years. This last condition, (Slide 5) under which a data base might be said to be large, would be in the number of fields and/or files associated with the intricate make-up of the individual data base.

We will be addressing all four of these areas - all four of these definitions of "large data bases" - during our discussion this morning. If, during the discussion, there is disagreement between what is said by two people, myself being possibly one of them, or two people within the audience during the question and answer period, it is quite possible that the difference lies in the assumed definition of some of the terms, including this one.

I would now wish to raise another question from a definition standpoint. (Slide 6) Who, or what, is the user? We have been inundated recently with calls for standardization of systems specifications of systems intercourse. There have been a wide variety of complaints about having to learn multiple disciplines of access for multiple large data bases. I wish to propose to the Conference a hypothesis and suggest that it be the subject of some discussion here, not formally, but over cocktails, if necessary. Is it not possible that the reason there is such a dichotomy of user interaction specifications is that the individual definitions of what and who the user is vary? The attributes that are ascribed to this person called "the user" by system-one vary and differ greatly from the attributes associated with this person called "the user" by system-two. Until we have some commonality of understanding, until we agree on the minimum set of attributes for this thing called "the user", we will continue to have a dichotomy of systems.

The understanding and agreement as to what a user is are important in still another area. The on-line interactive use of data bases, large or small, is in its infancy. As the user population grows, it is quite logical to assume that the profile (make-up) of that population will continue to change. Unless we, as an industry, understand this changing phenome-

A second example, already in use, is the AND condition (Slide 18). We are all fully aware of the relationship under standard Boolean logic of the "&" and the "OR" condition. Since there is an implied ambiguity, the expression "A and B or C" can have either of the following meanings (the word "AND" used linguistically rather than formally):

(A & B) or C

A & (B or C)

Classically the defined expression "A and B or C" has the meaning (A & B) or C, which linguistically is "the combination of A and B or the single element C". This is exemplified by the actual conversational expression "Doctor & Patient or Lawyer". A great deal of effort and a great deal of rigorous mathematical definitions have gone into all of the ramifications of this logic. Witness DeMorgan's law in which the reversal of operators is effected by the establishment of parenthetical (Slide 19) nesting of the NOT-operator:

NOT (A or B) is identical to (NOT A & NOT B)

People, on the other hand, do not always think in a Boolean manner. In fact, it is this author's opinion that a great many people think in the reverse form and, without any other specification, the expression "A and B or C" tends, in normal conversational discussion, to take on the meaning established in the second line above, spoken as "the single element A in combination with either B or C". A conversational example is "DOCTOR and HOSPITAL or OFFICE". This definition of logic has never been passed through the rigor of mathematical treatises, except in the use of parenthetical notation. No "not-reversal" (DeMorgan's duality) has been defined, and all of the rigor of the combination of the two has not been defined except by use of actual parenthetical notation. It is <u>believed</u> the "NOT (A and B or C)" would be, based on parenthetical notation, the same as saying "NOT A or the combination of NOT B and NOT C". It is time that we in the Information Science profession applied the rigor of logic to this preceding linguistic expression so that our users, who are not necessarily mathematically oriented, are able to express themselves in their language either way, rather than having our language imposed upon them. Based upon the opinion that both forms of the logical AND need to be defined and used by users, both a superior (to "OR") and inferior AND are definable in (DATA/CENTRAL). In this writing, the word "AND" is considered to be superior while the "&" is considered inferior:

A and B or C means A & (B or C)

A & B or C means (A & B) or C

A second area is one in which there is at least one anomaly in normal logic expressions, as found in general computer-oriented languages. Consider, for example, the two search requests shown in Figure 2:

1.#AUTHOR = JONES AND#DATE ≤ JUNE, '71

2.#AUTHOR = JONES AND#DATE —> JUNE, '71

Figure 2

(Notethat the pound sign refers to the field in which the conditions are to be satisfied, while the operators are defined as, for this example, ≤ meaning "less than or equal to" and the operator —> meaning "not greater than".)

In normal algebra, the two statements have identical meaning! It is possible that in data base systems they do not. Definition of negation: The logical definition (in normal expression) of the negation process requires the removal, from the set of possible answers, of any answer satisfying the positive form of the negative expression - for example:

"Any hotel but not the St. Francis" — means to find a list of "All Hotels" and remove from that list any hotel positively named "St. Francis".

Take a simplified form of the example. The entries are edited such that, on input, the field name DATE can contain only one value — the date of publication. That value, however, may be an actual date or it may be an indication of an unknown date. Please note that an unknown date is definitely not a date of zero time. An entry with an author named JONES with an "UNKNOWN" indication (absence of the date) would not satisfy the first statement; therefore, it would not be an answer. By not having a date, it can not be considered to be not greater than and it would not be removed from the set of answers. Therefore, the only criterion valid to determine the satisfaction to the search is the initial "JONES" criterion; it, therefore, satisfies the second statement.

7. (Field 1 ER Field 2) ⟨ operator ⟩ (Value 1 or Value 2) means that either Value 1 or Value 2 or both Value 1 & Value 2 must appear either in Field 1 or in Field 2 but not in both fields.

8. (Field 1 ER Field 2) ⟨ operator ⟩ (Value 1 & Value 2) means that both Value 1 & Value 2 must appear in either Field 1 or in Field 2 but both values are not to appear in both fields.

9. (Field 1 ER Field 2) ⟨ operator ⟩ (Value 1 ER Value 2) means that either Value 1 or Value 2 (but not both values) are to appear in either but not both fields.

We have run some preliminary studies attempting to define the results of both the positive and negative uses of these three connectors in nine combinations yielding 144 unique expressions. The definition of the negation operator (¬), it is believed, bears repeating: the negation operator, applied to a value, asks for any entry containing the presence of that value to be deleted from the set of answers to which it applies. The negation operator, applied to a field (or segment), asks for the complement of that segment's action with reference to the specified combination of values.

In the study, there were 16 combinations of entries that were evaluated. In Figure 3, the "N" refers to the non-presence of the value in the segment (or field) while the "P" indicates presence. The table was generated, based upon the way an implementation might process the data: Evaluate the righthand portion of request first, but separately for each segment (or field) specified on the left side (independent of the "sign" associated with the segment); reverse, if necessary, for the sign of the segment; and combine, based upon the connector between the segments. Ergo, for entry number 9, for the following request (one of the 144 expressions), the processing is specified in Figure 4.

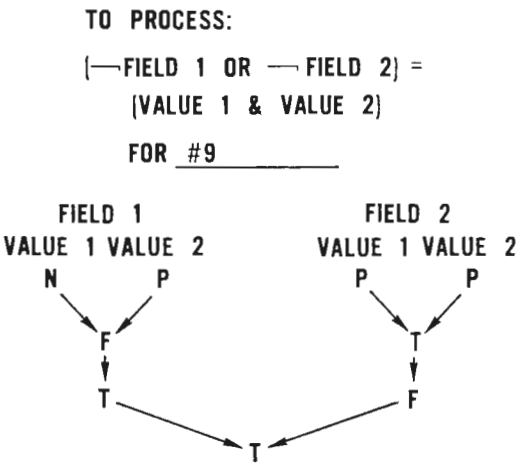


Figure 3

16 COMBINATIONS:

ENTRY	FIELD 1 VALUE 1	FIELD 1 VALUE 2	FIELD 2 VALUE 1	FIELD 2 VALUE 2
1	P	P	P	P
2	P	P	P	N
3	P	P	N	P
4	P	P	N	N
5	P	N	P	P
6	P	N	P	N
7	P	N	N	P
8	P	N	N	N
9	N	P	P	P
10	N	P	P	N
11	N	P	N	P
12	N	P	N	N
13	N	N	P	P
14	N	N	P	N
15	N	N	N	P
16	N	N	N	N

Figure 4

a whole series of "G-connectors" will need to be defined to aid the non-ADP oriented user in expressing a request for the solution to his problem.

For example, the defined GOR is an inclusive concatenation of the fields under consideration. It is highly probable that, as usage of information systems grows, the need for a definition of the exclusive concatenation, referred to as GER, would be required. It could be defined as:

"The inclusive concatenation "MINUS" those entries in which the same value appears in both segments. "

Additionally, the concatenation — intersection (G&) might well be required. In both of these connectors, the location of the value in the multi-valued segment may wish to be used as opposed to the normal "GER" and "G&" where only the presence of the word is used. As a final example, consider the following for a special G-negation:

An application is defined with one logical file containing 78 segments or fields. A predefined pseudo-segment called "SEG80" is defined as the concatenation of Segments 1, 7, 13, 17, and 18.

Thus:

Seg 80 ≡ (Seg 1 GOR Seg 7 GOR Seg 13 GOR Seg 17 GOR Seg 18)

A user, on an interactive terminal, wishes to access the file with the following request:

(Seg 1 GOR Seg 7 GOR Seg 13 GOR Seg 17 GOR Seg 18) = (Word & WD)

It is identical to:

Seg 80 = (Word & WD)

Another user might wish to access the same file except that, for this user, the concatenation should not include "Seg 13". The special G-Negation (G-) would allow this as:

(Seg 80 G- Seg 13) = (Word & WD)

As long as only the GOR and the G- are the only available "G-connectors", the processing formalism is not too complex; once, however, other "G-connectors" are required (especially the GER & the G&), the composite meaning may become most complex.

As can be seen, we've attempted to define four areas (Slide 28) in which additional man-machine conversational linguistic expressions can and, we believe, should be defined to make easier the way our users may communicate with the computer. To define his search strategy (especially as the user becomes more knowledgeable), the user will want ever-increasing capabilities with even more simplified forms of expression until, in the long run, the user will be able to converse verbally in natural language with the computer.

Thank you.

WHAT IS A "LARGE DATA BASE"

? ? ?

SLIDE 1

LARGE NUMBER OF ENTRIES!

SLIDE 2

LARGE NUMBER OF ENTRIES!

LARGE NUMBER OF CHARACTERS IN STORAGE!

SLIDE 3

LARGE NUMBER OF ENTRIES!

LARGE NUMBER OF CHARACTERS IN STORAGE!

LARGE NUMBER OF SELECTABLE ELEMENTS !

SLIDE 4

A AND B OR C

SLIDE 9

A AND B OR C

DOCTOR AND PATIENT
OR LAWYER

SLIDE 10

A AND B OR C

DOCTOR AND PATIENT OR LAWYER

DOCTOR AND HOSPITAL OR OFFICE

SLIDE 11

SEARCH-ONYMY

- BEYOND SYNONYMY
- DICTIONARY DISPLAY
- CONTROLLED THESAURUS
- BROWSING

SLIDE 12

EXCLUSIVE OR (CALLED er)

(A er B) $\equiv$

(A OR B) & $\neg$ (A & B)

SLIDE 17

A AND B OR C

DOCTOR AND PATIENT OR LAWYER

DOCTOR AND HOSPITAL OR OFFICE

SLIDE 18

De MORGAN'S LAW:

$\neg$(A OR B) MEANS ($\neg$A & $\neg$ B)

SLIDE 19

(FIELD 1 { & / OR / er } FIELD 2) =

(VALUE 1 { & / OR / er } VALUE 2)

SLIDE 20

(JOB-HISTORY-EXTERNAL

OR

JOB-HISTORY-INTERNAL

OR

CURRENT-POSITION-DESCRIPTION) =

(MISSILE & AIRCRAFT)

SLIDE 25

WHAT ABOUT:

(JOB-HISTORY-EXTERNAL OR
JOB-HISTORY-INTERNAL) =
MISSILE

AND

(CURRENT-POSITION-DESCRIPTION =
AIRCRAFT)

ETC.?

SLIDE 26

REALLY –

(JOB-HISTORY-EXTERNAL G OR
JOB-HISTORY-INTERNAL G OR
CURRENT-JOB-DESCRIPTION) =

(MISSILE & AIRCRAFT)

"G OR" IS DEFINED AS
A "CONCATENATION" CONNECTOR!

SLIDE 27

1. ADDITIONAL (COMPLEX)
   LOGICAL EXPRESSIONS

2. ARITHMETIC EXPRESSIONS
   THAT DEFY ALGEBRAIC LOGIC

3. MULTI-DIMENSIONAL EXPRESSIONS

4. CONCATENATION & OTHER
   G-CONNECTORS

SLIDE 28

letters to the editor

MORE ERGONOMIC "NO NO'S"

To the Editor:

Enclosed are a couple of pages in answer to your recent cover photo of ergonomic "no-no's" (May 1992 *ONLINE*).

Your answers list caught a number of the errors relating to glare and improper monitor contrast that I missed. But the answer list missed the one about carpal tunnel syndrome. Namely, that Heather's wrists were positioned below her keyboard, and not supported in any way. The stress of typing in this position would soon lead to ligament and muscle damage in her wrists.

The attached pages illustrate a well-designed computer table in which the monitor is both shielded from glare, and set low enough to prevent "document whiplash." That is neck and back fatigue brought on by having to rapidly transfer one's attention between a document on the desk and the document on the screen. It might appear that the designs have limited space for spreading out papers and books, but the company that makes the tables has always said they are happy to design custom systems for their clients.

Becky C. Davis
Librarian
Van Nuys, CA

Astron Computer Products, 2331 Abalone Ave., Suite #102, Torrance, CA 90509; 213/320-1503; 800/634-4884.

CAUTION: ERGONOMIC HAZARDS

To the Editor:

I loved your May 1992 cover and the article on ergonomics. I think the poor searcher illustrating the ergonomic problems has some serious safety hazards to contend with in her work environment. I'm amazed you'd let her get away with propping up her PC on two flimsy software boxes. If her "house of cards" tips over, that overhanging monitor and CPU could easily fall in her direction. And, if I were her, I wouldn't be real thrilled about sitting so close to the assorted cables from my neighbor's PC.

By the way, why does she have so many pencils? Is it for those times when the equipment collapses and she has to, heaven forbid, resort to manual methods of publishing?

Sincerely,
Sophie K. Hudnut

EPIC CAN "STOP THE CLOCK" ON SEARCHES, TOO

To the Editor:

ONLINE's article, "Stop The Clock, Hold The Search, And Take A Break" (May 1992), pointed out the benefits of being able to logoff for a bit, then return to find one's search strategy intact.

I'm happy to report that the EPIC Service, from OCLC, also has this feature as part of its **STOP** command. Here's how it works: When you want to interrupt a session, enter **STOP WAIT**. The system logs you off, with the message, **Retaining your search information.**

The system holds searches for ten minutes, and greets you with the message **Reconnected to EPIC.** Then just enter the **REVIEW** command to look at your search history and proceed.

The "wait" feature works with the EPIC logoff command **STOP** and all its synonyms, too: **logoff, bye, off,** and **quit.**

Tam Dalrymple
OCLC, Manager
Reference Services Marketing
Opportunities and Projects Section

DATA CORPORATION ONLINE OFFERING PRE-DATES DIALOG AND ORBIT

To the Editor:

This letter is in response to your article titled "Maxwell Online At The Crossroads" as published on page 29 of the May 1992 issue of *ONLINE*.

While I found the article to be most insightful, I must take issue with the content of one portion. In the middle of the second column on the first page of the article, you indicate that "In 1972, ORBIT and DIALOG became the first two commercial providers of online information." That statement is incorrect! Attached is a recap of the development effort with which I was associated in the late '60s. As you can see, two Mead Corporation subsidiaries were the first. (*Editor's Note: Attached list of dates is not reprinted here. —NG*)

The Data Central full-text technology was commercially demonstrated in the fall of 1968 at the ASIS convention in Columbus, Ohio, by Data Corporation, a Mead subsidiary. Starting then and continuing until the business was transferred to the new subsidiary, Mead Data Central, Data Corporation was in the commercial full-text online information retrieval business. The business, concentrated on private material, was supported from two commercial (Washington, DC and Dayton, Ohio) and one government center (Wright-Patterson Air Force Base).

In early 1970, the Information Systems Division of Data Corporation became the nucleus of the new subsidiary, Mead Data Central, Incorporated. This subsidiary expanded the business to include the selling of full text of public data to the legal profession; this business became what is now known as LEXIS/NEXIS.

→

If your date (1972) is correct, you can see that the Mead Corporation was at least two years ahead of either ORBIT or DIALOG. More importantly, by using full text, Mead was technologically ahead of those two whose business in those days was limited to bibliographic retrieval.

Richard H. Giering
Libertyville, IL

EUROPEANS "SPEAK" MANY SEARCH LANGUAGES

To the Editor:

In this letter I want to pay attention to a few aspects concerning the differences in the use of online systems between the U.S. and Europe, especially Finland. It should be noted that this letter reflects my personal view—somebody else might have a different view on the subject.

All information specialists are (or at least should be) frequently following the most important magazines in the industry. Every now and then one may read comments like "No longer is the online industry dominated by one or two online systems . . ." (Nancy Garman, *ONLINE*, March 1992, p. 8). I have to disagree.

The world is not limited to the land between the coasts of the U.S. Nor has the online industry been so limited here in Europe as one might think. Because of the differences in cultures, countries, and languages in Europe, we have been obliged to skip the comfort of using only one or two systems since the beginning.

This does not mean that information specialists do not have their personal favorites or The System-that they mostly use. It means that besides using different systems and search languages, we have to use even different natural languages. It is not exceptional for an information specialist in Finland to use five systems and three different languages during one working day. (Our information service has access to 60 different online systems!)

Using different natural languages is much easier than people usually think. Of course you can never be sure that you find all relevant information, but at least you find some—no matter what kind of lingual capabilities you have in a certain language. So, after all, it is only a matter of attitudes. If you are trying to find information on French nuclear power plants, you have to search the French online systems. (Guess if I can speak any French!)

The March issue of *ONLINE* included many interesting articles concerning the use of different systems. For the reasons mentioned above, I think that in some cases the ways to see the topics were too narrow. At least for me the key question for using different systems is not that one system gives cheaper print formats than another. I choose different systems because they simply include information that cannot be found anywhere else. That is why I sometimes feel conversations on which system is the best are quite frustrating. My customer is not interested in where I have found the information. He or she just wants to get it—even in Japanese—there are often pictures or tables in the articles.

Jaakko Anttila
M.Sc. (Eng.) Information Specialist
Technical Research Centre of Finland

Letters to the Editor are welcome, not only by conventional mail to 462 Danbury Road, Wilton, CT 06897-2126, but to any of our electronic mailboxes: OnTyme—CLASS.ONLINE; DIALMAIL—10045; CompuServe—76077,1320; or Fax 203/761-1444.

Watch for this box each month in the LETTERS TO THE EDITOR *column in both* ONLINE *and* DATABASE.
Clip and post it by your terminals, and remember to use this time to sharpen your searching skills.

FREE TIME

AUGUST

BRS
Physician Data Query Cancer Information (PDQB, PDQC, PDQD, PDQI, PDQP)—One hour free connect time; telecommunications and print charges still in effect

Data-Star
Bibliodata Fulltext Sources Online (FULL)—August 20, 21, 30 minutes free online usage (telecommunications charges still in effect)

D&B Country Files (D&B)—August 21, 27, one free connect hour; document and telecommunications charges still in effect

Financial Times Business Report (FTBR)—August 28, 30 minutes free online usage (telecommunications charges still in effect)

DIALOG
DIALOG CHRONOLOG NEWSLETTER (File 410)—Up to $15 free combined connect time and online display charges

EPIC
PsycINFO—Up to $35 free connect time; telecommunications and print/display charges still in effect

STN
LCASREACT—Free connect time; telecommunications charges still in effect

SEPTEMBER

BRS
PsycINFO (PSYC)—One hour free connect time; telecommunications and print charges still in effect

Data-Star
D&B Country Files (D&B)—September 4, one free connect hour; document and telecommunications charges still in effect

Predicasts Aerospace/Defence Markets and Technology (PTDT)—September 3, 4, 30 minutes free online usage (telecommunications charges still in effect)

DIALOG
Chemical Engineering and Biotechnology Abstracts (File 315)—Up to $75 free combined connect time and online display charges

Kompass Canada (File 594)—Up to $54 free combined connect time and online display charges

STN
LCASREACT—Free connect time; telecommunications charges still in effect

Compuscience—One free connect hour

These schedules are subject to change; check exact availability with each online service prior to use.

Search Strategies and User Interface†

RICHARD H. GIERING

Mead Technology Laboratories, Research Park, Dayton, Ohio 45432

Received August 30, 1974

This paper discusses what in The Mead Corporation have found to be certain aspects of search strategy with respect to searching large data bases and certain problem areas that remain. The paper defines some problem areas, especially problem areas in the freeing-up of the interface language between the user and the search process. Four specific areas in the interface language are defined, and a recommendation is made concerning basic research into some linguistic meanings for the four areas.

What is a large data base? I would venture a guess that although each one of us has our own definition of a large data base, there are some subtle differences between those definitions. We at The Mead Corporation are involved with large data bases in many classes, and I have found that the people involved with large data bases of the bibliographic class are far ahead of others in their thinking concerning the use of large data bases in general. But the problem areas for other types of large data bases (large for other reasons) also present some formidable difficulties.

Now, under what criteria can a data base be large? First, a data base can be large because the number or records or entries in that data base is large. Five hundred thousand entries in a data base represent a large data base, independently of the size of each of the individual entries. A data base containing merely bibliographic information, including a limited number of manually assigned index terms (on the 500,000 entries), could well require storage of a mere 50 to 100 million characters of information. Not a very large amount of storage in anyone's evaluation, but a large number of entries.

The second criterion for defining a large data base, then, could be the numbers of characters involved. Again, a bibliographic data base with only 10,000 entries in it, but with each entry being the full text of the document involved, represents a data base whose character count is on the order of a billion.

A third measure against which a data base could be considered is a measure very important to the search or selection process. If the number of selectable elements, possibly a controlled thesaurus, is on the order of a thousand, independently of the size of the data base in characters or entries, the data base could be said to be small. On the other hand, in a full-text operation, where every potential word and every potential value, including all variants, are search-selectable elements numbering in the hundred-thousand area, the data base could be said to be large, especially if the software involved required a serial search of those elements.

Now, finally, a fourth condition under which a data base could be said to be large, again independently of the previous three, appears primarily when we get away from bibliographic data bases—and I believe that this Conference should address this question—because, as bibliographic experts, we have led the field. We have been at the "head of the pack," so to speak. In defining Information Science activities, designers of management systems, for example, are just beginning to think about large data bases in the way we in the bibliographic area have been thinking of them for years. This last condition, under which a data base might be said to be large, would be in the number of fields and/or files associated with the intricate make-up of the individual data base.

We will be addressing all four of these areas—all four of these definitions of large data bases—during our discussion. If, during the discussion, there is disagreement between what is said by two people, myself being possibly one of them, or two people within the audience during the question and answer period, it is quite possible that the difference lies in the assumed definition of some of the terms, including this one.

I would now wish to raise another question from a definition standpoint. Who, or what, is the user? We have been inundated recently with calls for standardization of systems specifications of systems intercourse. There have been a wide variety of complaints about having to learn multiple disciplines of access for multiple large data bases. I wish to propose a hypothesis and suggest that it be the subject of some discussion. Is it not possible that the reason there is such a dichotomy of user interaction specifications is that the individual definitions of what and who the user is vary? The attributes that are ascribed to this person called "the user" by system one vary and differ greatly from the attributes associated with this person called "the user" by system two. Until we have some commonality of understanding, until we agree on the minimum set of attributes for "the user," we will continue to have a dichotomy of systems.

The understanding and agreement as to what a user is are important in still another area. The on-line interactive use of data bases, large or small, is in its infancy. As the user population grows, it is quite logical to assume that the profile (make-up) of that population will continue to change. Unless we, as an industry, understand this changing phenomenon, we will not be adequately ready to support the end user. I am going to describe this end-user now by *my* definition—this is the definition of the end user profile as we in The Mead Corporation see it. Up until recently, end users, that is, the users on the terminal, have had at least a smattering of training in the Information Sciences technology. They have been able to assimilate meanings of terms such as "Boolean logic," "operators," "arguments," etc. For the most part, we have been building the system for our own inner circle of users. That is not the definition of the user as The Mead Corporation sees it. The definition I am about to give you looks into the future where, *via* the advanced technology of such things as cable TV, the end users of both large and small data bases will be scared out of their wits by such terms as Boolean logic, even though in their normal day-to-day, natural language communication, they use an "AND" and an "OR" all the time.

† Presented in the "Conference on Large Data Bases," sponsored by the NAS/NRC Committee on Chemical Information, National Academy of Sciences, May 22–23, 1974.

Definition: the end user of an on-line interactive data base system has no knowledge of formal logic, of formal discussion, of formal anything, but rather is a professional in a profession *outside* our Information Science profession. He is in a profession we do not understand, and, more importantly, he does not understand our profession and *doesn't want to.*

This, then, leads into search strategies. What search strategy can an individual of this category use. Frankly, we are back to search strategies that we always refer to as "K-I-S-S," *Keep It Stupidly, Simple.* For the most part, the users always refer to the most universally accepted term—whether that's a manually assigned thesaurus term or a text term. He thus obtains a large set of answers, which he then reduces until he sees a set that he thinks he can work with. Note that I am talking about "*he* thinks," "*he* can work with," "*his* discussion," and "*his* term."

What about the help he needs. It is very critical, in our estimation and in support of search strategy process, that the user has access to an on-line, interactive, multi-level tutorial to help walk him through, so to speak, the process of selecting the answers he wishes to use.

Obviously, such a user—one who is not oriented in the intricacies of Information Science—would love to be able to communicate with the computer in something akin to natural language. We have found it necessary to start making some subtle additions to search terminology available to the end user. Let me give you an example. We all, I am sure, understand the true meaning of an "AND" conjunct with respect to an "OR." We understand the implied meaning of the expression "(A and B) or C" as it is expressed with parentheses—and, by the way, the end user hates parentheses. They have absolutely no meaning because he never speaks with parentheses, nor does he ever write with parentheses. Linguistically, that expression may be simplified as "(DOCTOR and PATIENT) or LAWYER." This is implicit—we know that DOCTOR and PATIENT go together.

Using the same linguistic expression, "DOCTOR and HOSPITAL or OFFICE," this exemplifies another form of "A and B or C," and it has a completely different implied meaning. Here, DOCTOR is associated with either of them, and the parenthetical notation of the meaning of that linguistic expression is different A and (B or C). (DATA/CENTRAL) allows for both expressions because, in a search strategy, a user needs to be able to communicate in the way *he* is thinking of the question. Again, basic search strategy is to keep it simple. Obtain a large set of answers and narrow it down to the appropriate set needed. But it is also important that the user be able not to lose any sets and, if his narrowing process takes him too far down, he must be able to backtrack, move in another direction, and to wipe clean the other areas because he doesn't want to be reminded of those areas—he doesn't want to be constantly shown his previous mistakes.

Now this brings us to another part of the definition of who the user is. A professional in his own profession, utilizing a very complex thing called a computer, is exceptionally peer conscious. What do I mean? He surely does not want his peers to see him make a mistake. The more flamboyant a terminal is, the less likely a man is to use it. A loud, clattering teletype is a very flamboyant terminal; a large screen blinking as a monster is a very flamboyant terminal. A much more usable terminal would be the private desk-top unit, an example of which is shown (Figure 1). This one happens to be on my desk, but it is this kind of terminal that we find to be extremely important for the end user, especially the ones who are extremely peer conscious. We can discuss the intricacies of search strategy, but the actual search strategy used is highly dependent upon the profile of the user, and the user is not likely to try anything special. He will not try any special kind of combinations that

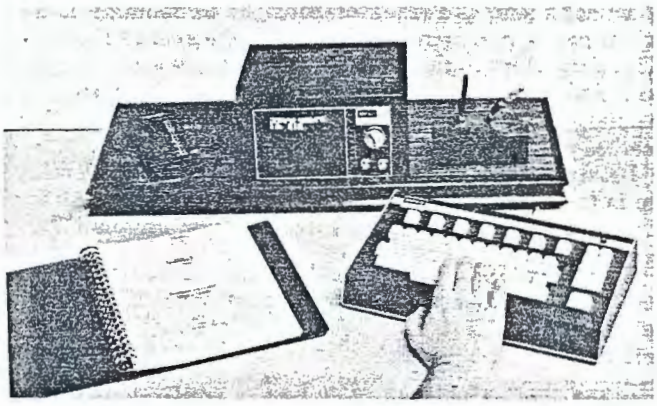


Figure 1.

he is not quite sure of if he feels that somebody might well be looking over his shoulder at what he is doing. I might add that this includes monitoring his activity. He doesn't want anybody to laugh at him, including those "stupid computer people," those "holier-than-thou Information Science experts." "I'll do it my own way before I do something that lets them get one up on me"—that's his attitude.

Now let me address another area of search strategy with respect to full-text systems. That area is synonymy or equivalences—we refer to it as "search-onymy." Of course, there are certain terms that will always and forever, within the context of a particular data base, be synonymous. These equivalent terms will, of course, be used by the appropriate data systems automatically. I would like to ignore that part of the problem since all of us have addressed that problem. People use them, they are used by the computer, and generally they are completely transparent to this terminal end user. He doesn't have to think about them; he knows about them when he sees the synonymous term "highlighted," but normally he doesn't have to worry about it. This applies to plurals and other forms of the word. What, however, about those terms that are synonymous sometimes, under certain contexts, and not synonymous under others. The mere display in alphanumeric order of the different word forms surrounding the word (selected for search) alphabetically is not fully satisfactory, since the display only answers the question of the various word forms and does not address the other question of synonymy: "is head synonymous with cranium?." One alternative solution is the display of a controlled dictionary or thesaurus in which broader and narrower terms are related to the individual term; the old controlled thesaurus used for manual indexing now has another use: aiding in search strategy. In fact, I would venture a guess that the use of the controlled thesaurus will increase in this area because the end user needs help. But the problem here goes well beyond that! What about the terms that have been forgotten in the controlled thesaurus? One cannot afford to continually update this thesaurus with every word seen in the text of the material on hand—it is just too expensive. We now come to the area that we have found to be extremely important in search strategies. And that is the ability to browse through a set of answers textually. The user can then subtly, and without realizing he is doing it, make use of a facet of natural language writing that indicates that the same term should not be used too many times in the same paragraph or set of paragraphs. Rather, synonymous expressions should be used to avoid monotony. As data bases, large data bases (and in this case I am using the term "large data base" especially with regard to the number of words (selectable words) in each entry and obviously the number of selectable terms in each data base), grow and they include more and more of the text, whether it be an abstract or the actual originating author's text, the ability to browse through the material, finding these searchonyms

as we call them, as a result of using one term, immediately will trigger a new trend of thought in the eyes of the user. This allows him to return to the search process and modify with additional constraints. We find this to be extremely helpful—probably as helpful to the end user as any other method of aid in this area.

Let us now consider the problems that I foresee showing up in the not too distant future, which I entitle A Call for Extended Logic for Use in Interactive Information Processes.

A CALL FOR EXTENDED LOGIC FOR USE IN INTERACTIVE INFORMATION PROCESSING

In the chapter entitled "The Rise of Abstract Algebra," Carl B. Boyer states:

> "The history of logic may be divided, with some slight degree of oversimplification, into three stages: (1) Greek Logic, (2) Scholastic Logic, and (3) Mathematical Logic. Whereas in the first two stages logic theorems were divided from ordinary languages, the third stage proceeds in a contrary manner—it first CONSTRUCTS a purely formal system, and only later does it look for an interpretation in everyday speech. It's floruit date (the stage) is really the year in which Boole's first book appeared."[1]

Since the days of Boole and DeMorgan[2] (Boole's first work was published in 1847), the succeeding century has been spent in the association with the fundamental laws[3] of logic defined in the formal system. All expressions can be defined in combinations of the basic fundamentals: intersection (&) and inclusive union (or).

With the rise of users of computer systems (which operate on logical circuits and/or programs), who themselves are not trained in the formalism of mathematic logic (Boole rightly claimed that logic is associated with mathematics rather than with metaphysics), it is necessary to close the circle described above by defining the complex elements of logic, so that the user might use a more simple form of expression in communicating his requirements. We may be entering into a fourth phase in the history of logic in which expansions of the basic two elements into formalized linguistic expressions are generated to allow for less ambiguity in communication—not only between man and automaton, but also man to man.

This period in the history of logic is not without its own problems in construction of formal expressions. As will be seen further in this presentation, an additional basic logic element may be in the process of being defined. Additionally, logic has been explained using sets (Venn diagrams). Sets have been, up to today at least, a one-dimensional concept. It is barely possible that we have embarked (using operators and expressions) on the definition of a second dimension in the set theory. We will discuss four problem areas requiring formal definitions:

1. Additional (complex) logical expressions
2. Arithmetic expressions that defy algebraic logic
3. Multidimensional expressions
4. Concatenation and other "G-connectors"

As interactive information processing becomes more used by non-information science professionals, it becomes apparent that extensions to logic are necessary for this unique set of information, especially when the information being processed is multivalued (*e.g.*, textural or periodic). It is readily admitted that there is no specification that cannot be properly written by the user or properly processed by computers using standard conventional Boolean logic. How many non-ADP professionals, who would have use of an information system, are readily able to discern the difference between the subtle ramifications in Boolean logic in order to enter semantically correct and syntactically definable elements of search? There are, of course, some, but their numbers are relatively few in comparison to the numbers of potential users of information systems. Extensions of logic, therefore, are needed in order to make easier the definition of the problem by nonprofessional users and, much more importantly, to make for efficient processing of the problem by the computer.

We have, in some sense, already done it for one expression. There is no need, in formal logic, to use the exclusive OR, defined, using the two basic elements, as:

$$(A \text{ ER } B) \equiv (A \text{ or } B) \ \& \neg \ (A \ \& \ B)$$

In programming, however, it has been found to be cumbersome to use only the fundamental elements; we have, therefore, engineered an exclusive OR operation for programmers to use. Other combination forms (called MACRO) for complexities of logic need to be defined, based upon combinations of the basic elements.

A second example, already in use, is the AND condition. We are all fully aware of the relationship under standard Boolean logic of the & and the OR condition. Since there is an implied ambiguity, the expression "A and B or C" can have either of the following meanings (the word AND used linguistically rather than formally):

(A & B) or C
A & (B or C)

Classically the defined expression "A and B or C" has the meaning (A & B) or C, which linguistically is the combination of A and B or the single element C. This is exemplified by the actual conversational expression "Doctor & Patient or Lawyer." A great deal of effort and a great deal of rigorous mathematical definitions have gone into all of the ramifications of this logic. Witness DeMorgan's law in which the reversal of operators is effected by the establishment of parenthetical nesting of the NOT operator:

NOT (A or B) is identical to (NOT A & NOT B)

People, on the other hand, do not always think in a Boolean manner. In fact, it is this author's opinion that a great many people think in the reverse form and, without any other specification, the expression "A and B or C" tends, in normal conversational discussion, to take on the meaning established in the second line above, spoken as "the single element A in combination with either D or C." A conversational example is DOCTOR and HOSPITAL or OFFICE. This definition of logic has never been passed through the rigor of mathematical treatises, except in the use of parenthetical notation. No "not-reversal" (DeMorgan's duality) has been defined, and all of the rigor of the combination of the two has not been defined except by use of actual parenthetical notation. It is *believed* the NOT (A and B or C) would be, based on parenthetical notation, the same as saying NOT A or the combination of NOT B and NOT C. It is time that we in the Information Science profession applied the rigor of logic to this preceding linguistic expression so that our users, who are not necessarily mathematically oriented, are able to express themselves in their language either way, rather than having our language imposed upon them. Based upon the opinion that both forms of the logical AND need to be defined and used by users, both a superior (to OR) and inferior AND are definable in (DATA/CENTRAL). In this writing, the word AND is considered to be superior while the & is considered inferior:

A and B or C means A & (B or C)
A & B or C means (A & B) or C

1.=AUTHOR = JONES AND=DATE< JUNE, '71

2.=AUTHOR = JONES AND=DATE ¬>JUNE, '71

Figure 2.

A second area is one in which there is at least one anomaly in normal logic expressions, as found in general computer-oriented languages. Consider, for example, the two search requests shown in Figure 2. Note that the pound sign refers to the field in which the conditions are to be satisfied, while the operators are defined as, for this example, ≤ meaning "less than or equal to" and the operator ¬> meaning "not greater than."

In normal algebra, the two statements have identical meaning! It is possible that in data base systems they do not. Definition of negation: The logical definition (in normal expression) of the negation process requires the removal, from the set of possible answers, of any answer satisfying the positive form of the negative expression, for example:

> "Any hotel but not the St. Francis" means to find a list of All Hotels and remove from that list any hotel positively named St. Francis.

Take a simplified form of the example. The entries are edited such that, on input, the field name DATE can contain only one value, the date of publication. That value, however, may be an actual date or it may be an indication of an unknown date. Please note that an unknown date is definitely not a date of zero time. An entry with an author named JONES with an UNKNOWN indication (absence of the date) would not satisfy the first statement; therefore, it would not be an answer. By not having a date, it cannot be considered to be not greater than and it would not be removed from the set of answers. Therefore, the only criterion valid to determine the satisfaction to the search is the initial JONES criterion; it, therefore, satisfies the second statement.

Admittedly, the above can be expressed in a different manner by the additional specification of the editing criteria that the field must contain either a valid date or it must contain the word UNKNOWN. Then the second search could be rephrased (to make it the same, logically, as the first statement) into:

#Author = Jones and #Date (¬>June '71 or = Unknown)

This becomes a human engineering or user-interface problem because it is beyond reason, in our belief, to expect the novice (or the non-information science oriented) terminal user to remember all of the various editing criteria for files that can and do contain in excess of a hundred fields and especially for files in which the fields were added at different times by different people. The solution, then, is to extend the language such that the concept desired can be expressed without resorting to looking up the edit criteria.

As information processing takes on more and more of the job of handling multivalued or textual data, we must understand the subtle difference between the two. Textual data here are defined as any data found in fields of individual entries (records) such that the structure of the data cannot be predefined; multivalued, on the other hand, refers to fields that can contain multiple separately searchable values. This, of course, includes the name of a person in a personnel record, his address, the name(s) of the school(s) he attended, etc., prior to the preestablishment of arbitrary codes for these data.

Now for a third area where extensions to logic (and especially their meanings for processing purposes) are necessary. (The "or" as used below is the normally used inclusive OR; the ER used below is the exclusive OR). In normal file handling logic, only two of the following nine possible combinations have defined meanings:

$$(\text{Field 1} \begin{Bmatrix} \& \\ \text{OR} \\ \text{ER} \end{Bmatrix} \text{Field 2}) = (\text{Value 1} \begin{Bmatrix} \& \\ \text{OR} \\ \text{ER} \end{Bmatrix} \text{Value 2})$$

Normally, the use of the conjunctive "&" connector and the exclusive OR-(ER) to the right of the operator is not defined. In the meaning below, the phraseology of the verb "appear" is used as though the ⟨operator⟩ is the logical appearance operator. The reader is reminded that the definitions have slightly subtle changes of meaning if the operator is different. Here we define all nine as follows:

1. (Field 1 or Field 2) ⟨operator⟩ (Value 1 or Value 2) means that the occurrence of either value (or both values) (a value may be a word, phase, or—if the operator is arithmetic—an arithmetic value) in either field (or both fields) satisfies the request and any such entry is considered valid for the retrieval and display process.* Example: Find all documents whose country-of-publication or country-of-nationality is Germany or France.

2. (Field 1 or Field 2) ⟨operator⟩ (Value 1 & Value 2) means that both Value 1 and Value 2 must occur in either Field 1 or Field 2 for the entry to satisfy the search and be available for retrieval and display. Example: Find all medical histories in which either Record-of-Treatment or Post-Operative-Care deals with both the heart and kidneys.

3. (Field 1 or Field 2) ⟨operator⟩ (Value 1 ER Value 2) means that either Value 1 or Value 2, but not both, must appear in either Field 1 or Field 2. Example: Find all medical histories in which either Record-of-Treatment or Post-Operative-Care deals with either the heart or the kidneys but not both.

4. (Field 1 & Field 2) ⟨operator⟩ (Value 1 or Value 2) means that either value must exist in both fields for the entry to satisfy the search. Example: Find all projects that had as both primary objective and methodology the use of either rockets or missiles.

5. (Field 1 & Field 2) ⟨operator⟩ (Value 1 & Value 2) means that both values must occur in both fields for the entry to satisfy the requirements of the search. Example: Find all chemical compounds that have both hydrogen and fluorine listed in the two fields: elements-used and elements-reacted.

6. (Field 1 & Field 2) ⟨operator⟩ (Value 1 ER Value 2) means that either Value 1 or Value 2, but not both, must appear in both Field 1 and Field 2.

7. (Field 1 ER Field 2) ⟨operator⟩ (Value 1 or Value 2) means that either Value 1 or Value 2 or both Value 1 and Value 2 must appear either in Field 1 or in Field 2 but not in both fields.

8. (Field 1 ER Field 2) ⟨operator⟩ (Value 1 & Value 2) means that both Value 1 and Value 2 must appear in either Field 1 or in Field 2 but both values are not to appear in both fields.

9. (Field 1 ER Field 2) ⟨operator⟩ (Value 1 ER Value 2) means that either Value 1 or Value 2 (but not both values) are to appear in either but not both fields.

We have run some preliminary studies attempting to define the results of both the positive and negative uses of these three connectors in nine combinations yielding 144 unique expressions. The definition of the negation operator (¬), it is believed, bears repeating: the negation operator applied to a value, asks for any entry containing the pre

* An example of the form when one value is textual and the other is arithmetic would be:

(Field 1 or Field 2) (= Value 1 & < Value 2)

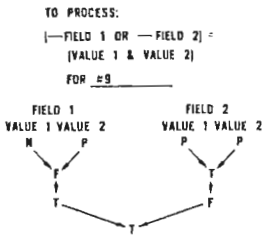


Figure 3.

16 COMBINATIONS:

ENTRY	FIELD 1 VALUE 1	FIELD 1 VALUE 2	FIELD 2 VALUE 1	FIELD 2 VALUE 2
1	P	P	P	P
2	P	P	P	N
3	P	P	N	P
4	P	P	N	N
5	P	N	P	P
6	P	N	P	N
7	P	N	N	P
8	P	N	N	N
9	N	P	P	P
10	N	P	P	N
11	N	P	N	P
12	N	P	N	N
13	N	N	P	P
14	N	N	P	N
15	N	N	N	P
16	N	N	N	N

Figure 4.

ence of that value to be deleted from the set of answers to which it applies. The negation operator, applied to a field (or segment), asks for the complement of that segment's action with reference to the specified combination of values.

In the study, there were 16 combinations of entries that were evaluated. In Figure 3, N refers to the nonpresence of the value in the segment (or field) while P indicates presence. The table was generated, based upon the way an implementation might process the data: Evaluate the right-hand portion of request first, but separately for each segment (or field) specified on the left side (independent of the sign associated with the segment); reverse, if necessary, for the sign of the segment; and combine, based upon the connector between the segments. *Ergo,* for entry number 9, for the following request (one of the 144 expressions), the processing is specified in Figure 4.

$$(\neg \text{Field 1 or} \neg \text{Field 2}) = (\text{Value 1 \& Value 2})$$

(It is noted that this request, as is the case for all of the above conditions, can be stated in terms of the basic Boolean elements. It is believed that this would be:

$$(\neg \text{Field 1} = \text{Value 1 \& Value 2})$$

or

$$(\neg \text{Field 2} = \text{Value 1 \& Value 2}).)$$

It should be apparent that many anomalies appear when the "not" operator is evaluated. Some of the propositions needing rigorous proof seem to be the following.

Proposition 1: The negative of an expression is not necessarily the negative of the parts.

Proposition 2: A not sign, applied to a complete Boolean request expression, can be applied only to the fields or to the values, but not both under discussion, and then DeMorgan's Law applied for values and does not affect the field specifications.

Proposition 3: When applying DeMorgan's Duality Law to the "taking in" of the NOT operator, parentheses may *not* be removed, as they normally define hierarchic relationships.

Now for the fourth area of discussion needing expansion. It has become apparent that many users of information systems (especially those processing text) will have a distinct requirement to specify relationships of the various segments (or fields) of information such that the fields can be considered to be concatenated without necessarily restructuring the physical file. A highly simplified example should suffice to define the problem. Consider a personnel file in which three of the fields or segments of that file are defined as Job-History-External, Job-History-Internal, Current-Position-Description. Please note that each of these fields contains large amounts of text. Consider the request: Search the records for anyone having experience in Missiles and also in Aircraft (disregard, for purposes of this problem, the additional words that might satisfy the requirement). It is not merely satisfactory to say:

(Job-History-External or Job-History-Internal or Current-Position-Description) = Missile & Aircraft

The true logic of the expression above states that both words, Missile & Aircraft, must appear in any combination of Job-History-External, Job-History-Internal, or the Current-Position-Description fields. The condition expressed below is, in fact, that which is implied by this verbalization of the problem:

((Job-History-External or Job-History-Internal or Current-Position) = Missile & Aircraft) or
((Job-History-External or Job-History-Internal) = Missile & Current-Position = Aircraft) or
((Job-History-External or Current-Position) = Missile & Job-History-Internal = Aircraft) or
((Job-History-Internal or Current-Position) = Missile & Job-History-External = Aircraft)

Obviously, the statement of the expression of the problem as defined above is a bit horrendous for a nonprofessional to attempt. One solution is to restructure the file with the three previously defined fields now concatenated into one. A terminal user could then request the occurrence of both words in the one new field. He would, of course, have to wait (at the terminal) while the restructure (just for him, probably) took place; he would most likely become frustrated. It is believed, therefore, that the best approach is to define a new Boolean connecting expression called a GOR. Using the GOR, the problem can be defined thusly:

(Job-History-External GOR Job-History-Internal GOR Current-Position) = (Missile & Aircraft)

As can be noted, the expression of the problem is now much simpler. Of course, the GOR has (currently) no meaning when used between two or more values to the right of the operator. It also has no meaning when the sign associated with any of the five elements (three segments, two values) is positive and the connector between the two values is OR. When either condition exists, then the answer is unique with respect to single connectors. This is caused by the fact that the concatenation (inclusive GOR) of the segments is based on positive values (presence of the value). Postulation: The negative of the concatenation is *not* the concatenation of the negative (nonpresence). Another way of saying it would be: absence of the same segment pair. It is believed that, as the science of automated information processing advances, a whole series of G-connectors will need to be defined to aid the non-ADP oriented user in expressing a request for the solution to his problem.

For example, the defined GOR is an inclusive concatenation of the fields under consideration. It is highly probable that, as usage of information systems grows, the need for a definition of the exclusive concatenation, referred to as GER, would be required. It could be defined as: "The inclusive concatenation "MINUS" those entries in which the same value appears in both segments."

Additionally, the concatenation–intersection (G&) might well be required. In both of these connectors, the location of the value in the multivalued segment may wish to be used as opposed to the normal GER and G& where only the presence of the word is used. As a final example, consider the following for a special G-negation:

An application is defined with one logical file containing 78 segments or fields. A predefined pseudo-segment called SEG80 is defined as the concatenation of Segments 1, 7, 13, 17, and 18.

Thus:

Seg 80 ≡ (Seg 1 GOR Seg 7 GOR Seg 17 GOR Seg 18)

A user, on an interactive terminal, wishes to access the file with the following request:

(Seg 1 GOR Seg 7 GOR Seg 13 GOR Seg 17 GOR Seg 18) = (Word & WD)

It is identical with

Seg 80 = (Word & WD)

Another user might wish to access the same file except that, for this user, the concatenation should not include Seg 13. The special G-Negation (G−) would allow this as:

Seg 80 G− Seg 13) = (Word & WD)

As long as only the GOR and the G− are the only avai' able G-connectors, the processing formalism is not t complex; once, however, other G-connectors are requirea (especially the GER & the G&), the composite meaning may become most complex.

As can be seen, we have attempted to define four areas in which additional man-machine conversational linguistic expressions can and, we believe, should be defined to make easier the way our users may communicate with the computer. To define his search strategy (especially as the user becomes more knowledgeable), the user will want ever-increasing capabilities with even more simplified forms of expression until, in the long run, the user will be able to converse verbally in natural language with the computer.

LITERATURE CITED

(1) Boyer, C. B., "A History of Mathematics," Wiley, New York, N. Y., 1968, p 633.
(2) Augustus DeMorgan, 1806–1871.
(3) Boole, G. S., "Investigation of the Laws of Thought," Dover Publications, New York, N. Y., 1854.

Data Security†

M. J. ORCEYRE

IBM Corporation, Poughkeepsie, New York 12602

Received December 13, 1974

Data security is a rich and complex subject dealing with the protection of the computing capability from all threats to its continuity. Some fundamental elements of the process of achieving a reasonable, prudent measure of that protection are considered.

Data security is a rich and broad topic, one that is of increasing concern to data-processing-oriented people at all levels, and is receiving more (and more formal) attention from users and manufacturers alike. The definition of data security should indicate the scope, complexity, and pervasive nature of the subject: it is simply the safety of data (and necessarily also of the system) from improper disclosure, modification, or destruction—whether these are accidentally or intentionally caused. Note that this definition applies as well to the manual as to the EDP operation. Note, too, that it is a global definition; no threat to the continued well-being of the operation is excluded. I intend this to be a complete—however brief—discussion, so you may expect me to deal with all sorts of situations that threaten the safety of data, ranging from technologically complex penetrations of computing systems by highly trained intruders, to earthquakes, to coffee spilled into the machinery, and so on.

† Presented in the "Conference on Large Data Bases," sponsored by the NAS/NRC Committee on Chemical Information, National Academy of Sciences, May 22–23, 1974.

In fact, when you yourselves deal with data security, keep this breadth of scope in mind. After all, to protect data you must understand the threats to those data. To protect data completely against all threats is an unrealizable goal; to protect data to some reasonable extent against reasonably predictable and probable threats is a prudent and practical goal. To accomplish the latter, you must undertake a risk assessment, which involves gaining the clearest possible understanding of the nature of that which you must protect and also of the relative probabilities of the events that threaten the well-being of what you must protect. If you do not have this understanding, you cannot assess risks; if you cannot assess risks, you cannot prudently determine protective measures; and if you cannot prudently undertake protection you cannot know that you are protected.

It is my intent to review some fundamental elements o the process of achieving protection. One of these elements is a clear understanding of the need for protection, and this need—which you all must have to one degree or another—springs from a number of sources:

DTN-67-9

TECHNICAL NOTE

ANALYSIS
OF
EXISTING AND PROPOSED
DATA HANDLING SYSTEMS

by

Richard H. Giering

Eastern Division
Data Corporation
1254 Jefferson-Davis Hwy.
Arlington, Virginia

23 October 1967

Data corporation

CONTENTS

INTRODUCTION

Since the information contained herein has been gathered from various sources, it is incumbent upon the reviewer to delineate his sources as an indication of the validity of the data. Additionally, Data Corporation (the reviewer) by its involvement in one of the reviewed systems, might be considered prejudiced in evaluating its own system. This possibility was recognized throughout the investigation and an attempt was made to remove any bias.

0.1 Sources

0.1.1 Information on the various IDHS systems was gathered through the following documents and additional data (to clarify questions) was recieved in direct verbal communication with members of the Systems Operations Support Branch at DIA. The documents are:

0.1.1.1 The DIAM 65-9 series titled "Intelligence Data Handling System (IDHS) 1410 Formatted File System (1410 FFS)" with various dates, published by the DIA. (This document covers Mark II).

0.1.1.2 Mark III 1410 Formatted File System, Preliminary User's Reference Manual dated 10 February 1967 by DIA.

0.1.1.3 A series of three manuals dealing with 7094 FFS revised during June 1966 and published by DIA. These manuals define Mod 7.5 7094 FFS.

0.1.1.4 IBM Document No. N-0661 titled "Mod 8 7094 FFS Program Capability Changes dated 5 May 1967 and published by IDHS, Contracting Agency, Rome Air Development Command, Griffiss Air Force Base, New York. (This document defines the changes to Mod 7.5 FFS to make it Mod 8.)

0.1.2 Information relative to the NIPS Systems were gathered primarily through discussions with individuals currently on the programming staff for the NIPS development effort. Clarification of details, however, was gathered from system planning manual SPM 1-67 dated 31 January 1967, relating to the program design approach for S 360/50 FFS as published by the NMCS.

0.1.3 An original IBM document No. E20-0179-0 relating to the description of the Generalized Information System (GIS) has been found to be erroneous. IBM representatives with whom discussions were held to gain additional information, reported this fact. The original GIS is now two separate systems, Document Processing System (DPS), the description of which may be found in the Application Description Manual No. 360A-CS-12X published about July 1967 by IBM. The remaining capabilities of GIS (once text processing specifications are removed) seem to be inherent in the second system currently called GIS.

0.1.4 Information on the Informatics Mark IV data handling system capabilities is not available. Information relative to this system contained herein was gained through technical discussions with Informatics personnel.

0.1.5 The source of (Data) Central information is self-evident.

0.1.6 Information on TDMS was found in an artical titled "Treating Hierarchical Data Structures in the SDC Time-Shared Data Management System (TDMS)" found in the Proceedings of 22nd National Converence, Association for Computing Machinery; "The Time-Shared Data Management System: A New Approach to Data Management", SDC No. SP-2747 dated 13 February 1967 by A. H Vorhaus and R. D. Wills; "A Data Management System for Time-Shared File-Processing Using a Cross-Index File and Self-Defining Entries," AFIPS Conference Proceedings dated 1966, Vol. 28, pags 79-86. Also available as SDC No. SP-2248 dated 21 April 1966, by E. W. Franks; "The Language Specifications for the Define Operations of TDMS," SDC No. TM-3370/003/00 dated 15 April 1967 by E. E. Grant and P. A. DeSimone; "COMPOSE/PRODUCE: A user-Oriented Report Generator Capability within the SDC Time-Shared Data Management System, "AFIPS Conference Proceedings, Spring 1967. Also available as SDC No. SP-2634, 8 February 1967, by W. D. Williams and P. R. Bartram.

1.0 Minimum Hardware Configuration

1.1 Central Processor Type

1.1.1 Mark II - FFS is operational on a 1410.

1.1.2 Mark III - FFS will be operational on a 1410.

1.1.3 NIPS 1410 is operation on a 1410.

1.1.4 NIPS Phase I will be operation on a 360/50. It has been reported that the system will be able to operate once it becomes operational on 360/40. This has not been verified in any documentation.

1.1.5 NIPS Phase II - The comments applicable to Phase I are also applicable to Phase II.

1.1.6 Mod 8 requires a 7094 Model 2.

1.1.7 GIS is to operate on a 360/40 or 360/50 although the information concerning the availability on a 360/40 is questionable.

1.1.8 DPS when it is released in November will be operational on a 360/40.

1.1.9 ATS will require a minimum of a 360/40 and possibly by implementation time (January 1968) will require a 360/50.

1.1.10 Informatics Mark IV is to be operational on a 360/30 and up.

1.1.11 (Data) Central is currently operational for a 360/30 and up.

1.1.12 TDMS is written to be operational on a 360/50 or 360/65.

1.2 Central Memory (Size)

1.2.1 Mark II - FFS requires 80K characters of minimum storage.

1.2.2 Mark III - FFS requires 80K characters of minimum storage.

1.2.3 NIPS 1410 requires 80K characters of minimum storage.

1.2.4 NIPS Phase I will require G (128K) memory, as long as no remote processing is desired. It will require H memory for remotes.

1.2.5 NIPS Phase II will require G (128K) memory, as long as no remote processing is desired. It will require H memory for remotes.

1.2.6 Mod 8 requires 32K words memory.

1.2.7 GIS requires either G or H memory.

1.2.8 Document processor will be operational with G memory non-remote. Memory requirements for remote processing are as yet undeterminable.

1.2.9 Memory requirements for ATS are undeterminable.

1.2.10 Informatics Mark IV can operate under E (32K) for DOS. However, when operational under OS will require F memory.

1.2.11 (Data) Central is operational under F memory.

1.2.12 TDMS requires H memory (256K bytes).

1.3 Magnetic Tapes (Number).

1.3.1 Mark II - FFS requires six tape drives.

1.3.2 Mark III - FFS requires six tape drives.

1.3.3 The 1410 NPIS requires a minimum of eight tape drives.

1.3.4 NIPS Phase I requires no tape except that if that portion of the output package defining tape output is used then one drive is needed for this feature.

1.3.5 NIPS Phase II requires no tape except that if that portion of the output package defining tape output is used then one drive is needed for this feature.

1.3.6 Mod 8 requires a minimum of twelve tape drives (17 are recommended) and these tape drives must be on three channels.

1.3.7 Information has not been released.

1.3.8 DPS requires no tapes, however, the documentation indicates that disk files are the normal mode of operation.

1.3.9 This information has not been released with respect to the implementation of ATS.

1.3.10 Informatics Mark IV per se does not require any specified number of tape drives as a minimum; however, since Mark IV is primarily a tape oriented system and for this the data is on tape as well as the output, normal operation will require a minimum of one for the old master, one for the new master, one for any sub-file, and one for output possibly.

1.3.11 (Data) Central per se requires no tape drives; however, it can use them for I/O if they are available.

1.3.12 TDMS also requires no tapes; however, the documentation indicates that disk files are the normal mode of operation.

1.4 Secondary (On-Line, Mass) Storage

1.4.1 Mark II - FFS requires a minimum of one 1301 disk for operation.

1.4.2 Mark III - FFS requires a minimum of one 1301 disk for operation.

1.4.3 NIPS 1410 requires a minimum of one 1301 disk for operation.

1.4.4 NIPS Phase I requires a minimum of four 2311 disk packs.

1.4.5 NIPS Phase II requires a minimum of four 2311 disk packs.

1.4.6 Mod 8 does not require any secondary mass storage; however, if a 2303 disk is available it will be used. In addition, if gross file and multi-file processing is desired a disk is required.

1.4.7 Information has not been released on GIS.

1.4.8 DPS per se requires no disk except that if a disk is available its use is preferred for storage of at least the inverted file.

1.4.9 This information has not been released with respect to the implementation of ATS.

1.4.10 Informatics Mark IV - The comments above in paragraph 1.3.10 (Magnetic Tape) apply here also.

1.4.11 (Data) Central is primarily a direct access system and requires direct access storage of sufficient quantity to cover the data base plus one 2311 (or greater) disk pack for system residency.

1.4.12 TDMS per se requires no disk except that if a disk is available its use is preferred for storage of at least the inverted file.

1.5 Remote Stations

1.5.1 Terminal Cluster - None of the checked systems seem capable of inputting from other than typewriter devices; ergo, a cluster if it is present is superfluous.

1.5.2 Teletype/Typewriter — NIPS 1410, NIPS 360, both Phase I and II, Administrative Terminal System, (Data) Central, and TDMS all may use these devices. Only (Data) Central and ATS requires them. IDHS, Mark II and III, as well as SAC Mod 8 are not contemplated to use remote terminals of any type. GIS, DPS, and Mark IV report a consideration being given to remote terminal operation of a typewriter type.

1.5.3 CRT - With the exception of (Data) Central and TDMS, no system requires nor can use CRT's. Both (Data) Central and TDMS are currently programming for the inclusion of CRT capability (estimated 1 January 1968 for (Data) Central and June 1968 for TDMS. There has been some discussion of the use of CRT's with NIPS; however, no plans have as yet been consummated to this reviewer's knowledge.

2.0 Operating System

2.1 With the exception of TDMS and SAC Mod 8, all systems operate under the operating system specified for the particular hardware. SAC Mod 8 operates as a separate non-communicative system under IBSYS and contains its own resident monitor. TDMS will operate under a SCD owned and maintained operating system called System 360 System on the 360 and is using the TSS-LUCID System on the An/MSQ-32 computer in test phase currently.

2.2 With the exception of (Data) Central, TDMS, and ATS, none of the systems are multi-programmed. ATS is to be multi-programmed in its deliverable state shortly after the first of the year. TDMS is currently operational in multi-programmed environment under TSS (see above). (Data) Central is currently not multi-programmed; however, is being multi-programmed currently for delivery approximately 1 January 1968.

3.0 Sub-Operating System

3.1 Recovery Procedures — All checked systems except (Data) Central, TDMS and ATS will have no direct recovery procedures but will by-pass to the next job in the job stream (this includes the batch and the remote terminal queue). (Data) Central has extensive recovery as part of its conversational mode operation to allow recovery to various levels of re-operation. TDMS reports that it will also have an extensive user recovery capability.

3.2 Operations Recording — Information relative to transaction recordings for GIS, DPS, and ATS is not available. All three IDHS (Mark II, Mark III and SAC Mod 8) have relatively extensive recordings procedures for printout only. These recording procedures cannot be considered as audit trails; however, because they are not maintained. (Data) Central has no direct operations recording and Mark IV reports audit trails of some type that can be considered as a form of operation recording. TDMS saves if requested a transaction tape recording all operations. Documentation as to the level of recording is meager.

4.0 File Format

4.1 Structured Data — With the exception of DPS and ATS, all systems are capable of handling structured data. DPS can handle structured data; however, since it is primarily designed to process text, its capability is relatively limited in structured data area. ATS is not designed to handle anything but voluminous documents and as such cannot process structured data.

4.2 Textual Data — All systems can retain textual material in formatted fields (or in the case of DPS in the field) and this paragraph deals with the direct textual processing capabilities primarily from a search and maintenance standpoint.

4.2.1 Searchable — Only the Document Processing System and (Data) Central has any capability to search textual material for selective retrieval. (Data) Central can do it on any field of the file, where the document processing system is limited to this capability in one available variable length field.

4.2.2 Maintenance Capability — With the exception of ATS no system has direct textual maintenance capability. However, it has been rumored that TDMS is planning for this capability sometime next year. ATS primarily to do this job has a rather extensive repertoire of operators for textual maintenance.

4.3 Field Length and Type

4.3.1 Mark II - FFS allows for fixed length fields of up to 910 characters, each with a maximum of 299 fields per logical entry plus the one variable length test field. Although fields may be defined up to 910 characters in length, during the query phase, only the first 56 characters are available for searching and during the output phase only the first 52 characters are available for use with the conditional statements of the output subsystem. All 910, however, are printable and for fields defined in excess of the above (52 or 56) parameters, the system will treat the data as pure unsearchable free text.

4.3.2 Mark III - FFS is the same as above.

4.3.3 NIPS 1410 allows for 99 fixed length fields (length unknown) with a maximum of 2700 characters per logical entry.

4.3.4 NIPS Phase I has the capacity to handle any number of 200 character fixed length fields per logical entry. A limit of approximately (available data is sketchy here) 250 fields are allowed for the fixed set data plus approximately the same number for each logical segment where a logical segment can be a group of one or more periodic sets. In addition one variable length text field may be assigned per segment.

4.3.5 NIPS Phase II — Same as above.

4.3.6 Mod 8 allows for 200 fixed length fields of 126 characters each (except that arithmetic data is limited to 36 bits) and allows for up to 5820 characters per logical entry.

4.3.7 A maximum of 750 field-names are usable across three files; e.g. 250 for each of 3; 375 for each of 2, or 1 file of 750 names. Each field is limited to 256 characters in length.

4.3.8 DPS may have up to 255 fields of data assigned per logical entry. The length of all but one field is restricted to 249 characters except that the total length for these fixed fields must be less than 1638 characters. In addition to these fixed fields, DPS allows for processing of free text (keywording) and phrasing against one variable length field.

4.3.9 There are no fields allowed for ATS and the complete document entered into the system is considered as one huge variable length field.

4.3.10 Informatics Mark IV reports no limit on the number of fixed length fields and the length of the logical record is limited only by core availability.

4.3.11 (Data) Central allows for 256 times 64 fields, all either variable or fixed in length with no restrictions placed on the number of characters per logical entry.

4.3.12 TDMS reports "some large number" of fixed length fields each a maximum of 256 characters per logical entry. The logical entry size is not restricted.

4.4 Field and Entry

4.4.1 Mark II - FFS allows for a maximum record length of 5400 characters.

4.4.2 Mark III - FFS allows for a maximum record length of 5400 characters.

4.4.3 NIPS 1410 is restricted to a maximum length of 2700 characters.

4.4.4 NIPS Phase I allows for approximately 13,000 characters per logical entry. The approximation is dependent on the use of segments in storage of physical records. NIPS allows for a multiplicity of segments where in general a segment is made up either of the fixed data or of one periodic set data and each segment is restricted to a maximum length of 1000 characters. The number of segments per logical entry at this stage in development has not been finalized but is believed to be on the order of 13.

4.4.5 There is no change contemplated in entry size restrictions for NIPS Phase II at this point in time.

4.4.6 Mod 8 allows for 5820 characters per entry.

4.4.7 It has been reported for GIS that there will be no direct limitation placed on the characters size per logical entry except that it is believed that one logical entry must be able to fit into core restricting it by core space. In addition, it has been reported that GIS will allow for segmentation and there will be theoretically no limit on the segments, except that each segment will have a length limitation dependent on equipment (e.g. disk track length).

4.4.8 DPS has no restriction on the size of the entry by virtue of the inclusion of one variable length by field. See paragraph 4.3.8 above.

4.4.9 There is no restriction imposed on the length of an ATS document.

4.4.10 See paragraph 4.3.10.

4.4.11 (Data) Central places no restriction on the number of characters per logical entry.

4.4.12 TDMS also reports no restriction on the number of characters per logical entry.

4.5 Maximum Number of Simultaneous Files

4.5.1 Mark II – FFS, all three versions of the NIPS, Document Processing System, Administrative Terminal System and TDMS all report no multi-file capacity.

4.5.2 Both Mark III – FFS and Mod 8 can process a multiplicity of files in one batch query; however, the actual process is one file at a time where data is held (and/or merged) between single file processings.

4.5.3 Informatics Mark IV reports the capability for simultaneous processing of up to five files.

4.5.4 (Data) Central processes simultaneously up to 64 files.

4.6 Keywording

4.6.1 With the exception of the Administrative Terminal System, all systems are capable of processing user supplied keywords in that the individual words (and/or phrases) are placed in separate fields for query purposes. The ATS system, because of its primary design for the maintenance of documents rather than the selective retrieval thereof, does not allow for this capacity.

4.6.2 Only the Document Processing System and (Data) Central report any current capability for processing system generated keywords. In both cases the textual material is broken down into keywords and separately stored for query. In addition, TDMS, in the proceedings of the 22nd Conference of the ACM, reports the planned inclusion of keyword ability. It is anticipated that this will not be available prior to the end of 1968.

4.7 Pre-stored Output Formats (in Object Form) — Mod 8, DPS, and ATS do not have this capability. All other systems either do or will have this capacity.

4.8 Unit of Measure Conversion (Quantified Data) — TDMS and (Data) Central both report automatic unit of measure conversion in the system. TDMS, however, does not report the level of conversion or the form the conversion takes (the conversion is currently being designed into the proposed system and should be available with it).

4.9 Security Codes

4.9.1 Mark II - FFS, Mark III - FFS, NIPS 1410, Mod 8, GIS, DPS, ATS, and Mark IV all do not use any internal security codes.

4.9.2 NIPS Phase I and Phase II, and TDMS have a file security mark, which prohibits the use of the file to anyone not appropriately cleared.

4.9.3 (Data) Central uses a security flag at the sub-field (word) level allowing for the inclusion of multi-level security information into a given field.

4.10 Retrieval (Substitution) Tables — TPS, Mark IV, ATS, and (Data) Central do not use retrieval tables, although (Data) Central is considering this for future inclusion. All other systems report the capacity for retrieving against internal tables or using internal tables for the substitution in printout of readable material for internal codes.

4.11 Repeating Data Sets (Periodicity)

4.11.1 Mark II - FFS allows for eight periodics, all at the same level.

4.11.2 Mark III - FFS when it becomes operational is to have provisions for up to 50 periodics, all at the same level.

4.11.3 NIPS 1410 also allows for eight periodics, all at the same level.

4.11.4 NIPS Phase I allows 1250 periodics, all at the same level.

4.11.5 NIPS Phase II allows 1250 periodics, all at the same level.

4.11.6 Mod 8 allows for a single level of nine periodics.

4.11.7 The information for GIS is not available. Discussion, however, reveals that multiple level periodicity (to about 15 levels) will be available; the total number of sets definable is not available.

4.11.8 DPS does not allow for any periodics and none is planned.

4.11.9 ATS allows for no fields of any kind.

4.11.10 Mark IV is designed to handle 99 periodics in any combination of up to nine hierarchical levels.

4.11.11 (Data) Central does not allow for any periodicity; however, system modification currently being considered will allow for 256 periodics for combinations of up to 15 levels of hierarchy.

4.11.12 Information has not been released.

5.0 File Maintenance

5.1 Update Language — All systems except DPS allow for an update language. DPS requires the complete entry to be deleted and re-entered without regard to a language per se.

5.2 Logical Maintenance — Logical maintenance here is defined as the capability of the system to modify all records (or entries) in the data base that set a given query condition. This differs from ordinary maintenance in that ordinary maintenance normally sets up a condition only on the record ID and logical maintenance allows for the condition on any field.

5.2.1 All versions of IDHS (Mark II, Mark III and SAC Mod 8) plus GIS, Mark IV, (Data) Central, and TDMS all allow for logical maintenance.

5.2.2 Both versions of the NIPS allow for logical maintenance from a linguistic standpoint as the only maintenance parameters.

5.2.3 Document processor and ATS have no logical maintenance capacity.

5.3 External File Conversion

5.3.1 ATS, Mark IV, Mark II - FFS, and Mark III - FFS allows for external conversion of position formatted data only. The above statement is a bit superfluous in the case of Mark IV in that Mark IV processes the data from a position formatted file directly and does not require conversion. The others actually convert to the internal structure.

5.3.2 Mod 8 allows for the external conversion of both position and comma formatted data.

5.3.3 (Data) Central and TDMS reports no restrictions.

5.3.4 All versions of the NIPS allow for external file conversion but only indirectly. The use of the assembly language POOL (and its 360 replacement) and the compiler level program language PCAL (and its 360 replacement) allows a programming user to convert external files, however, a separate program must be written.

5.4 Internal File Restructuring — NIPS 360 Phase I, DPS and ATS do not have any capacity for restructuring the data base. All other systems report that this capacity is inherent to the design.

5.5 User Supplied Standard Updates (Stored) — DPS and ATS do not have this capacity. (Data) Central is planning this for inclusion sometime next year. All other systems report that it is either operational for the second generation systems or will be operational when the system, itself, becomes operational (for third generation systems).

5.6 Audit Trails — TDMS and Mark IV are the only two systems that generate any type of audit trail, although a similitude of an audit trail is generated by all of the IDHS FFS's in their operational recording. No other audit trails are available.

6.0 Query (Search and Retrieval)

6.1 Mode

6.1.1 Conversationality — Only (Data) Central and TDMS have any type of direct conversationality in query. Although it could be said that ATS in maintaining the data base is really maintaining a document, the maintenance is in conversational mode. All other systems have no conversationality.

6.1.2 Queued Operation

6.1.2.1 Mark II - FFS allows for batch mode card input from local card readers only.

6.1.2.2 Mark III - FFS allows for batch mode card input from local card readers only.

6.1.2.3 NIPS 1410 allows for batch mode card input from local card readers and from remote terminals.

6.1.2.4 NIPS Phase I allows for batch mode card input from local card readers and from remote terminals.

6.1.2.5 NIPS Phase II allows for batch mode card input from local card readers and from remote terminals.

6.1.2.6 Mod 8 allows for batch mode card input from local card readers only.

6.1.2.7 GIS allows for batch mode card input from local card readers only.

6.1.2.8 DPS allows for batch mode card input from local card readers only.

6.1.2.9 ATS has no queued mode operation.

6.1.2.10 Informatics Mark IV allows for batch mode card input from local card readers only.

6.1.2.11 (Data) Central is currently planning to include, under its time sharing operation, a queued mode.

6.1.2.12 TDMS reports that it is primarily a conversational mode system and as such except for the necessity of a time sharing queue does not operate in queued mode (this means no batch capability).

6.2 Language and Capabilities — All systems with the exception of ATS have query language.

6.2.1 Stored Query

6.2.1.1 Mark II - FFS, Mark III - FFS, DPS, and ATS have no stored query capability.

6.2.1.2 (Data) Central does not currently have the capacity to store queries; however, this is being planned at this time.

6.2.1.3 All other systems have the capability to store standard user queries.

6.2.2 Search Method

6.2.2.1 Mark II - FFS is generally a serial search system. At file set up time, however, two fields of the file may be specified for indexing and the system will generate an index based upon the contents of these fields and for these fields then the system is index sequential.

6.2.2.2 Mark II - FFS is the same as the above.

6.2.2.3 NIPS 1410 is only a serial search system.

6.2.2.4 NIPS Phase I is serial except that it is index sequential for the record ID field.

6.2.2.5 NIPS Phase II is the same as the above.

6.2.2.6 Mod 8 is a serial search system.

6.2.2.7 GIS uses an index sequential system.

6.2.2.8 DPS uses a complicated set of inverted lists and parameters referring to inverted lists for its method of operation. It operates serially through the series of inverted lists and in this respect might be considered to be a form of chained mode where the inverted lists are themselves chained one to the other.

6.2.2.9 ATS is an index sequential search system.

6.2.2.10 Mark IV being primarily a tape oriented system is using a serial search.

6.2.2.11 (Data) Central utilizes one inverted file (inverted at the word within field level) for the whole data base.

6.2.2.12 TDMS utilizes an inverted file concept except that for hierarchical (periodic) data, separate chained inverted lists are maintained.

6.2.3 Multi-File

6.2.3.1 Mark II - FFS can process one file at a time, however, it is capable of merging the data from one file into a sub-file for future use.

6.2.3.2 Mark III - FFS is the same as the above.

6.2.3.3 NIPS 1410 processes one file at a time.

6.2.3.4 NIPS Phase I can process one file at a time, however, it is capable of merging the data from one file into a sub-file for future use.

6.2.3.5 NIPS Phase II is the same as the above.

6.2.3.6 Mod 8 can process one file at a time, however, it is capable of merging the data from one file into a sub-file for future use.

6.2.3.7 A maximum of 750 field-names are usable across 3 files; e.g. 250 for each 3; 375 for each of 2 or 1 file of 750 names.

6.2.3.8 DPS has no multi-file capability.

6.2.3.9 ATS has no multi-file capability.

6.2.3.10 Mark IV can process up to five files simultaneously.

6.2.3.11 (Data) Central can process 64 files simultaneously.

6.2.3.12 TDMS has no multi-file capability.

6.2.4 File Associated Tables

6.2.4.1 NIPS 1410, 360 Phase I, 360 Phase II and ATS do not use file associated tables.

6.2.4.2 All other systems can or do use file associated tables.

6.2.5 Operators Query Language

6.2.5.1 Logical Operators (AND, OR, NOT) — Only ATS which is not a querying system can not use all three operators directly.

6.2.5.2 Arithmetic Operators (LE, EQ, GT, etc.) — ATS is the only system that does not allow for arithmetic operation. All other systems allow for arithmetic searching. Both (Data) Central and TDMS however allow for automatic unit of measure conversion in arithmetic searching.

6.2.5.3 Between Operators — The Three NIPS systems and TDMS have the between operator for direct use against any arithmetic field. (Data) Central allows the use of the between operator (ALL) only for the record ID field. No other system uses between as a direct operator.

6.2.5.4 Geographic Searching — Only the military systems (all IDHS and NIPS systems) currently have the geographic search capability. (Data) Central is currently planning this inclusion for sometime immediately after the first of the year. No other system has this capability.

6.2.6 Language Features

6.2.6.1 Keyword-Key Phrase Capability — (Data) Central allows for full keywording and phrasing on all fields of the file. DPS allows for keywording and key phrasing on only the one variable length field defined. None

of the other evaluated systems have keywording except that TDMS reports planning this capability for inclusion approximately the end of 1968.

6.2.6.2 Partial word and Universal Character — Document Processing System has the partial word capability on suffixes specified. It does not have a universal character capability. (Data) Central has a universal character capability and the universal character when located at the end of a word can stand for any number of characters rather than only one. As a result the suffix word capability also is present. No other system has either capability.

6.2.6.3 Masking — Mark II - FFS, Mark III - FFS, the three NIPS, (Data) Central and DPS all allow for masking by use of the universal character. No other system has this capability.

6.2.6.4 Query Value Synonymy and Distance Searching — Both the DPS and (Data) Central have full synonym, equivalence and distance capability. The remaining systems have no capability in this respect.

6.2.6.5 Control Break Search — NIPS Phase I and II, ASTS, and, indirectly, the TDMS has control break search capability. Mod 8 has a control break capability except that it exists in the output package rather than the query package. All other systems have no capability in this respect.

6.2.6.6 Weighted Searching — None of the studies systems seem to have any direct weighted searching capability. TDMS as well as the output packages of the NIPS and the IDHS FFS's give the indirect capability by virtue of programming a counter.and outputting based upon the value of the counter. It is noted, however, that this is not a direct capability but rather an indirect capability.

6.2.6.7 Security Check — NIPS Phases I and II and the TDMS set up an overall file classification mode to prohibit at the file level the use of this file to unauthorized persons. (Data) Central establishes at the sub-field (word) level a security classification to prohibit individual items of data from unauthorized disclosure rather than prohibiting the complete file. All other systems have no direct security restrictive capabilities. It is noted, as in the previous paragraph, the capability

does exist indirectly by defining one or more security fields externally in the file and outputting based on the contents thereof.

6.2.6.8 Field Named Synonymy — NIPS 1410, NIPS Phases I and II, Mark IV and the TDMS allow for the use of field name synonymy. Other systems do not have this capability.

7.0 Maintenance Language

7.1 Language Type

7.1.1 Mark II and III - FFS, (Data) Central and TDMS have a compiler level maintenance language. Object decks are generated and used with the run.

7.1.2 The three versions of the NIPS systems have an assembly level and highly structured compiler level (very similar to COBOL language used for maintenance. In both cases, object decks are used.

7.1.3 The remaining systems use an interpretative operation and system maintenance language control cards or parameters (in all cases very highly structured) are interpreted at run time, and no object program is available or storable.

7.2 Language Functions

7.2.1 Delete — All systems have the capability to delete, in maintenance, complete entries from the file.

7.2.2 Replace — None of the defined systems utilizes the replace operator directly. The function is indirectly accomplished by a combination of delete and insert.

7.2.3 Insert (Add) — All systems have the capability of inserting (that is add) a new logical entry to the file.

7.2.4 Modify — The term "modify" as used here refers to the capacity of a system to modify an entry by the replacement of data in an existing field of the file. All systems studied have this capacity.

7.2.5 Randomizing and Conditional Logic in Maintenance — Conditional logic as used here is defined to relate to the combination of the function of logical maintenance and control break search. That is to say where under logical maintenance all entries of a file are to be maintained if they meet some conditional restriction(s). In conditional logic, this same function is accomplished only until an additional parameter has been satisfied (e.g. a control break) such as doing it only to the fields of

records until a record is found that meets the condition. It has been reported that Informatics Mark IV may have this capability. None of the other systems have either the conditional logic or the randomizing capability and in fact, systems that are not either index sequential or inverted search could not use the randomizing function.

8.0 Output Capabilities

8.1 Language

8.1.1 Type

8.1.1.1 Mark II and III, the three NIPS versions, (Data) Central and TDMS operate from a compiler level capability.

8.1.1.2 DPS and ATS have no output formatting capabilities directly.

8.1.1.3 Mark IV and Mod 8 utilize an interpretive mode of output language.

8.1.2 Pre-programmed or Stored Output Packages — An output package is considered stored in this sense regardless of its form (source or object). The systems listed above that operate at the compiler level can all store (ergo, pre-program the output package). Of the two interpretative mode operations only Mod 8 (not Mark IV) can store the output package in its source form. It is noted here, however, that the output package cannot be stored alone but must be stored as part of a query package.

8.1.3 Specified at Query Time — The subject of specification of output at query time relates to two separate functions; First the capability to specify which of the stored packages is desired (obviously if a package can be stored it must be able to be called) and second the capability to specify the contents of the package at run time. Of the compiler level operations only TDMS is capable of handling a "load and go" operation. The remaining compiler level systems are restricted to pre-compilation. The interpretative mode operations (Mod 8 and Mark IV) are always specified at one time.

8.1.4 System Generated — All systems that have any output capability (that is all except ATS and DPS) have the capability to allow for an internal system generated format, when no other format is specified.

8.1.5 User Controlled Volume

8.1.5.1 Mark II, Mark III and the three NIPS operations allow for user controlled volume by virtue of an automatically generated queriable page number. Volume is therefore controllable in page groups.

8.1.5.2 Under Mod 8, DPS, ATS and Mark IV, there is no user controlled volume capability.

8.1.5.3 In both (Data) Central and TDMS the conversationality of operation gives a complete user controlled volume at the record and/or field level.

8.2 Sort

8.2.1 Multi-File — Only Mark IV and (Data) Central have a multi-file sort capability.

8.2.2 Multi-Key

8.2.2.2 The three 1410 systems (Mark II, Mark III, and NIPS) allow for any combination of up to 25 characters in the sort key.

8.2.2.3 Mod 8 allows for specified seven different fields, one major and six minor, as the sort key except that for any field whose length is specified as being greater than 30 characters only the first 30 characters are used for the sort.

8.2.2.4 Mark IV allows for nine fields to be specified as a sort key and it is believed there is a character count restriction, although no documentation has been found to substantiate this impression.

8.2.2.5 Currently (Data) Central only allows for one sort key. A multiple key capability of unknown amount is currently being planned.

8.3 Generations

8.3.1 Sub-File — DPS and ATS cannot generate any sub-file. All other studied systems seem to have some capability in this respect.

8.3.2 Summary File — (Data) Central by virtue of its conversa-

tionality has not found a need for a summary file generation directly. ATS and DPS also do not have this capability. The remaining systems can generate a summary file during query phase for use during output phase.

8.3.3 Graphics

8.3.3.1 The current version of TSS-LUCID which SDC is operating as the developmental model for TDMS seems to have an extensive report generator package. It has been reported (although not confirmed) that this includes bar graphs, pie graphs, and a complete statistical evaluation sub-package.

8.3.3.2 Mod 8 as well as earlier models of 7094 FSS have a graphic display capability for printing only. This output package includes the capability to generate bar graphs and point graphs (both vertically and horizontally) with the scale of the graphs being system generated. In addition, it has the capability to generate minimal map overlays.

8.3.3.3 (Data) Central is currently programming a graphic capability (the CRT was delivered within the last month).

8.3.3.4 All other systems have no graphic capability.

8.3.4 System Controlled Volume

8.3.4.1 Mod 8 in system generation allows for the specification of a line item counter for system controlled volume.

8.3.4.2 TDMS, it is reported, (but it is not confirmed) also has system controlled volume.

8.3.4.3 NIPS 360 will, as a system, control the volume only during test phase.

8.3.4.4 The remaining systems have no system controlled volume.

8.4 Output Features

8.4.1 Output conditional logic — DPS and ATS have no output conditional capability. All other systems have this capacity.

8.4.2 Output Arithmetic Operators — The above comments apply here also.

8.4.3 Output Statistic Operators

8.4.3.1 If TDMS uses the aforementioned specifications for its output package, the statistical capabilities which include regression analysis, etc., will be vastly superior to any other RPG.

8.4.3.2 Both Mod 8 and Mark IV have the capability of specifying average only, taken over arithmetic fields.

8.4.3.3 (Data) Central using the operating systems RPG has a full range of statistical direct operators including average and deviation but no statistical analysis capability such as regression.

8.4.3.4 The remaining systems have no reported statistical capability.

8.4.4 Output Summary (Totals, Counts, Sub-Totals) etc.

8.4.4.1 Mark I, Mark III, the three NIPS all have total and count capability but no "sub" capabilities.

8.4.4.2 Mod 8 has the capability to summarize to two levels (e.g. counts and totals, sub-counts and sub-totals).

8.4.4.3 Mark IV has the capability for taking nine levels of totaling.

8.4.4.4 Both (Data) Central and TDMS have an unspecified number of levels for summarization capability.

8.4.4.5 DPS and ATS have no capability in this respect.

8.4.5 Sorting

8.4.5.1 All remaining systems have a measure of the capacity to specify the order of the output except that in both Mark II, and III and NIPS 1410 this capability is inherent to the query package rather than the output package.

8.4.6 Edit — All systems have a measure of this capacity.

8.4.7 LIST, DISPLAY, etc. — With the exception of DPS and ATS, all systems, dependent upon equipment availability, can specify various types of output such as list, punch, (referring either to cards or paper tape) and magnetic tape.

LEGEND:

Y = Yes

N = No

(P) = Planned

N = Note N

p = being programmed

S = sequential (serial)

INV = Inverted

IS = Index Sequential

A = Assembly

C = Compiler

I = Interpretative

Kc = 1000 characters

Ch = Characters

NOTES:

1. Information is not released.

2. "E" storage for DOS; "F" storage required for OS.

3. "G" storage for non-remote operations; "H" storage required when remote processing is available.

4. None for system per se except that if files take up all (sic) available disk space, then tapes are needed for sorting.

5. TDMS requires per se neither tape or DASD, Purchaser must have something, however, for both system and data base residency.

6. Resides as a separate system on the IBSYS system tape.

7. One variable length field only allowed per entry.

8. One variable length field allowed per segment.

9. Limited by core availability.

10. System is primarily a tape system — in addition a disk (2311) is needed if running under DOS.

11. Limited by DASD availability.

12. Via format number.

13. The only type allowed.

14. Position formatted files only.

15. Position and comma formatted only.

16. By writing a special (assembly or compiler) language program.

17. Index sequential on Record ID field only - otherwise serial.

18. Two fields can be specified at file set-up time for indexing and the system will allow for index sequential on these two fields.

19. Serial through a series of inverted lists — quite complicated set of lists.

20. Merge only — one file at a time into another.

21. By use of the universal character.

22. At the complete file level.

23. At the sub-field (word) level.

24. Essentially a control break search does exist except that it occurs in the "output" package rather than the query package.

25. Average only.

26. An extremely extensive RPG package exists here, which includes bar graphs, pie graphs, point plots and regression curves which implies a complete statistical evaluation sub-package must also be available.

27. Yes — answers here are dependent on equipment availability.

28. Although seven sort keys (fields) can be defined for sort purposes, only the first 30 characters in any of the specified fields are available for sorting.

29. IDS (Integrated Data Store) has been evaluated separately. IDS is not considered to be a data handling system in the same sense as the other specified systems. It is an addition of a series of VERBS to the COBOL language and to the COBOL compiler on the General Electric 600 series computers. It allows a PROGRAMMER (rather than a user) to affect the chained storage of data for any program that he is writing. All of the normal COBOL capabilities (and limitations) are present in the compiler; which affords the programmer a comprehensive set of tools with which to write a COBOL program for the processing of data. IDS, it is believed in the opinion of the viewer, should be evaluated with other programming systems (e.g. FORTRAN, RPG, JOVIAL, etc.) rather than with general purpose file-manipulating data handling systems.

30. Per segment rather than per entry.

31. A maximum of 750 field names are usable across 3 files; e.g. 250 for each 3; 375 for each of 2 or 1 file of 750 names.

HARDWARE				REMOTES			OPN SYS			FILE FORMAT																					MAINTENANCE						QUERY-RETRIEVAL								LANGUAGE QUERY LANGUAGE									
Central Processor Type	Memory Size	Mag Tapes Required	DASD Required	Cluster	TT/TTY	CRT	Batch	Multi-PGM	Recovery	Opn Rcd	Sectured Text (Sch)	Text (Maint)	Field/Entry Length	Nbr.	Max. Entry Size	# of Small Files	Key-Word User Gen	Key-Word Sys Gen	Stored (Object)	Unit of Measure	Security	Substit. Tables	Periodic Nbr/Level	Language	Logical	External Convers.	Internal Restr.	Stored Updates	Audits	Remote Inq. Conver.	Remote Inq. Qear	Language	Stored	Search Method	Multi-File	T.A. Tables	Op's Log/Arith.	Op's Bwn Geog	Op's Word & Phrase	Op's Partial	Op's Inv Ch.	Features Masking	Features Syn & Dist.	Features Ctl Break	Features Weight	Features Security	Features Fld Syn
1410	80K	6	1301	N	N	N	OP SYS	N	To NEXT	Y	Y	N	910 (7)	299	5440	1	Y	N	Y	N	N	Y	8/1	Y	Y	Y (14)	Y	Y	N	N	L	Y	N	S (18)	Y (20)	Y	Y	G	N	N	Y (21)	N	N	N	N	N
1410	80K	6	1301	N	N	N	OP SYS	N	To NEXT	Y	Y	N	N	910 (7)	299	5440	1	Y	N	Y	N	N	Y	50/1	Y	Y	Y (14)	Y	Y	N	N	L	Y	N	S (18)	Y (20)	Y	Y	G	N	N	Y (21)	N	N	N	N	N
1410	80K		1301	N	Y	N	OP SYS	N	To NEXT	N(+)	Y	N	N	99	2740	1	Y	N	Y	N	N	Y	8/1	Y	Y		Y	Y	N	N	R&L	Y	Y	S	1	N	Y	B/G	N	N	Y (21)	N	N	N	N	Y
1/50	G H (3)	8 (4)	4/2311	N	Y	N	OS	N	To NEXT	N	Y	N	N	200 (8)	250 (20)	@ 13KC	1	Y	N	Y	N	Y	Y	250/1	Y	Y (13)	Y (14)	N	Y	N	N	R&L	Y	Y	S (17)	Y (20)	N	Y	B/G	N	N	Y (21)	N	Y	N	Y (22)	Y
1/50	G H (3)	8 (4)	4/2311	N	Y	N	OS	N	To NEXT	N	Y	N	N	200 (8)	250 (20)	@ 13KC	1	Y	N	Y	N	Y	Y	250/1	Y	Y (13)	Y (14)	Y	Y	N	N	R&L	Y	Y	S (17)	Y (20)	N	Y	B/G	N	N	Y (21)	N	Y	N	Y (22)	Y
91 II	32K	12 or 3CH	2302	N	N	N	(6)	N	To NEXT	Y	Y	N	N	126	200	5882	1 INTO 7	Y	N	N	N	N	Y	9/1	Y	Y	Y (15)	Y	Y	N	N	L	Y	Y	S	Y (20)	Y	Y	G	N	N	N	N	N (24)	N	N	N
40 / 50 / 0	G H (1)	(1)	(1)	N	N	N	OS	I-N H-Y	To NEXT	(1)	Y	N	N	256	750 (31)	(9)	3 (31)	Y	N	N	N	N	(1)	15/15	Y	Y	(0)	Y	(0)	(1)	N	L	Y	(1)	15	3 (31)	(1)	Y	(1)	N	N	(1)	N	(1)	(1)	N	(1)
40	G	(5)	(5)	N	N	N	OS	N	To NEXT	(1)	Y(+)	Y (N)	N	249	255	1638 (+)	1	Y	Y	N	N	N	N	0	N	N	Y	N	N	N	N	L	Y	N	S-INV (17)	N	Y	Y	N	Y	Y-P N-UC	Y	Y	N	N	N	N
40 / 50 / 1440 (2)	(1)	(1)	(1)	N	Y	N	OS / OP SYS	Y	To USER	(1)	N	Y	Y	(9)	1	(9)	1	N	N	N	N	N	N	0	Y	N	Y (14)	N	N	N	R	N	N	N	15	N	N	N	N	N	N	N	N	Y	N	N	N
30 ↑	E (2) F	4(+) (10)	(16)	N	N	N	POS OS	N	To NEXT	AUDIT TRAILS	Y	N	N	NO LIMIT	(9)	(9)	5	Y	N	Y	N	N	N	99/9	Y	Y	Y (14)	Y	Y	Y	N	L	Y	Y	S	5	Y	Y	N	N	N	N	N	N	N	N	Y
30 ↑	F	0	ANY	N	Y	P	DOS OS	P	To USER	N	Y	Y	N (P)	ALL VERY	265	(11)	64	Y	Y	Y	Y	Y	N	25/15 (P)	Y (12)	Y	Y (12)	Y	(P)	N	R&L	P	Y	(P)	INV	64	Y	Y	B&G (P)	Y	Y (21)	Y (21)	Y	N	N	Y (23)	N
50 ↑ 1032	H	(5)	(6)	N	Y	P	OWN TSS	Y	To USER SELECTION	X	Y	N (P)	N (P)	256	(9)	(9)	1	Y	N (P)	Y	Y	Y	Y		Y	Y	Y (-)	Y	Y	Y	R&L	L	Y	Y	INV	N	Y	Y	B	N (P)	N	N	Y	N	Y (22)	Y

	MAINT. LANGUAGE						OUTPUT CAPABILITIES																
	FUNCTIONS						LANGUAGE					FEATURES											
	Type Ass'y Enter Compile	Delete	Replace	Insert (Add)	Modify	Randomizing	Condit.	Type	Pre-Pgm Std	Specif. Sys Gen	Usr Volume	Multi File	Sort # of keys	Sub File	Sum-File	Graphics	Sys Volume	Condit.	Arith.	Statistical	Summary	Sort	Edit	List, Display, etc
	C	Y	N	Y	Y	N	N	C	Y	Y	Y	Y	N	25 CH	Y	Y	N	N	Y	Y	N	Y-1	Y	Y	Y
	C	Y	N	Y	Y	N	N	C	Y	Y	Y	Y	N	25 CH	Y	Y	N	N	Y	Y	N	Y-1	Y	Y	Y
0)	A & C	Y	N	Y	Y	N	N	C	Y	Y	Y	Y	N	25 CH	Y	Y	N	N	Y	Y	N	Y-1	Y	Y	Y
) I	A & C	Y	N	Y	Y	N	N	C	Y	Y	Y	Y	N	210 CH	Y	Y	N	N	Y	Y	N	Y-1	Y	Y	Y
) II	A & C	Y	N	Y	Y	N	N	C	Y	Y	Y	Y	N	210 CH	Y	Y	N	N	Y	Y	N	Y-1	Y	Y	Y
8	I	Y	N	Y	Y	N	N	I	N	Y	Y	N	N	7 Flds (23)	Y	Y	Y	Y	Y	Y	Y (25)	Y-2	Y	Y	Y
		Y	N	Y	Y	N	N	⓪	①	①	①	①	①	①	①	①	①	①	①	①	①	①	①	①	①
	I	Y	N	Y	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	Y	N
	I	Y	N	Y	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	Y	N
	I	Y	N	Y	Y	N	?	I	Y	Y	Y	N	Y	9 Flds	Y	Y	N	N	Y	Y	Y (25)	Y-9	Y	Y	Y
	C	Y	N	Y	Y	N	N	C	Y	Y	Y	Y	Y	1-N(P)	Y	N	P	N	Y	Y	Y	Y-N	Y	Y	Y
	C	Y	N	Y	Y	N	N	C	Y	Y	Y	Y	N	?	Y	Y	Y (24)	Y	Y	Y	Y (26)	Y	Y	Y	Y
9)																									

that lumped together a number of different searchers. Only gross generalizations were available regarding whether end users were carrying out their own searches, whether searchers were frequently switching between data bases, or whether they tended to use video terminals instead of teletypes. After the site visits and before the workshop, matrices were drawn up summarizing the degree to which each of the eleven systems incorporated various features. At the workshop the matrices were used to structure discussion between designers. Revised versions of the matrices have been included as an appendix to the report. Revisions have been made partially at the request of designers and partially to make the matrices an accurate summary of the total report. System representatives have also reviewed the chapters of the report dealing with their systems and have made revisions so that the report accurately states the status of their systems as of April, 1973.

We would like to thank Donald Black and Robert Katter from System Development Corporation, Mark Radwin and Roger Summit from Lockheed, Lawrence Stevens and Howard Coleman from Informatics, Stanley Friedman from IBM, David Colombo and John Fried from Battelle Memorial Laboratories, Richard Giering from Mead Technology Laboratories, Donald Hillman and Louis Stern from Lehigh University, Benjamin Mittman and Wayne Dominick from Northwestern University, Richard Marcus from MIT, and Charles Goldstein from the NASA Lewis Research Center for their contributions to the workshop and report.

In addition to the written report, a twenty-minute color 16mm. film called "Access" has been prepared for introducing college level students to interactive searching. It is being distributed by the Extension Media Center, University of California at Berkeley. In the film, viewers are informed that searching is like looking for a needle in a haystack and that it

Report SU-COMM-ICR-74-1

A FEATURE ANALYSIS OF INTERACTIVE RETRIEVAL SYSTEMS

September 1974

Thomas H. Martin

Institute for Communication Research
Stanford University
Stanford, California 94305

Final Report for Period October 1972 - October 1974

Prepared for:

Office of Science Information Services
National Science Foundation
Washington, D.C. 20550

Research supported by the National Science Foundation under Research Grant NSF GN 36160 awarded to Stanford University in October 1972

Principal Investigator: Edwin B. Parker, Ph.D.

that lumped together a number of different searchers. Only gross generalizations were available regarding whether end users were carrying out their own searches, whether searchers were frequently switching between data bases, or whether they tended to use video terminals instead of teletypes. After the site visits and before the workshop, matrices were drawn up summarizing the degree to which each of the eleven systems incorporated various features. At the workshop the matrices were used to structure discussion between designers. Revised versions of the matrices have been included as an appendix to the report. Revisions have been made partially at the request of designers and partially to make the matrices an accurate summary of the total report. System representatives have also reviewed the chapters of the report dealing with their systems and have made revisions so that the report accurately states the status of their systems as of April, 1973.

We would like to thank Donald Black and Robert Katter from System Development Corporation, Mark Radwin and Roger Summit from Lockheed, Lawrence Stevens and Howard Coleman from Informatics, Stanley Friedman from IBM, David Colombo and John Fried from Battelle Memorial Laboratories, Richard Giering from Mead Technology Laboratories, Donald Hillman and Louis Stern from Lehigh University, Benjamin Mittman and Wayne Dominick from Northwestern University, Richard Marcus from MIT, and Charles Goldstein from the NASA Lewis Research Center for their contributions to the workshop and report.

In addition to the written report, a twenty-minute color 16mm. film called "Access" has been prepared for introducing college level students to interactive searching. It is being distributed by the Extension Media Center, University of California at Berkeley. In the film, viewers are informed that searching is like looking for a needle in a haystack and that it

2.0 THE SYSTEMS SELECTED

2.1 INTRODUCTION

Eleven systems (Battelle's BASIS, IBM's STAIRS, Lockheed's DIALOG, MIT's INTREX, NASA Lewis Research Center's NASIS, Lehigh's LEADER, Mead Technology Laboratories' DATA CENTRAL, Informatics' RECON, System Development Corporation's ORBIT II, Northwestern's RIQS, and Stanford's SPIRES II) were chosen for inclusion in the comparative analysis. The criteria used for selecting systems were that the system had to be 1) operational, 2) on-line and interactive, 3) able to handle multiple users simultaneously, 4) able to handle multiple data bases, 5) able to process data bases with variable length entries and elements, 6) demonstrable to the public, and 7) primarily oriented toward information storage and retrieval. In only one case was an exception made from these criteria: INTREX does not customarily handle more than one data base.

Many systems met the criteria of the comparative analysis but were not included. A number are close relatives of systems included in the analysis (ELHILL, NASA/RECON, BCN, LEXIS) so could be excluded without prejudicing the generality of the comparison. The versions included in the analysis were chosen because the person making the greatest contribution to the design of the system could represent it. Both DIALOG and RECON were included because Roger Summit of Lockheed and Larry Stevens (of NASA in 1968, of Informatics during the comparative analysis) worked together to establish the specifications for the original NASA/RECON.

Other systems were developed outside of the United States (QUIC/LAW of Canada, ISIS of Sweden) and it was felt that since system representatives

2.2.7 DATA CENTRAL was developed at Mead Technology Laboratories and was first put into service during 1968. At various times before April, 1973 it was used for searching Psychological Abstracts, Epilepsy Abstracts, and the case law of the state of Ohio. DATA CENTRAL is intended for use as a data base management system. Either the software can be leased or data bases can be placed on the parent system.

DATA CENTRAL

Since April, 1973 the system has been revised and is currently being reprogrammed to run on the DEC PDP-11 series of computers.

2.2.8 RECON was developed by Lockheed for NASA and is maintained by Informatics (in conjunction with their STIMS file maintenance package). Since 1968 the system has been extended in many different ways. RECON is installed not only at NASA, but also at the Department of Justice, the National Oceanographic and Atmospheric Administration, and other government agencies. Informatics supports RECON at these installations as well as selling service and making their computer available for storage of data bases.

RECON

Since April, 1973 The Environmental Protection Agency's ENVIRON data bases and George Washington University's POPINFORM data base have put onto the system. The TOXICON data base has been removed and has become the TOXLINE data base on ELHILL.

d. NTIS
Chemical Abstract's Condensates
* Science Information Associates handled marketing of searching service.

e. Chemical Abstract's Condensates
COMPENDEX (IEEE's Engineering Index)
** Metascience handled marketing of searching service.

2.3.3 In addition to the sites mentioned in 2.3.2, system software was also being used at the following locations.

a. ORBIT
b. RECON
c. STAIRS
d. DATA CENTRAL
e. NASIS

a. Department of State
Karolinska Institute (Stockholm, Sweden)
State University of New York (Syracuse)
b. Department of Justice (JURIS)
c. House of Representatives (Bill Status System)
d. Environmental Protection Agency
Wright-Paterson Air Force Base
Union Carbide
e. Carnegie-Mellon University

2.3.4 In addition to the data bases mentioned in 2.3.2, large data bases were available on the parent system for in-house use.

a. STAIRS
b. SPIRES
c. NASIS
d. INTREX

a. COMPENDEX
IBM's Technical Documents
b. MARC (Library of Congress's Machine Readable Cataloging data base)
c. ERTS (NASA's Earth Resources Satellite imagery data base)
d. the INTREX data base

2.3.5 As of April 23, 1973 the system was being used for some data bases that were neither bibliographic nor primarily textual.

RECON
DATA CENTRAL
BASIS
SPIRES
NASIS
RIQS

Report SU-COMM-ICR-74-1

A FEATURE ANALYSIS OF INTERACTIVE RETRIEVAL SYSTEMS

September 1974

Thomas H. Martin

Institute for Communication Research
Stanford University
Stanford, California 94305

Final Report for Period October 1972 - October 1974

Prepared for:

Office of Science Information Services
National Science Foundation
Washington, D.C. 20550

Research supported by the National Science Foundation under Research Grant NSF GN 36160 awarded to Stanford University in October 1972

Principal Investigator: Edwin B. Parker, Ph.D.

ild of the news/editorial department, the library can
est ally. If the library is crowded and disorganized, if
copies of articles, if subject classification is inconsis-
v, if photos are often missing, if reference books are
it's not surprising that users are dissatisfied. It's
ement tolerates the situation.

ive to an inadequate library. But good libraries don't
spaper's management chooses to commit itself to an
ion, then an improved library, as an integral part of
st be included in that commitment. The cost can be
the benefits immense as the information system

PGRADING THE LIBRARY

o valuable tools for improving a newspaper library
ewspaper Libraries, written by members of the
Special Libraries Association and published by and
erican Newspaper Publishers Association, and the
paper Division. The Division provides a consultation
ontinuing education seminars, annual conference
ape presentations. Aimed at increasing levels of
se activities are geared to practical working situa-

te books, equipment, and an adequate number of
el, and generally supporting library policy are keys
ient library operation which improves the quality of
uding the library in the planning process can help
eared to changing user needs.

an make the difference between a grudging half-an-
ic exta-effort which contributes the historical per-
to-research article, the newsworthy angle in an
ws story, or the information that sparks a feature.
day's paper were backed by library research? How
?

dustry, where the business is the dissemination of
o more important business than making that infor-
ccurate. And that's the business of the newspaper

B. ELECTRONIC APPROACHES

1980

by Ernest Perez
Librarian, *Chicago Sun-Times*

(Part of this article appeared in *Editor and Publisher*, January 12, 1980.

Several pioneering newspaper library operations are demonstrating that computerized library systems for news information and research are both practically and economically justifiable. Increased efficiency, reliability and recent cost decreases suggest that computerized library systems will be selected by many newspapers during the coming decade. Other factors are also making computer library systems attractive to newspaper management as an addition to and improvement of present editorial library information service capabilities.

The traditional newspaper research system has been the clipping file. Editorial staffs have used manual clip files since the 1800s, with varying degrees of success. The clipping file can be compact, systematic, convenient and timely — a fast information source limited only by physical filing capabilities. It meets the editorial requirement for convenient access to full text of new clippings (or other information) on a single subject, "Get me everything we have on..."

But the clipping file can also be bulky, inconsistent, cumbersome, and prone to misfiling, losses and physical deterioration. It also suffers from its very portability, the ability to wind up in the back of a reporter's desk, in his car, at home, or at his favorite bar.

Editorial research demands fast and convenient access to the complete text of previously printed stories or information file material. This full text may be viewed in several ways, including manual clipping files, microfilm images and reprints, or electronic VDT displays. Newspaper editorial staffs do not operate in the leisurely research fashion of academics or scientists, who are content with printed or computerized access to index listings of articles or publications. Although indexes can assist in research, most editorial people need the detailed facts and background information contained only in the complete text. Indexes, even those with headlines or abstracts, simply cannot convey the subtleties of information content which took a writer hundreds or thousands of words to cover in the original story.

dex listings, however sophisticated, also put substantial time and labor s at the research end of the system, unless there is automated access to ndexed material. A writer or library staff member must first manually te, display for viewing, and probably reprint, each item of possible e to see if it really does contain needed information. Lots of photocopies nade, money gets spent, and labor and researcher time invested, only scover that the desired information is not really there, after all.

mputer library systems are still evolving, but there are several de- lable and cost-efficient systems already in operation at major newspa- . These include the *Boston Globe*, the Louisville *Courier-Journal*, the nto *Globe & Mail* and the Philadelphia Newspapers Inc. (*Inquirer Daily News*).

rrently, two approaches are used for automated storage and quick ss to full text, microfilm images and electronic storage. One existing ry system uses microfiche photography for storage of text from velox eup copy or clippings of published articles; the other systems capture tronically readable text as a byproduct from the computerized editorial t-end and production system.

either case, the system software acts as an index locator or pointer to exact location of the full text of a particular item on microfiche or netic disk. This "internal index" then calls up full text from disk ge for VDT display or printout; or calls up the microphotograph of the red item on an automated microform reader/printer terminal.

MICROFORM STORAGE SYSTEMS

veral microform systems appear to be theoretically satisfactory, but nly actual installation is Info-Ky, developed at the Louisville *Cou- Journal* by their Dissly Research Corp. subsidiary. Info-Ky was in- ed at the *Courier-Journal* in 1972. After extensive testing, the *Cou- Journal* closed its clip file in 1976. The Info-Ky system has also been lled at the Allentown (PA) *Call-Chronicle* (1977), the Owensboro) *Messenger-Inquirer* (November, 1979) and is due for installation at Edmonton (Alberta) *Journal* in January, 1980.

fo-Ky uses a dedicated PDP-11 minicomputer. The smallest version a PDP-11/04 to control a single VDT and microfiche reader/printer inal inquiries. A typical installation for a major newspaper would de four VDTs, a PDP-11/34, a pair of disk drives, tape drive, line er and several of the computer-controlled microfiche reader/printer eval devices.

The Info-Ky system uses standard 4x6 microfiche with 192 frame format. Each frame measures 11" x 14" in original size and is reduced 42X by microphotography. The 11" x 14" frame can be filled with clippings or other desired items. These microfiche are then stored in computer-controlled microfiche retrieval devices with reader/printer capabilities. The microfiche retrieval devices, about the size of a large VDT, can store 780 microfiche. A conservative estimate of 2 clips per frame would allow about 300,000 clippings to be stored in a single device. Since only a single copy of each item is photographed, this will typically permit storage of approximately 3 years of clippings for a major newspaper in one microfilm retrieval device. When capacity is reached, a newspaper has options of adding more retrieval units to the system, or of storing infrequently used older fiche "offline" in other retrieval devices or manual files.

Info-Ky software stores an internal reference index which is manually created by library system personnel. Each item or article is examined and described by a skilled library classifier, then manually keyboarded into a VDT. The Info-Ky entry includes source, full headline, byline information, date and page citation, location on microfiche, and a short index description. The index description includes "keywords," meaningful names and terms appearing in the item text. Info-Ky classifiers also "enrich" the description with synonyms or important words that do not appear in the text. To aid in enrichment, Info-Ky has also developed a simplified classification system of about 120 defined broad terms and 33 news categories such as "editorial," "interview," "awards/honors," or "obituary." This simplified enrichment system is useful and powerful in getting efficient retrieval results. No story abstract is produced, since Info-Ky allows fast, automated display and reprint from the microfiche image of the original article.

Info-Ky is primarily designed for use by editorial staff, and features uncomplicated natural-language computer commands. To do research, a reporter goes to a work station consisting of an Info-Ky VDT next to an interfaced microfiche retrieval device. The reporter types in the terms and subject categories describing his information request. For example, the reporter could enter "Brown AND Ronstadt," to instruct the computer to locate only those articles indexed under *both* of those names. The system responds first with information about the items satisfying the query, and can then go on to display individual index entries. If desired, the user can then view articles displayed on the computer-controlled microfiche reader screen, and make any desired photocopies of the located items. If the number of items retrieved is too large or small, or not quite what is wanted, the user can modify the original query to limit, expand or redefine the search.

Info-Ky software automatically produces system use statistics which aid system management, problem area identification and information needs. stem statistics at the *Courier-Journal* show that 62% of all information arches on the system are being conducted directly by the editorial staff, indication of its acceptance and usability.

ELECTRONIC STORAGE SYSTEMS

Several electronic storage systems for newspaper text systems are eady operational or in development stages at major newspapers. The rliest operational installations were QL, Inc's "QL," at the *Toronto obe and Mail*, and Mead Data Central's "NewsMeadia" at the *Boston obe*", both fully operational in 1977. These systems were adaptations of text storage and retrieval systems originally used for legal case law and tutory texts, the American LEXIS system and the Canadian QL system. er modification to editorial research requirements, the systems were epted, and both newspapers have closed the old clipping files.

QL has also been selected by the *Los Angeles Times* and the *Philadel-ia Inquirer* and *Philadelphia Daily News*, and is in installation stage at se newspapers. The *Daily News* has been testing the system, found it eptable, and will close the clipping file the end of 1979. They will begin system operation with an electronic full text file going back to January, 3. The *Inquirer* will begin storing text on the QL system starting in uary, 1980, and will run a dual clipping file/computer text operation for a onth final testing period.

n early 1979, Mead Corp. announced that it is withdrawing the specia-d NewsMedia software package from the market. No other newspaper ept the *Boston Globe* had ever actually installed the system.

he *Boston Globe*, along with the *Chicago Sun-Times*, is currently ing a prototype system being developed jointly by Atex, Inc. and Infotex ociates (Dayton, OH.)

nother system, now in development stage at the *St. Louis Post-Dis-ch*, is UNIVAC's UNIDAS/1100 system. Although UNIDAS/1100 will e full text on disk, it will retrieve information by searching manually-gned index terms.

L and NewsMedia both require an IBM mainframe on a time-sharing s. UNIDAS/1100 will run in time-sharing mode on a UNIVAC 1100 puter. The Infotex system is designed for a dedicated PDP-11/34.

Infotex will be a stand-alone system, although it can be interfaced to editorial front-end systems. Atex is planning to create the interface to their front-end system when the Infotex testing phase nears completion. Such a system interface should be done by or in cooperation with a front-end system vendor to avoid negative effects upon front-end system performance.

An electronic full text system will normally process selected text captured from the front-end system for input into the library data base. Magnetic tape is an obvious method for transferring the text, although system interface could allow disk-to-disk transfer, and eliminate the separate tape input step.

The electronic full text systems will store full text of articles or information items on magnetic disk. During input processing, system software automatically indexes every meaningful word appearing in the full text and creates index pointers to location of full text in disk storage. Indexing of "noise" or meaningless words that would use up disk space for no valuable purpose is avoided by creation of a "a stop word" list by library system managers. The indexing software then ignores these defined stop words. A stop word list usually includes common articles, prepositions, conjunctions and terms like "whereof," "almost," "very," "perhaps," etc. Additional defined lists of synonyms, abbreviations and irregular plurals will direct the system to automatically locate equivalent terms.

Electronic full text systems software also allows text enrichment input by library staff, to permit addition of subjective concepts, synonyms, or terms not appearing in the text. For example, "Mafia" or "Syndicate" both could be added to a story about organized crime.

The internal index is not printed or meant to be used by a human researcher. It is a literal index to location of every word in the entire data base. The system software uses it to locate full text in answer to user information requests.

A TYPICAL ON-LINE SEARCH

A typical electronic storage system search begins with VDT entry of ords or phrases likely to have appeared in the desired stories, or to have :en added as enrichment terms. Complex term relationships can be :fined by use of "connectors" such as "AND," "NOT," "OR,"and "ADJ" djacent). The system will typically respond with display of the number of :ms in the data base which meet defined conditions. If the user is satisfied th search results, he can examine full texts or sections of texts to either t he desired information or discover necessity for refinement of his arch definition. For example, entry of a city councilman's name may trieve 497 items; adding "AND building code" may narrow results to 27 ms; further limiting search to articles dealing with a particular company project, or to a particular byline may restrict results to 5 or 7 items.

Electronic full text systems can then quickly display full text of the ired stories on special system VDTs located in the library or other as, on interfaced editorial system VDTs, or at remote locations via phone couplers. These systems can also output retrieved information in er forms, including hard copy printout from line printers, tape or disk ies for customized delivery of specialized files to other systems, for ut to phototypesetting systems for reprinting, or for production of rofilm via Computer Output Micrfilm (COM). The COM specialized s might be useful for system backup, for specialized editorial files on rofilm to be used at remote bureaus, on the road, etc., as an alternative ulky printout or clipping files.

he computerized newspaper library systems described are turn-key allations, but do have flexibility for customizing to a particular newspa- s input format or to defined information file specifications. The systems e multiple options available for VDT screen display format, command uage, display order, etc., and a system installation gets its normal rational appearance by the selection of automatic or "default" options. other optional modes are still available on request, but the default ces are used automatically, if no other command is given.

COSTS

Computer library system costs have come down to a point where auto mated newspaper library systems offer a realistic and perhaps superior alternative to the traditional manual library systems.

The Info-Ky system ranges from $150,000 to $300,000 depending upon configuration and complexity. This includes software, complete hardware installation and training in operation.

The QL retrieval software is available for $15,000/yr on lease basis, or for purchase cost of $50,000. Associated text editing software costs $5,000/yr for lease, or $15,000 for purchase. A percentage of lease payments can be applied to purchase cost. QL retrieval and editing systems run in time sharing mode in single partition of an IBM 360-30, an IBM 370-125, or larger IBM computers. QL also requires normal IBM peripheral hardware.

The Infotex system is still in pre-production testing stages. Total installed cost for a system capable of supporting a large metropolitan newspaper editorial operation is estimated at $400,000 to $500,000. This includes the complete package of software rights, PDP 11/34 and peripherals, system VDTs, disk drives sufficient to hold approximately 5 years of full text, complete installation customizing and staff training. A system suitable for a smaller newspaper would cost around $200,000.

UNIDAS/1100 is in developmental stage of newspaper text file applica- tion, and no definite price information was available.

These are the major computerized newspaper library systems in actual use and development at newspapers on this continent. Similar systems exist at publications in Europe. Utilization of today's more powerful mini-com- puters, hardware cost reductions and the dramatic and continuing reduc- tions in storage costs seem certain to make the computerized newspaper information file a practical and commonplace reality. All these develop- ments, especially storage technology advances, appear to be progressing at a rate which will prevent newspaper text library system operating cost from ever becoming a real problem.

There are operating and cost advantages immediately available from these systems. Obvious benefits are library time and labor savings from elimination of the manual file; space savings from stopping further manual file expansion, and reducing the old manual backfile to more permanent and organized microfilm form; the security and permanence of text file records, assuming the security backup of any well-designed automated system. This last factor can be contrasted to the inevitable lost, destroyed or stolen individual clippings and whole files in traditional newspaper libraries.

EDITORIAL RESEARCH BENEFITS

addition to the improved capabilities for published text access, com-
r library systems make possible efficiencies in maintenance and crea-
of valuable library peripheral file information. A system could include
ialized data bases for listings of physical library file holdings and dates
le coverage, photo and photo negative holdings, book or pamphlet file
ction listings, present obits, unpublished or overset material with
able information content, etc. Some computer library systems might
be able to function as location status files, with file location informa-
or indication of physical file checkout by editorial staff members.

reporter or editor could inquire as to "what we have" on "John Doe" at
rary system or interfaced editorial system VDT, and learn in a few
ients that "we have" microfilmed clippings on John Doe from 1957 to
, a photo file from 1957 to present, 14 envelopes of staff photographer
itives including photos of him (with a description of each photo assign-
t), a recent pamphlet about Doe's company, and the fact that George
s, the Photo Editor, checked out the photo file yesterday.

ıt computer library systems will immediately add new capabilities for
orial staff. These systems offer all the incredible power of online VDT
rmation retrieval to editorial staff previously limited by manual infor-
ion files. The systems will give power and speed to specific and general
rmation retrieval; allow multiple editorial users to have simultaneous
ss to the same material; provide immediate positive or negative
rmation about total file contents. Writers and editors will have near-in-
t access to all background or factual information in the library data
, or know immediately that "we haven't done anything on that."

mputer library systems will effectively automate journalistic re-
ch, the link that is now the slowest and most expensive part of the whole
orial and prdouction chain. It seems ironic that the power of online
rmation systems has been almost ignored by print journalists, the main
rmation professionals. But in the past, this has been due to the experi-
tal and unreliable nature of the systems, and high storage costs.

line news library systems will give new powers to ediorial staff,
nitting a new kind of journalistic investigative approach. Full text
ems, for example, can enable a reporter to quickly search through the
e library data base to find *any* occurrence of a name, or word
ping, to check for related facts or incidents using spelling variations,
et names or addresses, datelines, ages, residential areas, professions,
y other defined specifications. This power is something that journalists
never had, and will really have an effect on editorial content qual-

These systems can also automatically create all the specialized editorial specialist files that the library has never had the time or money or space to be able to offer. A full text electronic file would, in effect, "create" high school sports files; files under every name of divorce, marriage, obituary listings; files of editorials; action line columns, fashion and society columns; recipe files, etc.

All stories in these files would be identifiable by type, date of publication, length, wire service or local, byline, page or section, etc., as well as by every meaningful word in the story. A full text system would enable the Editorial Library to offer this kind of research power to every specialized department and individual user at no extra cost, and also relieve editorial personnel from the labor and time and space costs of maintaining all the special files that seem to exist everywhere in a newspaper. Editorial staff members can then devote more attention to the primary journalistic task, increasing their own productivity and improving the editorial product quality.

Newspaper librarians have a strong interest in the capabilities and features of the developing computer library systems, so that they will be able to offer the kind of system performance just described. For a number of years, members of the newspaper librarian professional group have conducted an extensive analysis of system features with vendors and with the users of the initial installations. The Automation Committee of the Newspaper Division, Special Libraries Association, chaired by Jim Scofield of the *St. Petersburg (FL) Times and Evening Independent*, has now approved a set of performance standards for electronic storage systems. This document, "Basic Specifications for A Full-Text, On-Line Newspaper Library System" will be available shortly from the Special Libraries Assn. (Editor's note: Plans were changed and it was prepared by the Division.)

NEWS FILE MARKETING

The final intriguing possibility is that of using the library information system as a new profit center, of marketing the library data base to outsiders via remote VDT and printer connections. The profitability of remote data base marketing has been demonstrated by systems such as the *New York Times* Information Bank, the Dow-Jones data base, the LEXIS and QL legal systems, Lockheed DIALOG, Systems Development Corporation's ORBIT, etc. The demands of our information-hungry society has made all of those operations profitable, even though most are only indexes, not full text, requiring specialized training and lots of text location labor for efficient information retrieval.

Most of these data bases are created by manual indexing and input, thus have a high cost for original data base creation. But full text newspaper data bases have the ability to generate new income from a service department normally regarded as an unavoidable overhead expense. Both microform storage library systems and the electronic storage full text systems can cost-justify on their own merits as efficient service department operations. But both types of systems offer the possibilities of functioning as profit centers, with potentially strong regional markets for information content, as well as lesser national markets. For example, the *Louisville Courier-Journal's* Info-Ky has already made several remote installations of their data base; and the *Toronto Globe and Mail's* "INFO GLOBE" subsidiary had 110 remote VDT users signed up in November, 1979, the first year of its marketing operations.

Passive and active information communications systems using cable television or broadcast transmissions are now being tested throughout the world. Distribution systems such as Prestel, Teletext and Qube offer the possibility of profitable mass-marketing of computer library system newspaper file information to private users, as well as to companies or organizations having sophisticated VDT remote access capabilities. Newspaper full text systems have an advantage for this kind of an information distribution, since they can offer remote display of information at no extra cost, as a by-product of their editorial front-end and library text systems. Library text systems could interface to the communications system and permit remote paid access to authorized "Public file" material for private citizens, home viewers, students etc.

News information stored primarily for editorial research value could thus be resold repeatedly, lowering system operating expense and very possibly generating profits. Inexpensive distribution systems delivering individualized text information to private users suggest that newspaper libraries could function as the "information utility companies" of the future. The electronic news library with mass distribution capability *could* outperform the daily publication as an information delivery outlet, because of the sheer volume of information contained. It could not compete with current publication in advertising revenue, at least not at this point ... these possibilities are yet to be explored.

ADDRESSES

Marie Wilmor, Info-Ky Systems
DRC,Inc.
620 S. 5th St.
Louisville, Ky. 40202
(502) 582-4655

Dick Giering
Infotex Associates
3103 S. Kettering Rd.
Dayton. OH. 45439
(513) 293-4173

Carol Marble, Newspaper Systems
QL Systems, Ltd.,
103 Sheraton Centre
90 Richmond St. W.
Toronto. Ontario.
M5H 2A3 Canada.
(416) 862-7656

BASIC SPECIFICATIONS FOR A FULL-TEXT ON-LINE AUTOMATED NEWSPAPER LIBRARY SYSTEM

Automation Committee
Newspaper Division/Special Libraries Association

TABLE OF CONTENTS

DTN-72-2

copy to Tomli

Chapt. 8

TECHNICAL NOTE

THIS IS DATA CENTRAL

(1972 TECHNICAL SPECIFICATIONS)

Richard H. Gierig

March 1972

Data corporation

DTN-72-2

TECHNICAL NOTE

THIS IS DATA CENTRAL

(1972 TECHNICAL SPECIFICATIONS)

by

Richard H. Giering
System Designer

March 1972

Prepared By

Data Corporation
3481 Dayton-Xenia Road
Dayton, Ohio 45432

Data corporation

FOREWORD

These technical specifications relate to a proprietary system and, as such, are themselves considered proprietary.

The form that was used in preparing this report was the form used by the "CODASYL SYSTEMS COMMITTEE" in its report titled "A Survey of Generalized Data Base Management Systems," non-copyrighted except that the source is hereby acknowledged and appreciation for the work of the committee is tendered by this author. Copies of the "CODASYL" committee report are available through the ACM, 1133 Avenue of the Americas, New York, New York 10036.

This report may be reproduced as long as appropriate bibliographic reference is made to this report. Additional copies are available upon request.

I. INTRODUCTION TO SYSTEM

1.1 Identification - Data Central Information Handling System is a proprietary software system developed and made available by Data Corporation (a subsidiary of The Mead Corporation), 3481 Dayton-Xenia Road, Dayton, Ohio 45432.

1.2 Status - Data Central is fully operational for System 360/370 computers. The system is complete with the use of major subsystems.

1.2.1 The on-line subsystem operates in an interactive manner from on-line terminals for the search and retrieval of material stored in Data Central data bases.

1.2.2 The update subsystem operates in a "batch" environment allowing for additions, deletions and/or changes to individual fields (segments) of the file entries.

1.2.3 DBD subsystem allows for a flexible specification of the file(s) in a data base.

1.3 System Background - In early 1966 a feasibility prototype of a total information system with an inverted file at the word level was operated. This proved the feasibility of approaching information processing in the manner Data Central now approaches it. Concentrated development effort was implemented in mid-1967. By late 1967 and early 1968, enough progress had been made in this developmental system to allow for it to be used operationally in support of corporate objectives. Since that time constant enhancement has taken place such that the system herein described is fully operational including all enhancements specified herein. It is being and has been used in many applications. A few examples are listed:

1.3.1 COMPUTERIZATION OF LEGAL RESEARCH - Mead Data Central, Inc. (another subsidiary of The Mead Corporation) applies the same technology to assist lawyers in research. A current contract with the Ohio State Bar Association allows Ohio lawyers to search all the statutory and decisional law in the State of Ohio. Additionally, a contract was signed with the New York State Bar Association, and several other state negotiations are in progress. The MDC corporate objective is to mount data bases comprised of the law covering the fifty states and the Federal Government. The first phase of the Federal data base is represented by a contract with the Internal Revenue Service. The data base, for Ohio alone, exceeds a half-billion characters of source data.

1.3.2 AMERICAN PSYCHOLOGICAL ASSOCIATION - The American Psychological Association has created a data base containing the past four years' Psychological Abstracts. APA performs a service to its members by allowing them to retrieve information from the Psychological Abstracts both in an on-line mode or batch mode. APA hopes to increase the content of their data base by offering other "soft sciences" files in addition to the Psychological Abstracts file.

1.3.3 DEPARTMENT OF HEALTH, EDUCATION AND WELFARE - HEW has created a data base that monitors an inventory (list) of all publications produced by any of its components (i.e., Food and Drug Administration, National Institutes of Health, etc.). A sample record describing a publication includes its title, its publication number, the issuing agency, a brief description, and cost information. These data are available to HEW through on-line query. Additionally, the capability to produce a computer "typesetting" interface has been created to allow HEW to produce quarterly and annual catalogues of these publications automatically from Data Central.

1.3.4 PERSONNEL (MEAD CORPORATION) - Mead Corporation is maintaining files of all their employees' personnel records, including each individual's safety record. This is a very dynamic file and represents a replacement for their manual record keeping system.

1.3.5 WALTER REED ARMY MEDICAL CENTER - A data base containing all relevant documents in the field of biological effects of non-electromagnetic radiation (BEER) has been created. This data base will allow medical researchers to discover quickly previously published research papers that are relevant to work they might be pursuing.

1.3.6 ENVIRONMENTAL PROTECTION AGENCY - This project entails a network of terminals spread throughout the entire nation in most EPA Regional Offices. The data base contains a wide variety of technical and management data. Some examples of these data are active research and development projects, proposed research projects, all oil and hazardous spills that have occurred nationwide, and information regarding all water quality monitoring stations throughout the nation.

1.4 Major Characteristics

1.4.1 The system provides for multiple level data structures. Groups may be defined in an N-level hierarchy or in single-level super-set configurations or both.

1.4.2 Generalized Processes Provided

1.4.2.1 For applications or systems oriented ADP personnel, the system provides for data base definition (a file is considered to be one hierarchy below data base), data base creation and updating and interface to specialized functions and formats written in any of the existing programming languages.

1.4.2.2 For non-ADP oriented terminal users the system provides for data base, file and/or field level interrogation at the "WORD" level (where a "WORD" is defined as a string of characters between two delimiters [see Section 3.2.1.3 below] in a field, and for on-line tutorial (or CAI) upon request.

1.4.3 Language Type(s) - Four different language types are employed by Data Central.

1.4.3.1 Data base definition is a specialized procedural language.

1.4.3.2 Data base creation and updating is a replacement-procedural data content oriented language.

1.4.3.3 Input and output/display format generation is allowed via any of the existing procedural programming languages.

1.4.3.4 Data base inquiry is a specialized procedural language employing high level procedural operators and Boolean connectors.

1.4.4 Language Form

1.4.4.1 Data base definition language is tabular.

1.4.4.2 Data base creation and/or updating is predominately string-oriented with tabular control.

1.4.4.3 Display formatting is either string (if language used is higher level) or tabular (if language used is Assembler).

1.4.4.4 The inquiry is completely string (quasi-English) oriented.

1.4.5 The internal file structure of Data Central is a completely controlled proprietary index sequential structure wherein the data content of the user-defined data base(s) is stored in both a sequential (serial-indexed) and inverted (word-indexed) mode (see Section 4 below). The structure is oriented toward terminal-inquiry response and the user has no physical control over the internal structure. The user has 100% control over the "apparent or logical" file structure with which he interfaces.

1.4.6 Modes of use

1.4.6.1 On-line use is available for interactive interrogation and display "browsing".

1.4.6.2 Batch use is available for all of the non-interactive functions (e.g., data base creation, definition, updating, etc.) and is also available for large-volume retrieval displays resulting from interrogation.

1.4.7 Data Central operates on any 360/370 equipment and supports the following terminal types: TTY, 1050, 2740, 2741, 2260, CC-30.

1.4.8 File Media

1.4.8.1 Data input is via any machine-readable media including tape, cards, OCR, cartridge, etc.

1.4.8.2 Data storage is primarily in direct access mode (disk or data cell) with tape providing backup.

1.4.9 Operating System Environment - Data Central operates under either DOS or OS with its own teleprocessing roll-out/roll-in executive.

1.5 Overall Philosophy - The basic design philosophy of Data Central is to enable a non-ADP class user to ask questions in an unsophisticated manner of an existing data base and receive rational and direct answers to that query in a reasonable amount of time.

As a result of this philosophy, the following design criteria have been embodied.

1.5.1 Interrogation Oriented - Data Central is an interrogation-oriented system. All other functions available within the design philosophy of Data Central are subordinate to the function of answering questions when questions are posed. It has been stated that people have literally spent billions of dollars building data bases, but were not able to use them to answer the questions when they were posed. It took time to get the answers. Relatively speaking, immediate response is a design criteria of an interrogation-oriented system.

1.5.2 The communication between man and machine must be in quasi-English and understandable form and it must be supported by a fully usable tutorial (or computer aided instruction) capability. The interrogators generally will be unsophisticated with respect to the use of ADP-oriented equipment.

1.5.3 The Data Spectrum - Since the human beings interrogating the system are unsophisticated with respect to ADP, the chances are they will have little or no use for any system fully predicated on codes of any type. Data Central is therefore capable of handling codes but is not limited to that. All words in a textual orientation of fields as well as arithmetic values are available for use as a search criteria.

1.6 Documentation - available in proprietary form.

1.6.1 Terminal Operator User's Manual (Black Book)

1.6.2 Data Base Management Manual (Blue Book)

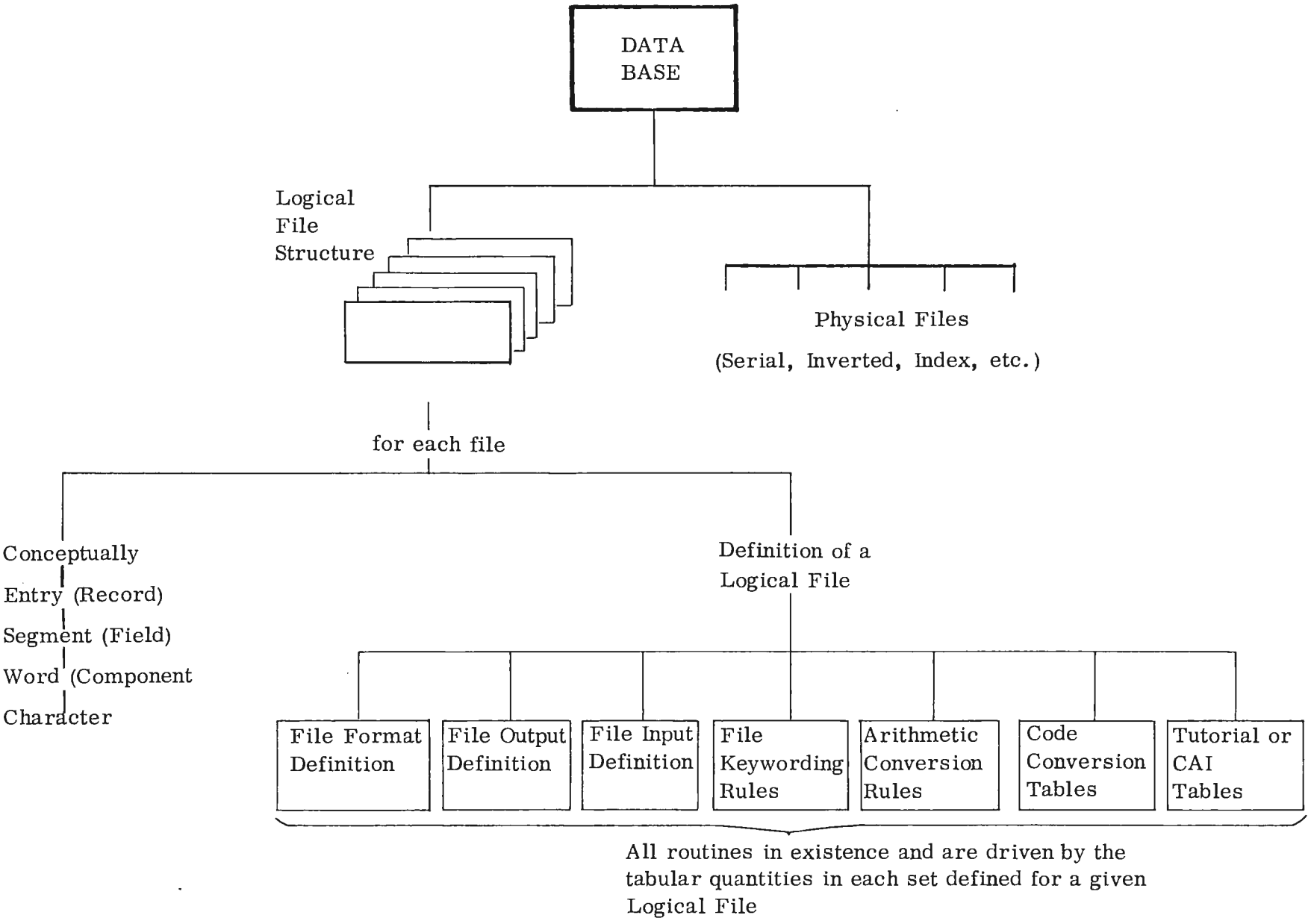
1.6.3 Computer Operator User's Guide (Green Book)

2. DATA STRUCTURE

Since the internal data structure of Data Central is unique and proprietary, the user's data structure (externally viewed structure) bears little correlation to the internal storage structure. The external vs. internal structure is exemplified in the figure on the following page. The physical file and structure are that with which the D.C. programmers operate; the logical structure (left side) is that viewed by the user.

2.1 Item

2.1.1 System's term for items: Segment.


DATA BASE
Logical File Structure
Physical Files
(Serial, Inverted, Index, etc.)
for each file
Conceptually
Entry (Record)
Segment (Field)
Word (Component
Character
Definition of a Logical File
File Format Definition
File Output Definition
File Input Definition
File Keywording Rules
Arithmetic Conversion Rules
Code Conversion Tables
Tutorial or CAI Tables
All routines in existence and are driven by the tabular quantities in each set defined for a given Logical File

2.1.2 Item Naming - Item names are strings of up to 40 alphanumeric characters in length. No embedded blanks or commas are allowed, nor is a leading dollar sign allowed. Item synonyms are allowed to any level and output "headers" are definable (synonyms may use the same header or may use different headers).

2.1.3 Data Item Types - Each segment is considered to be made up of words (defined as a string of characters - not necessarily a valid natural language word - between two delimiters as defined in the file keywording rules). Additionally, segments defined as being "arithmetic" may have as the "initial set of words" a group (or series) of values (numbers associated with a predefined or implied unit of measure) followed by "text." Using the rules established in the "arithmetic conversion rules," the data are transformed into a single searchable unit. Those acceptable in this mode include (but are not limited to):

Dates of the form
- dd/MM/YY (numeric)
- MM/dd/YY (numeric)
- ddmmmyy (month alpha)
- yyymmmdd
- YYY/MM/DD

Groups by table of units of measure (e.g., 5 meters, 16 CM, 2 MM)

Note that all data entered into the data base are considered to be directly displayable. External numeric data in any of the following forms are acceptable via an input formatting subroutine:

- Packed Decimal
- Hexadecimal
- Hexatridecimal (Base 36)
- Zoned Decimal (character)
- Binary (either fixed or floating point)

2.1.4 All segments are considered (by the system) to be of variable length. The data base designer has the option to constrain, at input time, a given segment to a given length. The variable length segments have a length limit of 65,535 non-trivial words.

2.1.5 The system treats multiple valued segments as repeating (periodic groups). See 2.2 below.

2.1.6 Sub-items. Each segment is considered to be made up of a multiple number of sub-items at the word, word group, or phrase level. Additionally, items defined as containing arithmetic data may have sub-levels (see 2.1.3 - Unit of Measure above).

2.2 Groups - Any segment may be defined as a segment or grouping segment. A grouping or pseudo segment may have up to 255 associated "real" segments. Grouping or pseudo segments may be defined as one of two types:

2.2.0.1 "OR Pointer" - The associated real segments for an "OR Pointer" pseudo segment are considered to be "OR" related in search. For example, assume two segments of a personnel file are "pay-location" and "work-location"; assume also a group (or pseudo) segment defines "location" as the "OR pointer" of both "pay-location" and "work-location." A search specified as "Find any entry whose location is Dayton in the state of Ohio," would be specified as:

Location = Dayton and Ohio

It is considered equivalent to:

(pay-location = Dayton and Ohio) or (work-location = Dayton and Ohio)

2.2.0.2 "Concatenation" - In this form of pseudo-segment the associated real fields are considered to be concatenated together into one "new" segment. For example, consider 2 other segments of the personnel file: "current-job-description" and "previous-job-history." A request to find all entries in which the "job-description" (defined as the concatenation of both segments) contains both the word Aircraft and the word Missile is stated:

Job-description = Aircraft and Missile

The "OR-Pointer" explosion of this request is:

(current-job-description = Aircraft and Missile) OR
(previous-job-history = Aircraft and Missile)

The "concatenation" explosion adds the additional criteria (since the customer is looking at the pseudo as a new single segment):

OR

((current-job-description = Aircraft) and (previous-job-history = Missile))

OR

((current-job-description = Missile) and (previous-job-history = Aircraft))

2.2.1 A grouping segment is known as a "pseudo segment".

2.2.2 Group structure (see 2.2 above for two types of pseudos) - There is no limitation on the number of pseudo-segments that may be defined in a logical file (pseudo segments may be cross-file oriented). The total number of segments of ALL types (real or pseudo) is $(2^{16}-4)$. There is no restriction on the number of pseudo-segments with which a real segment may be associated. Each unique pseudo-segment is identified by a segment name as are real segments.

2.2.3 Group relationships - The data base designer (who assigns the segment names) may use ANY method for defining intergroup relationships. Since segment names may contain parentheses, a common method is the implied subscript notation. This is not mandatory however and any method is allowed. Internally the real associations imply the hierarchic relationships.

2.2.4 Group identification is via real segment association during the data base definition phase.

2.2.5 Types of groups - see 2.2 above.

2.3 Entries

2.3.1 System Term - The system term for an entry is "ENTRY" - prior use included the term "DOCUMENT," although this term is slowly losing its use.

2.3.2 Entry Types - There is only one entry type (defined as the "File Format" of segments) per logical file; 255 logical files are allowed by Data Central in one physical (and logical) data base.

2.3.3 Entry Identification - The system assigns (or the computer may assign) a unique entry identification number to each entry. Entry number (E) may be such that:

$$1 \leq E \leq (2^{24}-1)$$

2.4 Logical Files (reference is made to discussions in paragraph 2 above and its associated figure).

2.4.1 System Term - Logical File(s)

2.4.2 File Type - Data Central allows for 255 unique file types. There is no restriction as to the make-up or interfile relationships. Additionally, pseudo files may be defined with real segment associations in a cross-file manner. No cross-file restrictions are placed on the data base designer.

2.4.3 File Identification - The file is defined via an acronym of up to 8 alphanumeric characters in length. No embedded blanks or comma(s) are allowed.

2.5 Data Structure Generalizations - See discussion of pseudo-segments in paragraph 2.2 above and discussion of pseudo-files in paragraph 2.4.2 above.

2.6 Data Security - A user accesses the on-line system via an "access key". Associated with this key is a list of logical files he may access and a pair of numbers indicating his "security range". For the user to have access to the data in that segment (either in search or retrieval) his range must include that segment's level. To recap, 3 levels of security are available. The access key obtains access to only certain data bases; certain files of the accessible data base may be "non-existent" to the user and certain segments in the available files may be "non-existent".

3. FUNCTIONS

The functions available to Data Central users are multi-leveled. Many of the answers below will be multi-faceted as a result. The functions are:

- On-line Interrogation
  - . Search Specification
  - . Retrieval Criteria
  - . Display Format
  - . Dialog Recursiveness
- Data Base Creation and Update
- Data Base Definition
- Interface Programming

3.1 Language Form

3.1.1 Search Language -On-line interrogation is via free form (quasi-English) with high level operators and Boolean connectors.

3.1.2 Retrieval and display specification is pre-stored in procedural tabular form; in use it is called for via an acronym.

3.1.3 Creation and update is procedure-implied tabular form.

3.1.4 Data base definition is via a procedure-oriented tabular language.

3.1.5 Interface programming is via any of the procedural languages available on the computer upon which Data Central is implemented.

3.2 Data Base Definition (reference is made to the discussion in Section 2 and the figure).

3.2.1 Definition of data items is via the assignment of an alpha-numeric string ($\leq$ 40 characters in length) as the identifier (acronym) and the assignment of a unique "segment number".

3.2.1.1 Definition of Data Item Type - All segments are considered to be "logical" or textual in nature. In the assignment of the acronym, however, the designer may indicate (by punching in one tabular column) those segments that are also to be considered arithmetic in form.

3.2.1.2 Item Length - Data Central requires no length specification.

3.2.1.3 Multi-valued definition is via the specification of the "keywording rules" (for textual items) and the arithmetic conversion rules (for arithmetic units of measure).

3.2.1.3.1 Keywording Rules - The rules for evaluating the strings associated with text (either in the update process against the source data or in the selection process against the arguments of the request) are known as keywording rules. They are defined in three parts:

3.2.1.3.1.1 Punctuation Rules - All characters available in the character set are defined as either:

- always part of a keyword candidate
- always a delimiter between words
- conditionally either of the above (e.g., under certain circumstances it MIGHT be desired to define the "comma" character as a conditional delimiter depending upon whether the preceding and/or following character(s) is/are numeric).

3.2.1.3.1.2 Depluralization Rules - Data Central has the ability (if desired) to attempt to depluralize the words found. This assures that both singular and plural form of keywords are selected independent of which is specified in the selection criteria.

3.2.1.3.1.3 Equivalences - Words, and word forms, can be made equivalent for the selection process. This requires that the equivalency be applied against the source data as well as against the selection specification arguments.

3.2.1.3.2 Arithmetic Rules - The data base designer, when specifying the design of the various files in the data base, has the ability to establish the rules under which arithmetic data will be processed. This includes the specification of both external (or source) forms accepted as well as internal forms desired. Possibly the most important feature in this regard is the automatic unit of measure conversion process. By creating a table of authorized units of measure and their relationship to an arbitrary internal "standard", the data base designer can trigger the system into automatically accounting for equivalency relationships without programming.

3.2.2 Groups - Groups are defined initially as any other segment name (see 3.2.1 above) and are followed by a list of real segment associations.

3.2.3 Entry type is implied by the file.

3.2.4 Definition of files is by definition of unique and/or synonymous tables of data (see section 2 and figure).

3.2.5 N/A

3.2.6 Security - Security level for a segment is defined by a number entered into a tabular position in the Segment Acronym Identification.

3.2.7 Data Validation - The initial program of the creation/update checks the "system validation" requirements discarding all segments whose character set includes invalid characters.

3.2.8 Data Base Definition Revision - Revision of the DBD is available in Data Central and varies in work load complexity from a simple redefinition (when merely another pseudo is defined) to a complex reloading of the data base (e.g., the keywording rules change the "word position" relationships). It must be noted that, under normal circumstances, data base definition revision does NOT require the actual revision of the physical files. This cost saving capability is explained in paragraph 2.2 above.

3.3 Data Central has complete internal control over storage structure.

3.4 Interrogation - This is the heart of the Data Central system (see Design Philosophy - paragraph 1.5.1 above).

3.4.1 Selection Criteria

3.4.1.1 Atomic Condition - terminology: There are four basic semantic components in the interactive specification of search criteria. These are:

- Segment name - this is the specification of the segment or group of segments (see item 2.2.0 on pseudo segments above) in which the condition specified is to be checked.

- Operator - this is the definition of the type of condition for which the check is made.

- Argument - this is the expression of the "argument(s)" to be checked against.

- Connectors - with Boolean connectors it is possible to combine atomic conditions.

3.4.1.1.1 Comparative conditions on item values - In Data Central there are two basic comparative conditions available in the search specifications: logical (or word) conditions and arithmetic conditions. Both require the specification (either explicit or implicit) of segment name, operator and argument.

3.4.1.1.1.1 Logical (textual) comparative conditions - when the operator associated with a "simple search specification" is a logical operator, the argument following the operator is considered to be a word (see item 3.4.1.1.3.3 on phrases below) in the instructional portion of the segment. The logical operators and their meanings are:

- EQUAL (semantically entered as "=" or "EQU") - With this operator, the component(s) following (until the existence of either a segment name or another operator) the operators are to exist in the specified segment.
For example:

$JOB-RESUME = MISSILE

means that the word "MISSILE" is to exist (be found) in the segment (field) known as "JOB-RESUME."

- NOT EQUAL (semantically entered as "/=" or "NEQ") - With this operator, the component(s) following the operator are not to exist (not to be found) in the specified segment.
For example:

$NAME NEQ RICHARD

means that the word "RICHARD" is not to exist (be found) in the segment known as "NAME." A special restriction exists upon the specification of this atomic condition - since it negates from a universe - it can only be used when connected to some positive specification that pre-defines a unique universe.

3.4.1.1.1.2 Arithmetic conditions - When the operator associated with the "simple search specification" is one of the list below, the

component(s) following the operator are considered to be arithmetic in form and, in general (except for dates which are handled separately), to be of the form n...n (decimal point) n...n a ... a where n represents a numeric volume and alpha represents a unit of measure. Up to three of these specifications are allowed.

For example:

$BROAD-JUMP-LENGTH > 1.0 YD, 1 FT, 8.1 IN

the operators associated with this type of process (which includes arithmetic unit of measure conversion and combining) are:

Arithmetic Equality (semantically "A=" or "AEQ")

Arithmetic Inequality (semantically "A/=" or "ANEQ")

Greater than (semantically ">" or "GTR")

Not Greater than (semantically "/>" or "NGT")

Less than (semantically "<" or "LSS")

Not less than (semantically "/<" or "NLS")

3.4.1.1.2 Non-existent data is handled in the form defined by "NEQ" in 3.4.1.1.1.1 above.

3.4.1.1.3 Other Atomic Conditions

3.4.1.1.3.1 Distance Search - In this specification two logical components can be linked with one of four special operator-connectors as follows:

- (Wn) - this is exemplified by:

$PROJECT-STATUS = BALLISTIC (W6) MISSILE

In this condition the two logical components must occur Within six words of each other.

- (WMn) - This is exemplified by:

$PROJECT-STATUS = BALLISTIC (WM6) MISSILE

In this condition the first component must appear the specified number of words in front of (Minus) the position of the second specified component.

- (WPn) - this is exemplified by:

$PROJECT-STATUS = MISSILE (WP6) BALLISTIC

This is the exact reverse of the aforementioned condition.

- (WPnMn) or (WMnPn) - This is exemplified by:

$PROJECT-STATUS = BALLISTIC (WP2M4) MISSILE

In this condition the length of each directional distance is separately specified (e.g., BALLISTIC must appear Within Plus 2 words or Minus 4 words of MISSILE).

3.4.1.1.3.2 Pseudo Segments - In the specification of what otherwise seems to be a "simple search criteria" (or atomic condition) with one segment name, Data Central allows for the implied complex specification of segment definitions. These pre-defined (and stored) definitions do NOT require a restructure of the physical file - Data Central accounts for the "REAL" structure internally. For the definition of the two types of pre-defined pseudo segments see Section 2.2 above.

3.4.1.1.3.3 Phrase Default - If two or more arguments are concatenated together in the atomic specification, the system assumes them to be a phrase specification, e.g.:

$SEGMENT-NAME = BALLISTIC MISSILE

3.4.1.1.3.4 Universal Character (and/or Root Condition) - The use of a universal character (asterisk *) in the argument allows for varients in spelling - e.g., SM*TH implies SMITH, SMYTH, etc. The use of multiple universal characters appended to a root word allows for root/stem expansion, e.g., TAX***** implies TAX, TAXABLE, TAXPAYER, TAXPAYEE, TAXATION, etc.

3.4.1.2 Item Conditions - The use of Boolean connectors (defined below) to connect arguments for the same segment name is allowed, e.g.

$DATE > JUN70 AND < JAN71

Additionally, Boolean connectors are allowed between arguments without the respecification of the operator, e.g.

$NAME = RICHARD OR ROBERT

3.4.1.3 Entry Selection Criterion - There are three basic ways in which total conditions for entry selection may be built up.

3.4.1.3.1 Boolean Connectors - The system allows for two levels of conjunctivity and one level disjunctivity. They are defined as follows:

A and B or C means A & (B or C)

A & B or C means (A & B) or C

3.4.1.3.2 Dialog Recursiveness - specifically "MODIFY" - Data Central is an interactive system, using the command "MODIFY," the user may, at any point in the process (e.g., retrieval, display, etc.) return to the specification of selection criteria (called search mode). In doing so he is not required to restate his already defined and pre-stored selection criteria; he merely adds the additional criteria either increasing the universe of selection by entering the criteria with an "OR" or he may decrease the universe of selection with an "AND." The recommended procedure is this latter form is to start with a simple criteria (which creates a large universe) and "NARROW" the universe with "AND" modifications.

3.4.1.3.3 Data Recursiveness -Although cross-file and cross-entry logic is an inherent capability of Data Central, the direct multiple atomic cross-file conditions are normally of the "OR" type as entries usually belong only to one file. Data Central has the ability to use data generated from existing entries in the generation

of either subsequent multiple atomic conditions in "AND" modifications of the existing multiple conditions or as "stand-alone" multiple atomic conditions. This process however, because of its complexity, requires the generation via any existing procedural language of the computation (or other processing) requirements necessary to generate the specific new set of conditions against which the search is to operate.

3.4.1.4 Since Data Central operates in an interactive mode and the user is NOT required to view ALL entries (see "ENTRYING" in Section 3.4.2.2.7.2 below), no capability for weighting of selection condition exists.

3.4.1.5 Each entry in the Data Central data base is defined by a unique Data Central assigned accession number. If this number is known, access to the data directly (bypassing the normal search criteria) can be made by issuing the condition:

$NR $\begin{Bmatrix}\text{EQU}\\ \text{BTN}\end{Bmatrix}$ X, X, X, ..., X (where X stands for the known accession numbers). If the operator is "EQU" then each specified accession number is set for retrieval. If the operator is 'BTN" then the numbers are expected in pairs and each pair represents a range of accession numbers to be retrieved.

3.4.2 Data Extraction - Reference is made to "File Output Definition" in the diagram associated with paragraph 2 above. In Data Central the output definition, in general, contains four parts:

- extraction or retrieval requirements
- sorting or sequencing requirements
- the subroutine defining the display format (generally output unit dependent)
- the device (terminal or printer) normally to receive the output

3.4.2.1 Extraction Features - The extraction criteria in Data Central are based upon the entries which satisfy the multiple atomic selection criteria above coupled with the list (pre-stored as part of the output definition) of segments to be retrieved. Since Data Central is an on-line interactive system designed to be used by a non-programmer,the pre-stored set of segments to be retrieved is handled by ADP personnel and it is activated by the use of an output acronym. In specifying the segments to be extracted certain special criteria can be specified:

3.4.2.1.0.1 "ASK.TERM" - In addition to the list of "always to be retrieved segments" a special segment name known as "ASK.TERM" can be specified. When this is specified, the terminal user has the option to specify additional segment (names) to be retrieved.

3.4.2.1.0.2 Pseudo-Segments - (see paragraph 2.2 above for definition of pseudo segments). Even though no data may exist in the file for pseudo segments, they may be specified for retrieval. All real segments associated with the definition of the pseudo are retrieved. Additionally, if the pseudo has been defined as a "concatenation" type, the system will change the segment identification for the retrieved data such that they may be viewed as a true concatenation (a new simple segment).

3.4.2.1.0.3 "IF HIT" - One possibility of establishing the selection criteria is via pseudos. Under this condition the user may only want retrieval of the segments in which the "hit" occurred. As a segment is defined (either in its original form or as it is specified for retrieval) it may be specified as an "IF HIT" segment and retrieval of segment occurs only if the segment satisfied any of the atomic conditions (above).

3.4.2.1.0.4 Personalized Abstract - Although most segments of a logical file are relatively fixed (short) in length, many segments may contain voluminous lengths of data. When segments are specified for retrieval, it

may be desired that all data NOT be retrieved. If this is specified, a personalized abstract is created based on the arguments used in the selection criteria. Only those words used in the selection specification and a pre-defined number of words surrounding it (them) is retrieved. This abstract of the actual data is personalized to the terms of the user's request.

3.4.2.1.1 Sorting - Part of the output definition called for (via acronym) is sorting information. Sorting may be specified as MANDATORY, OPTIONAL, or NOT-TO-BE-DONE based on the number of entries selected during the search process. If the sorting is mandatory or optional, the sorting parameters of segment, length, and mode (ascending or descending) are included from major to minor order. If the segment specified for a sequencing parameter is "ASK.TERM" (see paragraph 3.4.2.1.0.1), the terminal user may enter his own special set of parameters, again from major to minor order. The specification of sequencing parameters is independent of retrieval specifications. Although the system imposes no restrictions, certain implementations limit the amount of sequencing because of storage space available.

3.4.2.1.2 Item properties which are extractable - none.

3.4.2.1.3 Discrete extraction sets per selection criteria - known in Data Central as "Recursive Output." At any time in the browsing or display process the user may order the command "OUTPUT." With this command, the output definition (and/or device) may be changed. The user therefore may receive any number of "extraction sets" per selection criteria. This is an inherent capability of the interactive mode of operation.

3.4.2.2 Report Capability - The report capability in Data Central must be defined in two parts. The primary report capability lies in the ability of Data Central to interface with formatting subroutines written in any of the existing procedural languages. This capability and interface (known as the OIF - Output Interface Format) allows for ANY report function to be utilized. Additionally, Data Central makes available one generalized subroutine.

3.4.2.2.1 Content Lines - Data Central is designed to be operated by non-ADP personnel and report format specification is considered to be a programming function. Other procedural languages (COBOL, PL/1, RPG, etc.) are used.

3.4.2.2.2 Titles - Either a standard or a non-standard title is included in the OIF. The output definition defines which (see Section 3.4.2 above and 3.4.3.2 below).

3.4.2.2.3 Heading lines and footing lines - Pre-stored heading and footing lines are available (on call by the subroutine) from the Data Central system. This is handled via the pre-stored code conversion tables (see figure associated with paragraph 2 above).

3.4.2.2.4 Other user specified text - conversion of codes to textual data is available (on call by the subroutine) from the Data Central system.

3.4.2.2.5 Editing and Formatting - This is handled procedurally via one of the existing languages.

3.4.2.2.6 Derived Data - This is handled procedurally via one of the existing languages.

3.4.2.2.7 Other

3.4.2.2.7.1 Color - The system (if defined in the output definition) has the capability to insert, at appropriate points in data, codes to effect color display on the currently available color CRT devices.

3.4.2.2.7.2 ENTRY-ing - During the display or browse phase, the Data Central system allows the user to "skip" the remaining data in the report for one entry and move immediately to another entry's report. This is handled via the command:

$$\left\{\begin{matrix}+\\-\end{matrix}\right\}\left\{\begin{matrix}\emptyset\\1\\\cdot\\\cdot\\\cdot\\m\end{matrix}\right\}\ \text{ENTRY}$$

where "+" means move forward and "-" backward; the number indicates the number of entries to which movement is desired.

3.4.2.2.7.3 Abstract explosion (see 3.4.2.1.0.4 above) - A command exists which causes Data Central to expand or contract the size associated with the context of the keywords in the personalized abstract.

3.4.2.3 Extraction of files for use outside the system - Using existing procedural languages this is possible. For example, Data Central output has been interfaced to automatic typesetting equipment. There is no restriction to the capability for using Data Central stored information external to the system.

3.4.2.4 Extraction of files for system's use - This capability exists in the following forms:

3.4.2.4.1 Dialog Recursiveness - Through interactive dialog commands the following extracted files are available for the system's use:

3.4.2.4.1.1 "MODIFY" (see 3.4.1.3.2 above) - The file containing the user's entered request specification is extracted and used by the system to concatenate the semantic elements of the modification as the user enters them. This capability is commanded by the one-word command "MODIFY."

3.4.2.4.1.2 "OUTPUT" (see 3.4.2.1.3 above) - The file containing the list of selected entries (and their associated parameters) relating to the proper modification level are extracted and used by the system to change the form, mode or location of output desired. Commanded by the single word "OUTPUT".

3.4.2.4.1.3 "PRINT" - This recursive command also causes extraction of the file containing the list(s) of selected entries for use in redisplay. This is similar to "OUTPUT" described in the preceding paragraph except that the form, mode and location parameters are unchanged.

3.4.2.4.1.4 - "ENTRY-ing" - (see 3.4.2.2.7.2 above).

3.4.2.4.2 "OFF-LINE PRINT" - During the extraction phase, if the device specified is "PRINTER", the extraction criteria are applied to the files,creating a new file for subsequent physical formatting and outputting by the system in a batch mode.

3.4.2.4.3 DBD - Those portions of the DBD necessary for processing are extracted and used by the system. These files are built and maintained by the data base designer.

3.4.3 Many elements of the meaning associated with an interrogation are stored. Two examples are:

3.4.3.1 Equivalency (see Section 3.2.1.3 above) - These equivalence relationships are stored as part of either (or both) of the keywording rules and/or the arithmetic rules.

3.4.3.2 Output Definition - (see paragraph 3.4.2 and 3.4.2.2 above) - Not only the report format (identified by the name of the formatting subroutine desired) but all items of the extraction definition are stored as pre-defined parts of the interrogation process. Note that additive modification (if specified as allowable) is an inherent capability of the system.

3.4.4 - Other features of the interrogation process

3.4.4.1 - Special Retrieval Criteria - A capability exists to pre-define and store certain special retrieval criteria. This exists in two parts: actual character set translation parameters and the name of subroutines to do special output or retrieval editing while using a "standard" format.

3.4.4.2 Special Sort Criteria - Although the segment, length and mode for sorting may be specified directly, the system allows for the specification (pre-stored) of the name of a subroutine to create special sort criteria. This sort criteria subroutine can create ANY sort parameters desired from a simple extraction of embedded data to a complex calculation of two or more segments.

3.5 Update - The update subsystem in Data Central affects all of the processing to effect an update of entries of the data base. Updates of system (DBD) files are covered in paragraph 3.2.8. This processing for data base update includes the building of all indices, of the inverted file(s), etc. It does not include the update reformatting (and/or editing) of the source data. The Data Central system converts the external data via a procedural language written subroutine. This subroutine is responsible for converting the form and format of the external data into Data Central format and passing it on to the interface in the defined form. This passing is referred to as the External Input Interface Format (EIIF). The system converts to Internal Input Interface Format (IIIF) and completes the update process. In creating the Data Base Definition for a given file of the data base, the designer can specify not only the input formatting subroutine (that routine that converts from the actual format coming into the EIIF) but also the name associated with input editing criteria called SEC (Special Edit Criteria).

3.5.1 Selection Criteria - The EIIF specifies the selection (entry and segment) selected.

3.5.2 Update Specification - Old data entering the update subsystem must be identified as to which entry and which segment is being affected. Insertion or deletion of either the entry or segment level is allowed as long as the segment of the file is defined in the DBD. A sub-level subroutine makes the "current" source data available to the programmer for "intra-segment" effects. This is known as "GETSF".

3.5.3 "System" triggered updates are handled procedurally as a special input of the Input Interface Subroutine, written in ANY existing procedural language.

3.5.4 Input validation is handled in one of two methods, both of which are dependent upon the procedural language compiler(s) available. First, the editing can be done as part of the input formatting subroutine. As an alternative, Data Central could make available any "Special Edit Criteria (SEC)" subroutine.

3.5.5 Invocation of pre-defined update - Available via procedural routines written in their own language.

3.5.6 Audit trails - a series of listings (including, in the worst case, a list of all EIIF data) is available on call as part of the update subsystem.

3.6 File creation is handled as an update of a "NULL" file except that the DBD for the file specified must have been generated (see Section 3.2 above).

3.7 Global Functions

3.7.1 Arithmetic Computations

3.7.1.1 Unit of Measure - see paragraphs 3.2.1.3.2 and 3.4.1.1.2 above.

3.7.1.2 Output Computations - see paragraph 3.4.2.2 above.

3.7.2 Own Code - see paragraph 3.4.2.1, 3.4.2.2, and 3.5 above.

3.8 Other Functional Capabilities

The on-line tutorial or CAI ("WHAT") capability of Data Central has not been previously defined. As part of the DBD (see Section 2 above) the data base designer has the option of specifying either or both of:

3.8.1 The actual semantics of messages from the computer to the user.

3.8.2 A CAI type dissertation about both what the message means as well as the options for response. If desired, the capability may be presented to the user in multiple levels based upon the user's capability. All or none (or any intermediate level of tutorial) may be called upon by the user.

4. STORAGE STRUCTURE - The internal physical storage structure is not apparent to the user. He views the data base based on the logical definition(s) in the DBD (see paragraph 2.2 above). He therefore needs (and has) no control over the physical storage structure.

4.1 Item Level Storage Representation - All data in an item (segment) of an entry in the data base is considered to be made up of a string of text of variable lengths.

4.2 Entry and Group Level Storage Structure

4.2.1 Entry Level - A heading for the identification of the entry is provided ahead of the first occurring segment in the entry.

4.2.2 Group Level - Since the group is a pseudo-identification and the association of real segments, paragraph 4.1 applies to each real segment in a group.

4.3 File Level Storage Structure - While paragraph 4.1 and 4.2 applies to the displayable data, inverted data and indices exist in a proprietary form. The inversion is at the argument (word or value) level.

4.4 Multiple File Storage Structure

4.4.1 Data Central is a multiple physical file system. Two files and their indices are explained above. The third physical file (and its index) relate to the DBD. This file of data represents the manner of transformation from the user-specified logical concept to the physical data. As above, this physical file is NOT apparent to the user; it is built and maintained by the data base design (applications) engineer.

4.4.2 The ability to switch data bases is a capability definable in one section of the DBD. This allows the user to switch from one set of multiple files for one data base another set of multiple files for another data base.

5. OPERATIONAL ENVIRONMENT - Data Central operates with its own executive program (TTAM). With this executive Data Central appears, to the hardware/operating system complex, as a single never-ending job. TTAM itself makes identical copies of Data Central available to the terminal user; in this manner, and with roll-out/roll-in, TTAM time shares Data Central.

5.1 Hardware Parameters

5.1.1 Minimum Basic System - Data Central is operational on IBM 360-40 and up (including IBM 370 computers). The minimum core requirement is dependent upon the operation system and number of communication lines desired. It is possible to run Data Central in support of one terminal on a 65K machine (assuming a DOS supervisor). It requires two tape drives (for update), one disk (plus data base storage - dependent upon application data base size). Larger core and/or more disk space on larger systems effect greater efficiency of operation.

5.1.2 Storage Media - Data Central supports ALL IBM direct access devices including:

- 2311 disk
- 2314 disk
- 2321 data cell
- 3330 disk

5.1.3 Terminal Equipment - The TTAM executive supports, in a dial-up mode:

- Model 33/35 TTY
- IBM 1050
- IBM 2740
- IBM 2741
- IBM 2260
- CC-30 (Computer Communications, Inc. terminal in color or black-and-white)
- CC-70 (Computer Communications, Inc. front-end processor)

NOTE - A given installation MAY restrict usage to a sub-set of the above by virtue of the hardware "port" availability.

5.1.4 Hardware Transferability - The system operates on any IBM 360/370 computer (see 5.1.1 above). The internal design (and proprietary documentation) is in a machine-independent form. Data Corporation will entertain the requirement to make the system available for other manufacturer's equipment.

5.2 Operating System Parameters - While most operating modules of Data Central are operating system independent, the modules that affect the interface (primarily I/O) are operating system dependent.

5.2.1 Basic Required Operating System - Data Central operates under either DOS or OS. Operation in either mode is identical and therefore the operating system is transparent to the terminal user.

5.2.2 Significant Features - Since Data Central operates in any OS or DOS environment all capabilities are important; since it operates in an overlay structure in a roll-out/roll-in environment the use of the library is critical.

5.2.3 Transferability Between Operating Systems - Although all data files (data sets) of Data Central are transferable between OS and DOS, the programs themselves are not all transferable. Only certain modules (see 5.2 above) are transferable.

5.3 Restart and Recovery - The TTAM executive "traps" all errors and issues a "technical trouble" message to the terminal generating the error. That terminal user is immediately restarted at the sign-on stage. All other time-shared terminal users are unaware of any error as they are recovered and their processing continues uninterrupted.

5.4 System Operation Reports

5.4.1 Usage Reports - The system (via TTAM) records and logs the time-on and time-off parameters for each user in the system.

5.4.2 The system can (under control and only if desired) render message/response timing statistics.

(Data corporation)®

7500 old xenia pike, dayton, ohio 45432

DTN-68-2

copy to Turski

TECHNICAL NOTE

Chap. 8

INFORMATION PROCESSING AND THE DATA SPECTRUM

Richard H. Giering

Oct. 1967

Data corporation

INFORMATION PROCESSING AND THE DATA SPECTRUM

by

Richard H. Giering

October 1967

5.4 System Operation Reports

5.4.1 Usage Reports - The system (via TTAM) records and logs the time-on and time-off parameters for each user in the system.

5.4.2 The system can (under control and only if desired) render message/response timing statistics.

INFORMATION PROCESSING AND THE DATA SPECTRUM

Over the years, the state-of-the-art in information processing has gone from the edge-punched card to the 80-column card to the multiple-position single-record type file to the complex structured file "systems" we know of today. Each time that the discipline has advanced, it has done so in spite of the anguish of the "believers" of the preceding state. It is the author's opinion that another breakthrough is upon us!

This time the change in the discipline deals with the manner in which we look at the data spectrum. How many of us have actually sat down and looked at the TOTAL universe of data that we are processing or that our customers wish us to process? The cries of the believers can now be heard: "what is this guy talking about; I know what DATA is; what's he trying to sell?" The question is raised because, contrary to popular belief, we in the information processing game have not been considering the broad range of the data spectrum in the design of information processing systems in this, the third generation of the state-of-the-art. Without realizing it, each of us has been specializing in only one small portion of the spectrum- one type of data; it is now time to put away the second generation ideas concerning the "economic feasibility" of the whole of the data spectrum in the same manner that we had to put away the "80-column" ideas some time ago.

What, then, is this data spectrum? In order to give examples in the description of this data spectrum let us discuss it, using but two of the functional characteristics of information processing: data search and data retrieval. (Figure 1).

On one extreme of the spectrum one finds the meaningless (to a human) codes attached to some portion of the information universe. This would include classification codes and even account codes (numbers). All of us, at one time or another, have run into one of these coding schemes that either did not fit our unique situation or was so comprehensive as to require an expert with many years experience to use it. Witness some of the job aptitude codes.

As one proceeds down the spectrum, he finds next the "standard abbreviations" used in many systems. These codes (and, for all intents and purposes, they are codes) are at

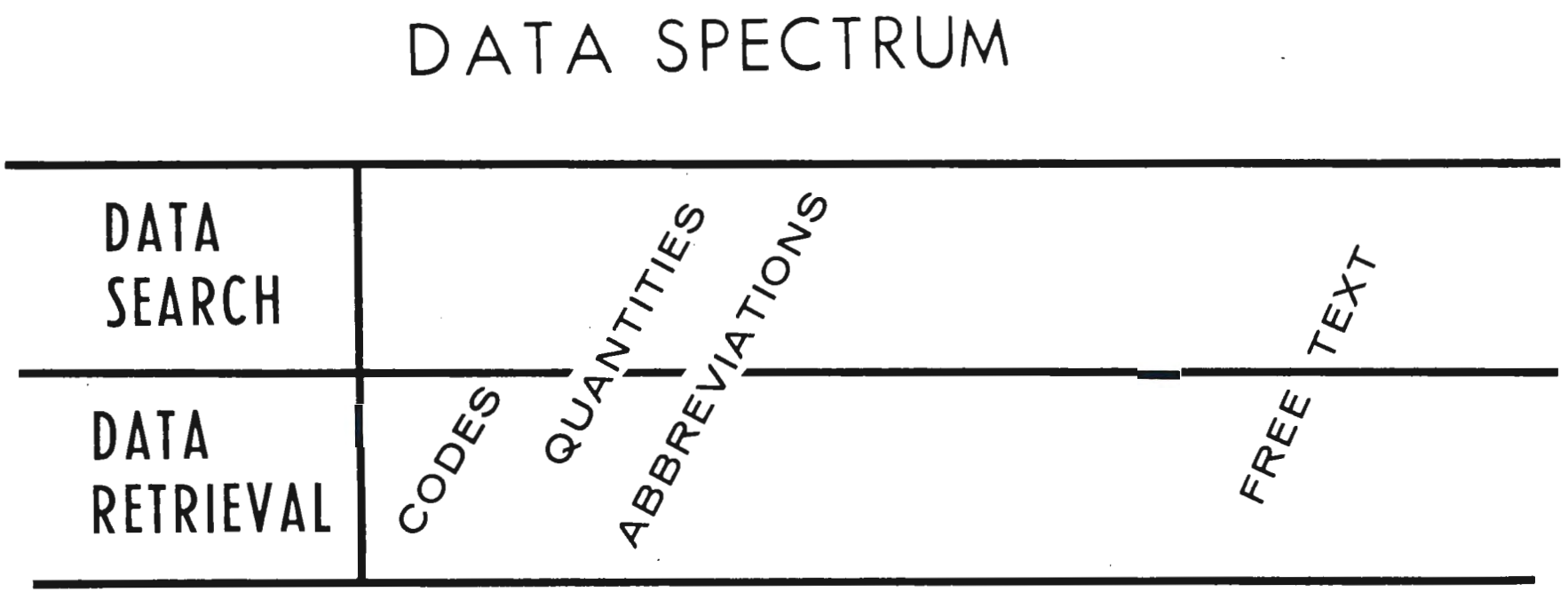


Figure 1. Contents of the Data Spectrum

least meaningful to the lay user. This portion of the spectrum includes the abbreviations for state, country, etc. These codes are less structured than are the previously defined codes; and, at least in the opinion of this author, they will remain a part of processing requirements as long as humans are a party to setting those requirements. Being less structured, however, their meanings are sometimes ambiguous when taken by themselves. They must be "read" in context.

On the other extreme of the data spectrum is the completely free text data. Here is where complete documents lie, including letters, memoranda, messages, etc.

For the sake of discussion let us categorize the spectrum into the two parts defined above. On one hand there is the structured data (both highly and lightly structured). This data is generally found in formatted fields. On the other hand one has the completely unstructured data (free text) and this is non-fielded.

Wait a moment!!! Does all data fit into one of these two broad categories? (Figure 2)

Actually, there is a third category. This third category has been largely ignored by the information processors to date. Into this category falls a mixture of the two previous categories; this is textual (unstructured) data, but it is in fielded form. One approaches the concept of processing this data with the "document processors" where there is a field for 'TITLE', another for 'AUTHOR', another for keyword descriptors, etc. In this example, search processing of this fielded data must be the same as the search processing of non-fielded text, e.g. keywording.

But what about the generalized file processors that have been touted for so long?

They are actually designed to process either the structured data in formatted files or (using the context and/or keywords) the completely unstructured data of a non-fielded file. We have too long ignored the center of the spectrum! What about the position title for individuals (whose data is maintained in your personnel formatted files); what about the actual description of a part (in your inventory control files); what about the name of the account (in your accounts receivable files)?

Considering the state-of-the-art of a few years ago, one designed with only one

DATA SPECTRUM

DATA SEARCH	STRUCTURED DATA	UNSTRUCTURED DATA	UNSTRUCTURED DATA
DATA RETRIEVAL	FORMATTED FIELDS	FORMATTED FIELDS	NON-FIELDED

Figure 2. Categories of the Data Spectrum

of the aforementioned extreme groups in mind (from a data search standpoint). For example, the systems designed to handle formatted files had (and, to a large degree, still have) no keyword search capability! On the other hand, text processors were (and, to a large degree, still are) incapable of processing structured, fielded data. Even the work currently going on in the third generation systems considers only one of the extremes.

Well then, what is the breakthrough? It is the consideration of the center of the Data Spectrum!

Upon study it becomes apparent the extremes of the spectrum are merely special cases of the center. In maintenance and retrieval everything is handled at the field level; that is, a complete field (or segment) is changed (in maintenance) or pulled and formatted on a page (in retrieval). Search, on the other hand, treats every non-common word (in this sense -- a code in a field is also considered a word) in all fields of the entry as a "keyword" or query handle. This approach differs from other systems in which only words from specified fields (e.g. Title only) or words deliberately and usually manually "coded" serve as keywords. As many of today's scientific disciplines are constantly in the throes of a breakthrough into a new technology, the above concept is of extreme importance. Prior to a breakthrough, a given set of words or phrases has one implied meaning (to the reader) while, after the breakthrough, the same set of words or phrases may have a completely different implication. The above concept then allows for the actual words, or phrases, throughout the complete record (report, document, or any other ADP record) to be used as query handle.

The approach also differs from "accepted practice" in formatted file processing where no keywords are used in search -- a match across the complete field is required to achieve a "good hit."

It was stated earlier that processors took into account "only one small portion of the spectrum." That is, formatted file processors do not allow for the use of keywords in search -- text processors do not consider the requirement for fielding (and obviously formatting). It has been said that "it is not economically feasible to process across the spectrum." The one notable exception to these previously accepted rules is (Data)

Central -- a generalized information system programmed and implemented by Data Corporation of Dayton and Washington.

(Data) Central processes across the data spectrum! And in an on-line conversational mode! How is this done? Doesn't keywording require a lot of main-frame time? The answer to the second question is "no." In order to answer the first question, it is necessary to get into a discussion of how (Data) Central works!

Figure 3 indicates how (Data) Central looks at implementations. One customer's application becomes only one of many applications of the system, and this is referred to as a data base. Each application is made up of many files, and each file is made up of many entries. During the second generation these entries were referred to as records because they were all logically stored as a single entity which, from a storage standpoint, is called a record. In order to differentiate between the storage notation of a record and the logical conglomeration of facts relating to a single subject, which previously was also called a record, (Data) Central refers to the conglomeration as either an entry, or a document, and uses the term "record" only to refer to storage problems. Each of these entries or documents is made up of a number of fields or, as we call them, segments. The terms are synonomous. You will note then that each field is made up of keywords. A keyword is defined as any set of characters between two delimiters as long as the word does not appear in an exclusion list. Now that means that highly structured or coded data goes into this category because obviously a code in a field defines the field as containing only one keyword, the code itself. Under certain unique combinations we could also process down to the character and bit level. Under normal circumstances, (Data) Central processes at the word, or sometimes called data value, level.

Now what is this processing? (Figure 4). When one has a serial file, the normal or standard approach has been the definition of fixed length fields, shown in the figure as fixed block length character strings. This means that if the field is not used up, it is just padded with a bunch of extraneous material. Our approach to the serial file is to have one variable length character string per entry, preceded by what might be called a Table of Contents. This reduces storage drastically and allows us to process any variable length data.

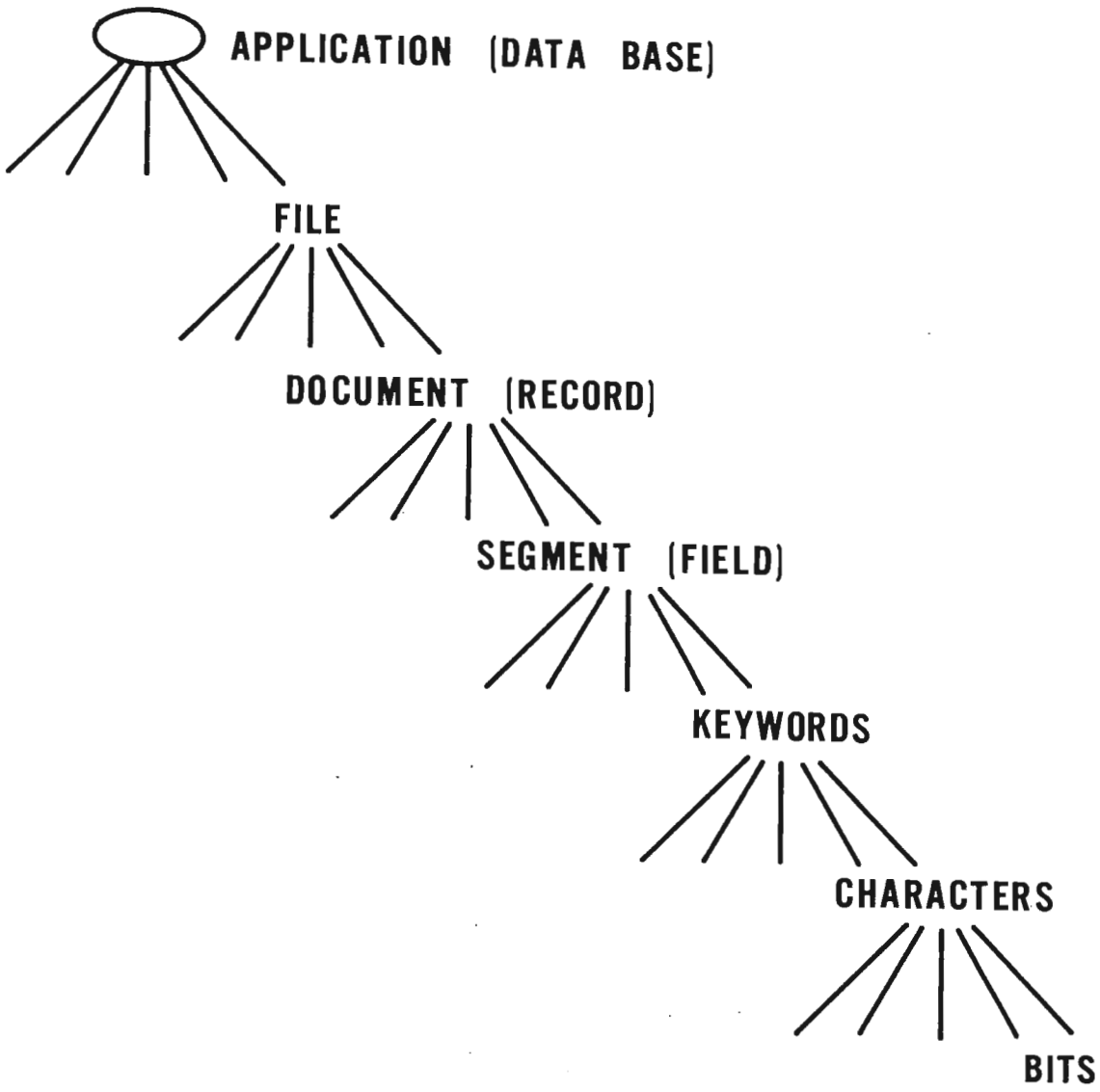


Figure 3. Data Central System Approach

STANDARD TECHNIQUE

DATA CORPORATION TECHNIQUE

1. Data Base
2. Document Segment
3. Character Count To Next Control Word

Figure 4. File Organization Serial File

However, serial processing does not allow for the fast acquisition of data. For this the system required an inverted file (Figure 5). There are two methods available for inverting a file: one is an inversion at the field level, and this is considered the standard method; the (Data) Central method inverts the file at the word level.

At the field level, inversion is merely the establishment of a separate index by each field value. That is, all values for the field "name" would be stored together, and attached to each of these values would be the accession number or entry identifier to identify which document contains this value in the field called "name." This still makes it difficult to do a keywording type of search on the field and allows generally for only the processing of structured data, where structured data now takes on the meaning of requiring a character for character match across the entire field.

The (Data) Central method, as shown in Figure 5 takes every word (except noise words) in the file and places it in the inverted file in alpha-numeric order, making it a keyword. Associated with each keyword -- and I might add at this point a reminder that highly structured, coded or formatted-file type data is also considered a keyword under this concept -- is a series of associated information strings which specifies not only in which entry the keyword appeared but in which field, in which file, in which sentence of the field, and what its relative position was within that sentence. All this allows the system to reconstruct, if necessary, complete phrases. Other information is carried for each occurrence of a keyword, including security classifications and information relating to maintenance. This maintenance information allows the system to know first when the data entered the file and also when it was last checked or updated; this affords a handle to manage the maintenance of the data base and also to determine Selective Dissemination responses.

Let's look now at an example of document entry (Figure 6). This is the upper half of a DD form 1498. It includes all types of data across the data spectrum; for example, the funding agency can be considered as coded or structured data; the contract amount as quantified data and such fields as the technology utilized or the title or the name of the organization, etc. as variable length, unstructured, yet fielded data. How would (Data) Central access a file of this type? Assume, for purposes of

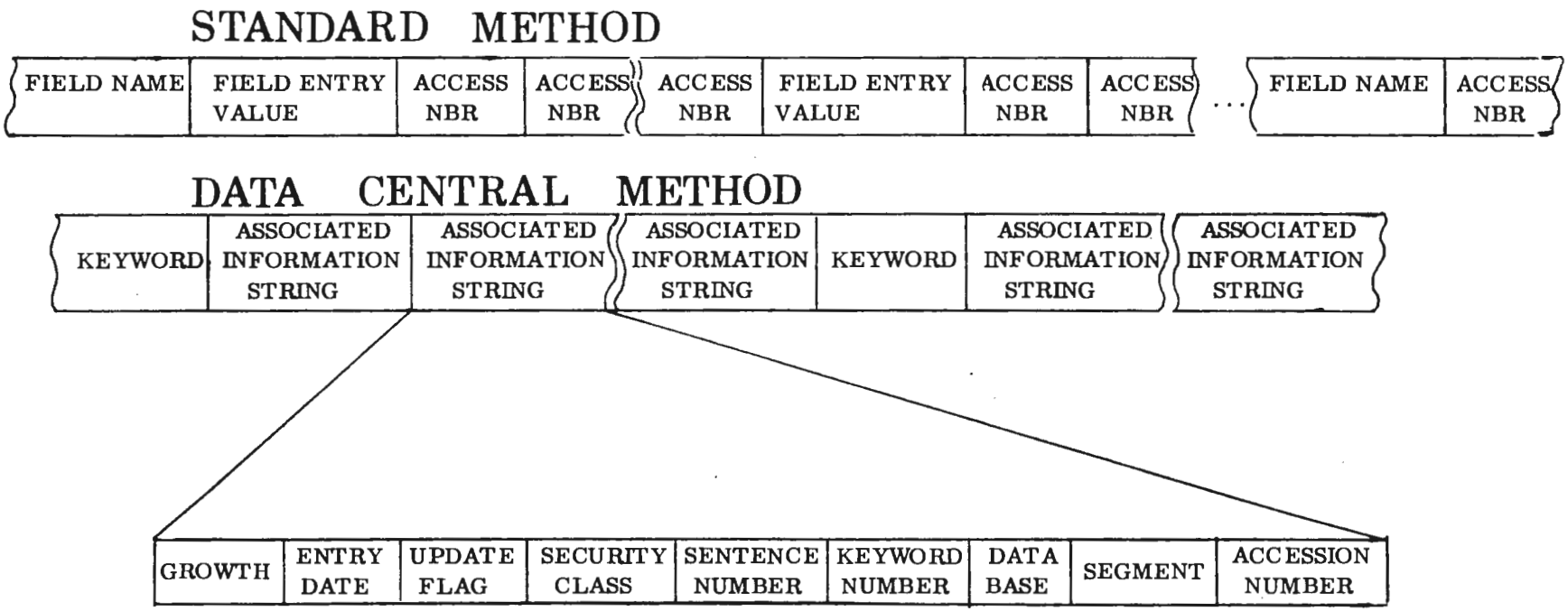


Figure 5. File Organization Inverted File

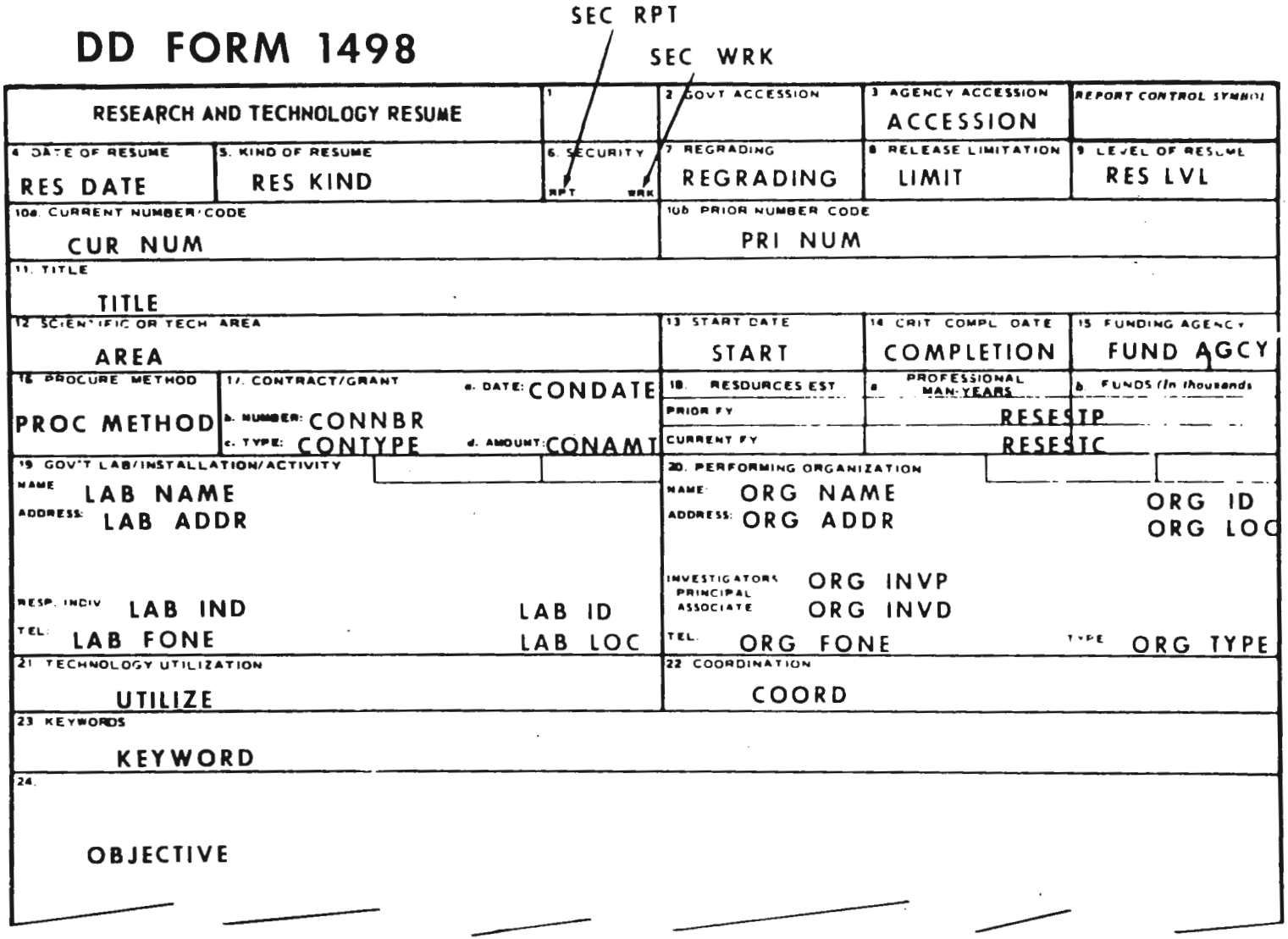


Figure 6. Example of Document Entry

discussion, that a file exists containing entries for 1498's. The file is defined as the "work unit" file and is assigned an acronym "WKUN."

The first thing a user sees when he punches into the system (Figure 7) is the request by the computer to "Enter Format, File, Output Device." In answer to this question the user enters a format number, in the example a format No. 5; the file acronym, in this case "WKUN"; and the word "Console."

Here we have an example of the minimal ADP training required to operate this system. We do not expect that the customer-users will have to program a special type of format output. If they want a special format, it is a job for the programmers. A programmer utilizing the report generator package or COBOL, etc. would generate the format package and insert it under a temporary format number. The user need not code any type of program but need only refer to a format number. This format NBR also is used to connect output to other sophisticated processors such as Fortran

```
ENTER FORMAT, FILE, OUTPUT DEVICE
5, wkun, console
ENTER REQUEST
$utilize equ 'mathematical model' and $title neq 'laser research'
THERE ARE 0002 DOCUMENTS THAT SATISFY YOUR REQUEST.
DO YOU WANT TO PRINT THE ANSWERS? - YES OR NO.
yes
DO YOU WANT THE DOCUMENTS SEQUENCED BY ONE OF THE RETRIEVED FIELDS?
YES OR NO.
yes
ENTER THE NAME OF THE FIELD TO BE SORTED
curnum
SET PAPER TO PRINT ON PAGE BREAK PRESS SPACE BAR TWICE AND CARRIAGE
RETURN
```

Figure 7. Computer/Console Dialog

statistical packages, etc. Later in this report there are two special formats defined that are universally used. This is an example of the differentiation (as far as the (Data) Central system is concerned) between programmers and customers. In the example the user also indicates that once he has processed a query against the file called WKUN, he wants the answer back on his console.

Since format 5, by its design, requires certain fields of data to be retrieved, the system need not ask for this information. As a result, the system comes back immediately with the question, "Enter Request." In the case at hand a simple request is generated: the user wants all documents or entries for which the field contains the phrase -- maybe alone or with others -- 'mathematical model', and the title of project field does not contain the phrase 'laser research.' Referring back to the inverted file concept, it is not necessary for the system to serially search every entry in the file. It is only necessary to go to that portion of the inverted file that relates to the words specified and determine which entries had the words used in the proper field and in the proper order. If it was used in another field, it would not fit the requirement. From the time the request is transmitted to the time the computer prints out the statement that "There are 002 Documents That Satisfy the Request." is somewhere on the order of five seconds, regardless of the number of entries in the data base. An additional four or five seconds of computer time is necessary for the remaining portion of the dialogue shown in the figure which also allows the computer to actually generate the report; the last part of the report is shown in Fig. 8.

Figure 9 is another example of a query. In the Recon Central data base or application there are three files: one dealing with 1498's called WKUN, one dealing with project abstracts called RCCN, and one dealing with photographic equipment data called PHBK for Photo Bank. Let us assume that an integrated query is necessary and that Figure 9 is that integrated query. Assume also that the field name "date" refers to the date in the 1498's, the "author" refers to the abstract author, and the "source" to that field in the photo equipment bank of data. In addition, the retrieval field "title" will also refer to the 1498's. Prior to entering the request a pseudo-file called INFO is defined containing this combination of fields, all currently existing in

TECHNICAL PROGRAM LISTING

RC 018208

N/A CURRENT NUMBER/CODE: 6.25.09.01.R 5955 AO377 CD000

TITLE:
(U) AIRBORNE LASER

PROFESSIONAL MANYEARS		FUNDS	
PRIOR FY 1965	0.00	PRIOR FY	$120000.
CURRENT FY 1966	0.00	CURRENT FY	$0.
		CURRENT FY+1	N/A

CUMULATIVE TOTALS

PROFESSIONAL MANYEARS		FUNDS	
PRIOR FY	.10	PRIOR FY	$211,000.00
CURRENT FY	2.60	CURRENT FY	$ 66,000.00

Figure 8. Example of Format 5

NOTE

The data contained in this figure, as well as elsewhere in this document, are fictitious and are not related to actual data appearing on official 1498 forms.

the other files. Currently this pseudo-file definition must be done at the computer site. We are currently working on placing this capability on-line. Once the pseudo-file INFO has been generated, this query can be run. Again, the first question to the user is "Enter Format, File, Output Device." In the case at hand this user is asking for format 99, and this is one of the special universal formats mentioned above.

Format 99 essentially tells the computer that the user does not care about the actual format per se and, in fact, may not be aware of any format that adequately satisfies his immediate requirements. He just wants a series of fields displayed in order, without regard to format. As a result, the next thing the computer asks for are the fields wanted or as shown on Fig. 9 "Enter Types to Be Displayed." The answers

ENTER FORMAT, FILE, OUTPUT DEVICE

99, INFO, PRINTER

ENTER TYPES TO BE DISPLAYED IN ANSWERS

TITLE, SOURCE, TEXT, AUTHOR, DATE

ENTER REQUEST

$SOURCE EQU 'TEXAS INSTRUMENTS' AND $AUTHOR EQU JONES OR SMITH AND $DATE GTR 67/01/01 AND LSS 67/03/31

THERE ARE 0012 DOCUMENTS THAT SATISFY YOUR REQUEST. DO YOU WANT TO PRINT THESE ANSWERS – YES OR NO?

YES

→ DO YOU WANT THE DOCUMENTS SORTED BY ONE OF THE RETRIEVED FIELDS – YES OR NO?

YES

ENTER THE NAME OF THE FIELD TO BE SORTED.

DATE

Figure 9. Computer Dialogue

that the user has asked for are the title, source, text, author, and date, the complete set of fields for the pseudo-file called INFO. The computer is now ready to entertain the request. Please note in the request the uses of the Boolian operators as well as of the arithmetic operators. Again, the logic employed in solving this problem does not necessitate a serial search of all entries or even of all fields specified (as in other inverted file concepts); but rather the portion of the inverted file dealing with the word "Texas" is immediately accessed, and all uses of the word "Texas" in the field name "source" are held as possible solutions, then all instances of the occurrence of the word "instruments" in the same field in word position one greater than the word position of the word "Texas" are noted, which in fact reconstitutes the phrase "Texas Instruments," reducing the number of documents or entries that satisfy the request. The same thing is done for the words "Jones" or "Smith" in the field named "author."

Finally, a sub-portion of the inverted file is used for the arithmetic data search. One of the unique capabilities of (Data) Central is the unit of measure conversion. Every field in the system is logical in nature. In addition, certain fields are defined as being also arithmetic; and a certain standard unit of measure is established for all arithmetic fields in the file. Data, upon entry, is converted to this standard unit of measure and entered in numeric sequence in the arithmetic portion of the inverted file. The user's arithmetic request parameters are also converted to this same standard unit of measure; and this converted figure, rather than the quantity entered, is used for the search. This means that one user might have entered into a focal length field the value of six inches and at query time asked for focal length greater than 25 millimeters. The quantity six is less than the quantity twenty-five, but the value six inches is greater than the value twenty-five millimeters. Having converted to a standard unit of measure internally, the system is able to accurately satisfy the request. The answer that would be received in this case would be six inches because in the serial portion of the file, which is used for retrieval, the data remains in its original -- not its converted -- form. The user never sees the converted form.

It might be added, also, with respect to the dialogue in the figure, that this user knew in advance that there would be more than just a few answers. Note that there are really twelve. As a result, he specified that he wanted the answers printed on the high-

speed printer in the computer room; and, in fact, that's what happens. The capability to specify "tape", using this same principle, would allow for data exchange, in machine readable form.

What additional features are available in working with data in this third generation system? One of the capabilities is that there is an internal synonym dictionary; and, in fact, when the term "Aircraft" is used as a query parameter, the phrase "flying machine" and "airplane" also become search parameters. In addition, there is a universal character which means that the words "Smith" and "Smyth" can be searched for by specifying SM*TH, where the asterisk indicates to the system that any character in this position is legal. Thirdly, since the system obviously can reconstitute phrases by virtue of word position, this system also has, inherent to its design, a distance searching capability; that is, it is possible to search for the occurrence of the word "ballistic" and the word "missile", requiring that they be within some number of words of each other. Other capabilities to include root words, geographic searching, etc. are planned but not implemented at this point in time.

Data (Central) processes across the data spectrum! And in an on-line conversational mode! The fact that it is operational and that it has been successfully (and profitably) implemented belies the statement concerning "economic feasibility;" the fact that the data is not only available but is currently in existing files indicates that there is, indeed, a "requirement."

This then is the breakthrough. The concept of processing by context rather than by code is upon us; we can structure any and all data bases the way the customer thinks; we do not have to force man to think the way the computer does. In addition, new data bases are not needed; the new concept allows for the processing of existing automated files and, in fact, allows for the searching of fields in these existing files that were NOT previously searchable. Files can be either truncated in nature (like abstracts) or, at the option of the user, in a voluminous form.

For years, we in the ADP profession have been talking about this cybernetic society of ours. We like to think that we have ascribed to the automaton many anthropomorphic qualities. In fact, because of the many involved coding schemes it is

actually the opposite; we have ascribed to the human user characteristics normally associated with the computer. In the past, this has been because the state-of-the-art would not allow us to do otherwise; this is no longer true and we, as true professionals, can return to the original path -- LETS MAKE THE MACHINE AN EXTENSION OF MAN, NOT THE REVERSE.

Information Systems Division

3481 Dayton–Xenia Road, Dayton, Ohio 45432

Phone: DAYTON 513 / 426–3111 WASHINGTON 703 / 979–1000

DTN-67-9

TECHNICAL NOTE

ANALYSIS
OF
EXISTING AND PROPOSED DATA HANDLING SYSTEMS

by

Richard H. Giering

Eastern Division
Data Corporation
1254 Jefferson-Davis Hwy.
Arlington, Virginia

23 October 1967

INTRODUCTION

Since the information contained herein has been gathered from various sources, it is incumbent upon the reviewer to delineate his sources as an indication of the validity of the data. Additionally, Data Corporation (the reviewer) by its involvement in one of the reviewed systems, might be considered prejudiced in evaluating its own system. This possibility was recognized throughout the investigation and an attempt was made to remove any bias.

0.1 Sources

0.1.1 Information on the various IDHS systems was gathered through the following documents and additional data (to clarify questions) was recieved in direct verbal communication with members of the Systems Operations Support Branch at DIA. The documents are:

0.1.1.1 The DIAM 65-9 series titled "Intelligence Data Handling System (IDHS) 1410 Formatted File System (1410 FFS)" with various dates, published by the DIA. (This document covers Mark II).

0.1.1.2 Mark III 1410 Formatted File System, Preliminary User's Reference Manual dated 10 February 1967 by DIA.

0.1.1.3 A series of three manuals dealing with 7094 FFS revised during June 1966 and published by DIA. These manuals define Mod 7.5 7094 FFS.

0.1.1.4 IBM Document No. N-0661 titled "Mod 8 7094 FFS Program Capability Changes dated 5 May 1967 and published by IDHS, Contracting Agency, Rome Air Development Command, Griffiss Air Force Base, New York. (This document defines the changes to Mod 7.5 FFS to make it Mod 8.)

0.1.2 Information relative to the NIPS Systems were gathered primarily through discussions with individuals currently on the programming staff for the NIPS development effort. Clarification of details, however, was gathered from system planning manual SPM 1-67 dated 31 January 1967, relating to the program design approach for S 360/50 FFS as published by the NMCS.

1.0 Minimum Hardware Configuration

1.1 Central Processor Type

1.1.1 Mark II - FFS is operational on a 1410.

1.1.2 Mark III - FFS will be operational on a 1410.

1.1.3 NIPS 1410 is operation on a 1410.

1.1.4 NIPS Phase I will be operation on a 360/50. It has been reported that the system will be able to operate once it becomes operational on 360/40. This has not been verified in any documentation.

1.1.5 NIPS Phase II - The comments applicable to Phase I are also applicable to Phase II.

1.1.6 Mod 8 requires a 7094 Model 2.

1.1.7 GIS is to operate on a 360/40 or 360/50 although the information concerning the availability on a 360/40 is questionable.

1.1.8 DPS when it is released in November will be operational on a 360/40.

1.1.9 ATS will require a minimum of a 360/40 and possibly by implementation time (January 1968) will require a 360/50.

1.1.10 Informatics Mark IV is to be operational on a 360/30 and up.

1.1.11 (Data) Central is currently operational for a 360/30 and up.

1.1.12 TDMS is written to be operational on a 360/50 or 360/65.

1.2 Central Memory (Size)

1.2.1 Mark II - FFS requires 80K characters of minimum storage.

1.3.5 NIPS Phase II requires no tape except that if that portion of the output package defining tape output is used then one drive is needed for this feature.

1.3.6 Mod 8 requires a minimum of twelve tape drives (17 are recommended) and these tape drives must be on three channels.

1.3.7 Information has not been released.

1.3.8 DPS requires no tapes, however, the documentation indicates that disk files are the normal mode of operation.

1.3.9 This information has not been released with respect to the implementation of ATS.

1.3.10 Informatics Mark IV per se does not require any specified number of tape drives as a minimum; however, since Mark IV is primarily a tape oriented system and for this the data is on tape as well as the output, normal operation will require a minimum of one for the old master, one for the new master, one for any sub-file, and one for output possibly.

1.3.11 (Data) Central per se requires no tape drives; however, it can use them for I/O if they are available.

1.3.12 TDMS also requires no tapes; however, the documentation indicates that disk files are the normal mode of operation.

1.4 Secondary (On-Line, Mass) Storage

1.4.1 Mark II - FFS requires a minimum of one 1301 disk for operation.

1.4.2 Mark III - FFS requires a minimum of one 1301 disk for operation.

1.4.3 NIPS 1410 requires a minimum of one 1301 disk for operation.

1.4.4 NIPS Phase I requires a minimum of four 2311 disk packs.

1.4.5 NIPS Phase II requires a minimum of four 2311 disk packs.

1.5.3 CRT - With the exception of (Data) Central and TDMS, no system requires nor can use CRT's. Both (Data) Central and TDMS are currently programming for the inclusion of CRT capability (estimated 1 January 1968 for (Data) Central and June 1968 for TDMS. There has been some discussion of the use of CRT's with NIPS; however, no plans have as yet been consummated to this reviewer's knowledge.

3.0 Sub-Operating System

3.1 Recovery Procedures — All checked systems except (Data) Central, TDMS and ATS will have no direct recovery procedures but will by-pass to the next job in the job stream (this includes the batch and the remote terminal queue). (Data) Central has extensive recovery as part of its conversational mode operation to allow recovery to various levels of re-operation. TDMS reports that it will also have an extensive user recovery capability.

3.2 Operations Recording — Information relative to transaction recordings for GIS, DPS, and ATS is not available. All three IDHS (Mark II, Mark III and SAC Mod 8) have relatively extensive recordings procedures for printout only. These recording procedures cannot be considered as audit trails; however, because they are not maintained. (Data) Central has no direct operations recording and Mark IV reports audit trails of some type that can be considered as a form of operation recording. TDMS saves if requested a transaction tape recording all operations. Documentation as to the level of recording is meager.

4.3.4 NIPS Phase I has the capacity to handle any number of 200 character fixed length fields per logical entry. A limit of approximately (available data is sketchy here) 250 fields are allowed for the fixed set data plus approximately the same number for each logical segment where a logical segment can be a group of one or more periodic sets. In addition one variable length text field may be assigned per segment.

4.3.5 NIPS Phase II — Same as above.

4.3.6 Mod 8 allows for 200 fixed length fields of 126 characters each (except that arithmetic data is limited to 36 bits) and allows for up to 5820 characters per logical entry.

4.3.7 A maximum of 750 field-names are usable across three files; e.g. 250 for each of 3; 375 for each of 2, or 1 file of 750 names. Each field is limited to 256 characters in length.

4.3.8 DPS may have up to 255 fields of data assigned per logical entry. The length of all but one field is restricted to 249 characters except that the total length for these fixed fields must be less than 1638 characters. In addition to these fixed fields, DPS allows for processing of free text (keywording) and phrasing against one variable length field.

4.3.9 There are no fields allowed for ATS and the complete document entered into the system is considered as one huge variable length field.

4.3.10 Informatics Mark IV reports no limit on the number of fixed length fields and the length of the logical record is limited only by core availability.

4.3.11 (Data) Central allows for 256 times 64 fields, all either variable or fixed in length with no restrictions placed on the number of characters per logical entry.

4.3.12 TDMS reports "some large number" of fixed length fields each a maximum of 256 characters per logical entry. The logical entry size is not restricted.

4.4.10 See paragraph 4.3.10.

4.4.11 (Data) Central places no restriction on the number of characters per logical entry.

4.4.12 TDMS also reports no restriction on the number of characters per logical entry.

4.5 Maximum Number of Simultaneous Files

4.5.1 Mark II – FFS, all three versions of the NIPS, Document Processing System, Administrative Terminal System and TDMS all report no multi-file capacity.

4.5.2 Both Mark III – FFS and Mod 8 can process a multiplicity of files in one batch query; however, the actual process is one file at a time where data is held (and/or merged) between single file processings.

4.5.3 Informatics Mark IV reports the capability for simultaneous processing of up to five files.

4.5.4 (Data) Central processes simultaneously up to 64 files.

4.6 Keywording

4.6.1 With the exception of the Administrative Terminal System, all systems are capable of processing user supplied keywords in that the individual words (and/or phrases) are placed in separate fields for query purposes. The ATS system, because of its primary design for the maintenance of documents rather than the selective retrieval thereof, does not allow for this capacity.

4.6.2 Only the Document Processing System and (Data) Central report any current capability for processing system generated keywords. In both cases the textual material is broken down into keywords and separately stored for query. In addition, TDMS, in the proceedings of the 22nd Conference of the ACM, reports the planned inclusion of keyword ability. It is anticipated that this will not be available prior to the end of 1968.

4.11.5 NIPS Phase II allows 1250 periodics, all at the same level.

4.11.6 Mod 8 allows for a single level of nine periodics.

4.11.7 The information for GIS is not available. Discussion, however, reveals that multiple level periodicity (to about 15 levels) will be available; the total number of sets definable is not available.

4.11.8 DPS does not allow for any periodics and none is planned.

4.11.9 ATS allows for no fields of any kind.

4.11.10 Mark IV is designed to handle 99 periodics in any combination of up to nine hierarchical levels.

4.11.11 (Data) Central does not allow for any periodicity; however, system modification currently being considered will allow for 256 periodics for combinations of up to 15 levels of hierarchy.

4.11.12 Information has not been released.

5.4 Internal File Restructuring — NIPS 360 Phase I, DPS and ATS do not have any capacity for restructuring the data base. All other systems report that this capacity is inherent to the design.

5.5 User Supplied Standard Updates (Stored) — DPS and ATS do not have this capacity. (Data) Central is planning this for inclusion sometime next year. All other systems report that it is either operational for the second generation systems or will be operational when the system, itself, becomes operational (for third generation systems).

5.6 Audit Trails — TDMS and Mark IV are the only two systems that generate any type of audit trail, although a similitude of an audit trail is generated by all of the IDHS FFS's in their operational recording. No other audit trails are available.

6.2.3.10 Mark IV can process up to five files simultaneously.

6.2.3.11 (Data) Central can process 64 files simultaneously.

6.2.3.12 TDMS has no multi-file capability.

6.2.4 File Associated Tables

6.2.4.1 NIPS 1410, 360 Phase I, 360 Phase II and ATS do not use file associated tables.

6.2.4.2 All other systems can or do use file associated tables.

6.2.5 Operators Query Language

6.2.5.1 Logical Operators (AND, OR, NOT) — Only ATS which is not a querying system can not use all three operators directly.

6.2.5.2 Arithmetic Operators (LE, EQ, GT, etc.) — ATS is the only system that does not allow for arithmetic operation. All other systems allow for arithmetic searching. Both (Data) Central and TDMS however allow for automatic unit of measure conversion in arithmetic searching.

6.2.5.3 Between Operators — The Three NIPS systems and TDMS have the between operator for direct use against any arithmetic field. (Data) Central allows the use of the between operator (ALL) only for the record ID field. No other system uses between as a direct operator.

6.2.5.4 Geographic Searching — Only the military systems (all IDHS and NIPS systems) currently have the geographic search capability. (Data) Central is currently planning this inclusion for sometime immediately after the first of the year. No other system has this capability.

6.2.6 Language Features

6.2.6.1 Keyword-Key Phrase Capability — (Data) Central allows for full keywording and phrasing on all fields of the file. DPS allows for keywording and key phrasing on only the one variable length field defined. None

6.1.2.11 (Data) Central is currently planning to include, under its time sharing operation, a queued mode.

6.1.2.12 TDMS reports that it is primarily a conversational mode system and as such except for the necessity of a time sharing queue does not operate in queued mode (this means no batch capability).

6.2 Language and Capabilities — All systems with the exception of ATS have query language.

6.2.1 Stored Query

6.2.1.1 Mark II - FFS, Mark III - FFS, DPS, and ATS have no stored query capability.

6.2.1.2 (Data) Central does not currently have the capacity to store queries; however, this is being planned at this time.

6.2.1.3 All other systems have the capability to store standard user queries.

6.2.2 Search Method

6.2.2.1 Mark II - FFS is generally a serial search system. At file set up time, however, two fields of the file may be specified for indexing and the system will generate an index based upon the contents of these fields and for these fields then the system is index sequential.

6.2.2.2 Mark II - FFS is the same as the above.

6.2.2.3 NIPS 1410 is only a serial search system.

6.2.2.4 NIPS Phase I is serial except that it is index sequential for the record ID field.

6.2.2.5 NIPS Phase II is the same as the above.

6.2.2.6 Mod 8 is a serial search system.

6.2.2.7 GIS uses an index sequential system.

does exist indirectly by defining one or more security fields externally in the file and outputting based on the contents thereof.

6.2.6.8 Field Named Synonymy — NIPS 1410, NIPS Phases I and II, Mark IV and the TDMS allow for the use of field name synonymy. Other systems do not have this capability.

records until a record is found that meets the condition. It has been reported that Informatics Mark IV may have this capability. None of the other systems have either the conditional logic or the randomizing capability and in fact, systems that are not either index sequential or inverted search could not use the randomizing function.

8.1.5.1 Mark II, Mark III and the three NIPS operations allow for user controlled volume by virtue of an automatically generated queriable page number. Volume is therefore controllable in page groups.

8.1.5.2 Under Mod 8, DPS, ATS and Mark IV, there is no user controlled volume capability.

8.1.5.3 In both (Data) Central and TDMS the conversationality of operation gives a complete user controlled volume at the record and/or field level.

8.2 Sort

8.2.1 Multi-File — Only Mark IV and (Data) Central have a multi-file sort capability.

8.2.2 Multi-Key

8.2.2.2 The three 1410 systems (Mark II, Mark III, and NIPS) allow for any combination of up to 25 characters in the sort key.

8.2.2.3 Mod 8 allows for specified seven different fields, one major and six minor, as the sort key except that for any field whose length is specified as being greater than 30 characters only the first 30 characters are used for the sort.

8.2.2.4 Mark IV allows for nine fields to be specified as a sort key and it is believed there is a character count restriction, although no documentation has been found to substantiate this impression.

8.2.2.5 Currently (Data) Central only allows for one sort key. A multiple key capability of unknown amount is currently being planned.

8.3 Generations

8.3.1 Sub-File — DPS and ATS cannot generate any sub-file. All other studied systems seem to have some capability in this respect.

8.3.2 Summary File — (Data) Central by virtue of its conversa-

8.4.2 Output Arithmetic Operators — The above comments apply here also.

8.4.3 Output Statistic Operators

8.4.3.1 If TDMS uses the aforementioned specifications for its output package, the statistical capabilities which include regression analysis, etc., will be vastly superior to any other RPG.

8.4.3.2 Both Mod 8 and Mark IV have the capability of specifying average only, taken over arithmetic fields.

8.4.3.3 (Data) Central using the operating systems RPG has a full range of statistical direct operators including average and deviation but no statistical analysis capability such as regression.

8.4.3.4 The remaining systems have no reported statistical capability.

8.4.4 Output Summary (Totals, Counts, Sub-Totals) etc.

8.4.4.1 Mark I, Mark III, the three NIPS all have total and count capability but no "sub" capabilities.

8.4.4.2 Mod 8 has the capability to summarize to two levels (e.g. counts and totals, sub-counts and sub-totals).

8.4.4.3 Mark IV has the capability for taking nine levels of totaling.

8.4.4.4 Both (Data) Central and TDMS have an unspecified number of levels for summarization capability.

8.4.4.5 DPS and ATS have no capability in this respect.

8.4.5 Sorting

8.4.5.1 All remaining systems have a measure of the capacity to specify the order of the output except that in both Mark II, and III and NIPS 1410 this capability is inherent to the query package rather than the output package.

NOTES:

1. Information is not released.

2. "E" storage for DOS; "F" storage required for OS.

3. "G" storage for non-remote operations; "H" storage required when remote processing is available.

4. None for system per se except that if files take up all (sic) available disk space, then tapes are needed for sorting.

5. TDMS requires per se neither tape or DASD, Purchaser must have something, however, for both system and data base residency.

6. Resides as a separate system on the IBSYS system tape.

7. One variable length field only allowed per entry.

8. One variable length field allowed per segment.

9. Limited by core availability.

10. System is primarily a tape system — in addition a disk (2311) is needed if running under DOS.

11. Limited by DASD availability.

12. Via format number.

13. The only type allowed.

14. Position formatted files only.

15. Position and comma formatted only.

16. By writing a special (assembly or compiler) language program.

17. Index sequential on Record ID field only - otherwise serial.

18. Two fields can be specified at file set-up time for indexing and the system will allow for index sequential on these two fields.

19. Serial through a series of inverted lists — quite complicated set of lists.

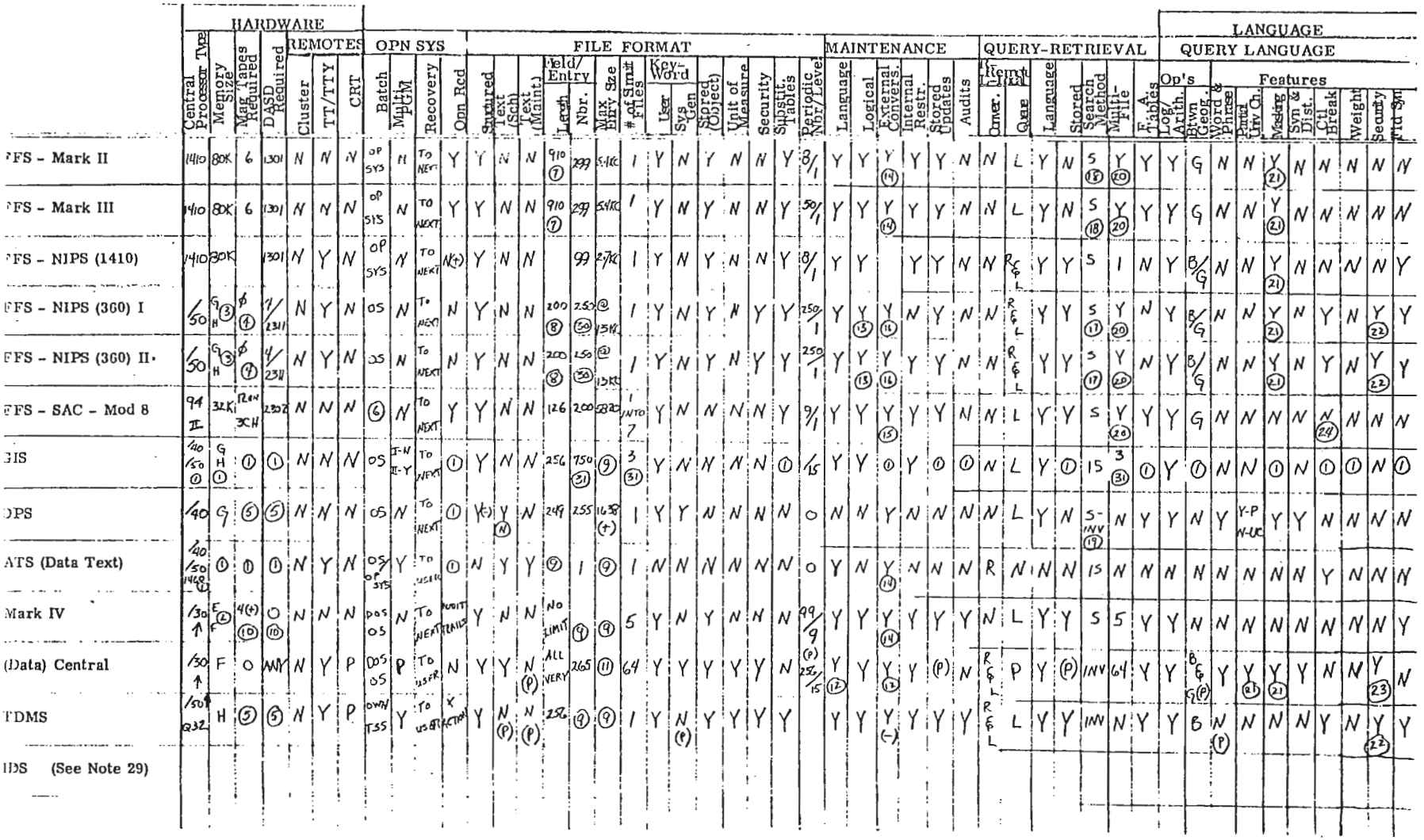
20. Merge only — one file at a time into another.

21. By use of the universal character.

22. At the complete file level.

23. At the sub-field (word) level.

24. Essentially a control break search does exist except that it occurs in the "output" package rather than the query package.


HARDWARE
REMOTES
OPN SYS
FILE FORMAT
MAINTENANCE
QUERY-RETRIEVAL
LANGUAGE
QUERY LANGUAGE
Features
Central Processor Type
Memory Size
Mag Tapes Required
DASD Required
Cluster
TT/TTY
CRT
Batch
Multi PGM
Recovery
Opn Rcd
Structured Text (Sch)
Text (Maint.)
Field/ Entry
Length
Nbr.
Max Entry Size
# of Smrt Files
Key Word
User
Sys Gen
Stored (Object)
Unit of Measure
Security
Substit. Tables
Periodic Nbr/Level
Language
Logical
External Convers.
Internal Restr.
Stored Updates
Audits
Temp Dat
Quer
Language
Stored
Search Method
Multi-File
Tables
Op's
Log/ Arith.
Btwn Geog.
Word & Phrase
Print Univ Ch.
Masking
Syn & Dist.
Ctl Break
Weight
Security
Std Sm
PFS - Mark II
PFS - Mark III
PFS - NIPS (1410)
PFS - NIPS (360) I
PFS - NIPS (360) II
PFS - SAC - Mod 8
GIS
OPS
ATS (Data Text)
Mark IV
(Data) Central
TDMS
IDS (See Note 29)

Journal of Chemical Information and Computer Sciences 15:1 (February 197
pp. C-11

RICHARD H. GIERING

Search Strategies and User Interface†

RICHARD H. GIERING

Mead Technology Laboratories, Research Park, Dayton, Ohio 45432

Received August 30, 1974

This paper discusses what in The Mead Corporation have found to be certain aspects of search strategy with respect to searching large data bases and certain problem areas that remain. The paper defines some problem areas, especially problem areas in the freeing-up of the interface language between the user and the search process. Four specific areas in the interface language are defined, and a recommendation is made concerning basic research into some linguistic meanings for the four areas.

What is a large data base? I would venture a guess that although each one of us has our own definition of a large data base, there are some subtle differences between those definitions. We at The Mead Corporation are involved with large data bases in many classes, and I have found that the people involved with large data bases of the bibliographic class are far ahead of others in their thinking concerning the use of large data bases in general. But the problem areas for other types of large data bases (large for other reasons) also present some formidable difficulties.

Now, under what criteria can a data base be large? First, a data base can be large because the number or records or entries in that data base is large. Five hundred thousand entries in a data base represent a large data base, independently of the size of each of the individual entries. A data base containing merely bibliographic information, including a limited number of manually assigned index terms (on the 500,000 entries), could well require storage of a mere 50 to 100 million characters of information. Not a very large amount of storage in anyone's evaluation, but a large number of entries.

The second criterion for defining a large data base, then, could be the numbers of characters involved. Again, a bibliographic data base with only 10,000 entries in it, but with each entry being the full text of the document involved, represents a data base whose character count is on the order of a billion.

A third measure against which a data base could be considered is a measure very important to the search or selection process. If the number of selectable elements, possibly a controlled thesaurus, is on the order of a thousand, independently of the size of the data base in characters or entries, the data base could be said to be small. On the other hand, in a full-text operation, where every potential word and every potential value, including all variants, are search-selectable elements numbering in the hundred-thousand area, the data base could be said to be large, especially if the software involved required a serial search of those elements.

Now, finally, a fourth condition under which a data base could be said to be large, again independently of the previous three, appears primarily when we get away from bibliographic data bases—and I believe that this Conference should address this question—because, as bibliographic experts, we have led the field. We have been at the "head of the pack," so to speak. In defining Information Science activities, designers of management systems, for example, are just beginning to think about large data bases in the way we in the bibliographic area have been thinking of them for years. This last condition, under which a data base might be said to be large, would be in the number of fields and/or files associated with the intricate make-up of the individual data base.

We will be addressing all four of these areas—all four of these definitions of large data bases—during our discussion. If, during the discussion, there is disagreement between what is said by two people, myself being possibly one of them, or two people within the audience during the question and answer period, it is quite possible that the difference lies in the assumed definition of some of the terms, including this one.

I would now wish to raise another question from a definition standpoint. Who, or what, is the user? We have been inundated recently with calls for standardization of systems specifications of systems intercourse. There have be a wide variety of complaints about having to learn multi disciplines of access for multiple large data bases. I wish to propose a hypothesis and suggest that it be the subject of some discussion. Is it not possible that the reason there is such a dichotomy of user interaction specifications is that the individual definitions of what and who the user is vary? The attributes that are ascribed to this person called "the user" by system one vary and differ greatly from the attributes associated with this person called "the user" by system two. Until we have some commonality of understanding, until we agree on the minimum set of attributes for "the user," we will continue to have a dichotomy of systems.

The understanding and agreement as to what a user is are important in still another area. The on-line interactive use of data bases, large or small, is in its infancy. As the user population grows, it is quite logical to assume that the profile (make-up) of that population will continue to change. Unless we, as an industry, understand this changing phenomenon, we will not be adequately ready to support the end user. I am going to describe this end-user now by *my* definition—this is the definition of the end user profile as we in The Mead Corporation see it. Up until recently, end users, that is, the users on the terminal, have had at least a smattering of training in the Information Sciences technology. They have been able to assimilate meanings of terms such as "Boolean logic," "operators," "arguments," etc. For the most part, we have been building the system for our own inner circle of users. That is not the definition of the user as The Mead Corporation sees it. The definition I am about to give you looks into the future where, *via* the advanced technology of such things as cable TV, the end users of both large and small data bases will be scared out of their wits by such terms as Boolean logic. even though in their normal day-to-day, natural language communication, they use an "AND" and an "OR" all the time.

† Presented in the "Conference on Large Data Bases," sponsored by the NAS/NRC Committee on Chemical Information, National Academy of Sciences, May 22–23, 1974.

[6] C. W. CLEVERDON, *J. Docum.* 1974. 30(2), 170–180.
[7] T. A. DOLOTTA, *Datamation* 1971, 17(11), 11–13.
[8] E. SWARTHE and J. L. LITTLE, Society for Information Display. NBS-SID Soft Copy Conference. U.S. National Bureau of Standards, Washington (1971).
[9] R. V. KATTER and D. A. BLANKENSHIP, On-line interfaces for document information systems: considerations for the biomedical communications network. System Development Corporation, Santa Monica, California (1969).
[10] R. H. GIERING, *J. Chem. Inf. Comp. Sci.* 1975, 15(1), 6–11.
[11] R. V. KATTER, *Ann. Rev. Inform. Sci. Technol.* 1969, 4, 31–70.
[12] S. W. MILLER, *Proc. Am. Fed. Inform. Processing Socs.* 1968, 33(1), 267–277.
[13] W. S. COOPER, *J. Am. Soc. Inf. Sci.* 1973, 24, 413–424.
[14] K. H. COOK, *Proc. Am. Soc. Inf. Sci.* 1970, 7, 111–114.
[15] H. S. HEAPS, *J. Chem. Inf. Comp. Sci.* 1975, 15(1), 32–39.
[16] A. D. BOOTH, *Inform. Control* 1967, 10, 386–393.
[17] W. MOLL, *Bull. Med. Libr. Assoc.* 1971, 59(3), 458–462.
[18] B. T. STERN, *Aslib Proc.* 1972, 24(10), 562–564.
[19] G. MAGYAR, *J. Docum.* 1974, 30(1), 32–40.
[20] Massachusetts Institute of Technology Project Intrex. Semiannual activity reports. 15Sept. 1971–15Mar. 1972 and 15 Mar. 1972–15Sept. 1972. M.I.T., Cambridge, Massachusetts, 1971, 1972. (ED060899 and ED066207).
[21] C. W. CLEVERDON, 2nd. *Int. Conf. Mechanized Storage and Retrieval*, Cranfield 1969, July.
[22] C. W. CLEVERDON, *J. Docum.* 1970, 26(1), 55–67.
[23] D. C. VEAL and B. K. WYATT, *Newsidic* 1974, (12), 18–19.
[24] P. L. HOLMES, Library and information networks with special reference to the U.K. Paper presented at AGARD Conference: National and international networks of libraries, documentation and information centres, Brussels, 2–3 October 1974.
[25] *Newsidic*, 1974, (15), 10–11).
[26] *Newsidic*, 1975, (16), 12–13.
[27] D. R. SMITH, R. O. BEAUCHAMP, J. L. GARBER and M. A. DAUGHERTY, *J. chem. Docum.* 1972, 12(1), 9–13.
[28] C. M. LEE, Comparative analysis of computer retrieval by Excerpta Medica and Medlars of prostaglandin citations (1969).
[29] R. R. BLANKEN, *Encyclopedia of Library and Information Science*, Vol. 8, pp. 261–282. Marcel Dekker, New York (1972).
[30] M. E. WILLIAMS, *J. Chem. Inf. Comp. Sci.* 1975, 15(1), 3–5.
[31] G. W. ADAMSON and J. BOREHAM, *Inform. Stor. Retr.* 1974, 10, 253–260.
[32] C. T. YU, *Inform. Stor. Retr.* 1974, 10, 243–251.

PLANNING AND DESIGN OF ON-LINE SYSTEMS FOR THE ULTIMATE USER OF BIOMEDICAL INFORMATION

R. R. BLANKEN and B. T. STERN
Excerpta Medica, P.O. Box 1126, Amsterdam, The Netherlands

(Received 24 July 1975)

Abstract—Following a general discussion on the philosophy and design of information systems, with particular attention to the definition, needs and psychology of the ultimate user of systems providing on-line access to biomedical information, the role of the documentalist, the differences between document retrieval and true information retrieval and the operational characteristics of on-line systems which affect their cost and hence their design and acceptability, the authors make some tentative predictions as to the future demand for such information retrieval services and their probable organizational form. A brief report is then presented on the principal findings and conclusions of a user's study of the Excerpta Medica system, the key features and history of which are briefly described. Based on the conclusions of this study, particularly as regards the complexity of the average search question, the role of the search formulators in determining the results of computer searching, the importance of secondary concepts for retrieval and the optimal level of specificity of a computer thesaurus, some of the changes in the Excerpta Medica system which are in the planning stage and will be incorporated into the system's Mark II version are outlined, as are the principal features of the two systems currently offering on-line access to the Excerpta Medica database in Western Germany and the U.S.A. Finally, attention is given to the planned partial hierarchic structuring of the Excerpta Medica thesaurus (Malimet), a project which is to be based largely on frequency counts of the existing database and the elimination of over-specific terms by posting under broader concepts. The results of some of the initial steps in this direction (i.e. frequency counts of portions of the database and the structuring of some of the terms used in the cancer field) are presented by way of illustration.

1. INTRODUCTION

On-line access to large literature databases has been available since the end of the 1960s and the experience gained has been proving useful in the major redesigns that are, or will be, taking place within these services. Such design amendments must be seen as distinct from the continuing modifications introduced as a result of hardware changes, increased geographical or time availability and the greater volume of material handled.

With more than 1·5 million references in its computer files Excerpta Medica is one of the world's largest literature services and is now actively engaged in the design and implementation of a new system which will retain the best features of the old, whilst at the same time introducing improvements based on both user experience and in-house production needs.

Excerpta Medica's present Mark I system has been amply described in the literature on several occasions[1–4]. However, one difference between Excerpta Medica and other services should be emphasized. It is a service designed for access by the ultimate user and not by a documentalist intermediary. For this reason, the design of the Mark II system, with a likely increase in on-line access, demands greater care than with other services if the available information is not to be locked into the system as a result of a "system barrier" caused by non-familiarity.

All systems have their peculiarities which need to be appreciated by the user, just as dictionaries or other retrieval tools prove most valuable when there is some familiarity with their intellectual patterns. Documentalists have rightly served as the skilled intermediary between system and user as they have the necessary familiarities with a variety of sources to obtain the best results. As the services are refined and tend towards a single optimum pattern, a so-far hidden barrier to effective retrieval will become critical. This is the unconscious change in the concept transmitted by the user to the documentalist when posing his question, caused by the latter trying to "fit" the question to the system, in order to obtain the highest possible number of relevant references.

For Excerpta Medica, the ultimate user is the person who actually absorbs (and attempts to apply) the ideas from the original article. It is not sufficient, in our view, to provide a system

INFOTEX ASSOCIATES
3101 Kettering Blvd.
Dayton, Ohio 45439

(513) 293-4173

May 9, 1980

Mr. Charles P. Bourne
C/O Lockheed Information Services
3460 Hillview Avenue
Palo Alto, California 94304

Dear Charles:

Thank you for your letter on the history of online services and please accept my apologies for the delay in responding. I am glad to be included. The following information is the best I can recall; as I am no longer associated with Mead, I don't have access to the records so I don't have the total facts. I am afraid that some of the information differs from that you have. First a question however: In reference to your first set of information dealing with SDC in ?/60 - It was my impression that, while the full-text could be retrieved for display in that demonstration, the search was on manually assigned index terms. Please correct me if I'm wrong. Now for the Data Central history:

1. Pre-1964 - A manual library (primarily using KEYDEX) was set up and was contractor operated in support of the Reconnaissance Laboratory (Recon Lab) operations. It (the library) was named RECON CENTRAL.

2. Mid/64 - The contractor operating the RECON CENTRAL facility was Data Corporation of Dayton. The contract was expanded to include the determination of the feasibility of using full-text automated techniques to aid in the operation of RECON CENTRAL. People involved: Bob Roalof and Len Crouch (Recon Lab) and William Gorog (Data Corp). There were, of course, others involved, but I'm not sure I know who they were.

3. 1965 - As theoretical feasibility was determined, the contract was expanded to include the construction of a feasibility (breadboard) model. This was the standard way in which the laboratory worked in the development of Recon-oriented hardware.

4. Mid-1966 - Feasibility software demonstrated in-house at Data Corporation. Scope was expanded to prepare for on-site feasibility testing. Computer involved was an IBM 360-40, one bin of an IBM 2321 data cell (40 million bytes) two 2311 disks (7 megabytes each) and the online console typewriter.

TI sponsored Form 1498's (RD status reports), a demonstration personnel file, etc. It is believed that this was the first public demonstration of a system specifically designed for full-text search.

14. late 68 - Union Carbide started feasibility testing using chemical compound files. Feasibility resulted in a purchase of this breadboard model by Union Carbide. The system was installed in their Charleston, West Virginia plant.

15. Late 68 - The OBAR data base was tested in law firms using teletype terminals. Development of the commercial system continued.

16. Early 1969 (about Feb.) Data Corporation contracted with A. D. Little for a market research survey concerning the viability of a data base business. Main participants were H. Donald Wilson, Jerome Rubin for ADL and Wm Gorog, R H Giering for Data Corporation.

17. Summer 69 - Version one of the commercial system was demonstrated and used for both the legal service from Dayton, Ohio and for a number of government contracts from Data Corporation's data center in Arlington, Virginia. Contracts included (but were not limited to):

    1. COSATI (Committee on Scientific and Technical Information).
    2. BEER (Biological Effects from Electromagnetic Radiation).
    3. EARS (Epilepsy Abstracts Retrieval System).
    4. PADAT (Psychological Abstracts Direct Access Terminal).

    Version one of the system included high speed (1200 baud) black and white CRT terminals in addition to a wide variety of other terminal types. Boolian searching was available including the ability to indicate distance search specifications, universal character and truncated words. The system also had arithmetic search capabilities on structured fields.

1970?

18. Feb 1969 - The Information Systems Division of Data Corporation (Mr Giering and Mr Vann were co-directors of the division) became the nucleus of a new Mead corporate subsidiary: Mead Data Central Incorporated (MDCI). Mr Wilson, formerly of ADL was named President, Mr J. Rubin (also formerly of ADL), Mr. Giering, Mr Vann and Mr. Welch (all from Data Corp.) were named Vice-Presidents. The mission of the new corporation was to exploit the new business(es) available because of the existence of the Data Central system. Development on added capabilities for Data Central continued under Mr. Giering's direction.

19. Mid 1970 - The name of Recon Central was changed to Avionics Central (the laboratory's name was changed). Avionics Central contracted with MTL for installation and use of Data Central. Additionally they began to serve outside clients from other organizations (rather than limiting the support to those within the laboratory). The first client was RPVSPO (Remotely Piloted Vehicle Special Projects Office) of Hq, Aeronautical Systems Division. The files included Reliability forms and trouble reports. Over the next few years Avionics Central grew with, among others, the following applications:

    1. AUGEN Central (For the Air Force Auditor General staff at Norton AFB, Calif.). Files included Reports of audit, Auiit plans, Time (personnel time) studies, etc.

    2. Hq., Dept of the Navy, Pentagon. File was a composite of all Navy regulations.

    3. Defense Audio-Visual Agency (DAVA). Files included a cross-service inventory of Audio-Visual items (Slides, motion pictures, etc) and equipment (both for showing A-V items and for producing them), a personnel capabilities file, a facilities file and a budget file.

    4. Directorate of personnel, ASD. A personnel file.

    5. Headquarters, Systems Command. The file was made up of project information. Named MASIS, it was a Management And Scientific Information Systen.

    6. Naval Training Command. A file of lesson plans was maintained together with a personnel file for instructor capabilities.

    7. Electronic Systems Command, Hanscomb Field, Mass. The files included Preventative Maintenance instructions and reports of the actual PM's.

20. Fall 1970 - The KWIC release of the Data Central system was introduced. This version of the system allowed the online user to obtain a "personalized abstract" of the documents retrieved. This entailed the display of the occurrences of the terms used by the online researcher in the search and the surrounding context (KWIC = Keyword In Context) of those terms with the remaining text replaced by ellipses marks. At the same time, the full-text display included the use of color highlighting (KWIC also means KeyWord In Color).

21. The Legal data base continued to expand and the legal search service was expanded into many more law offices in Ohio. About this same tine, the business expanded by contracting with the New York State Bar Association for support of the New York legal cases.

22. Fall 1971 - The businesses of MDCI were split. The legal search service remained with MDCI and the non-legal contracts were returned to Data Corporation. The Data Central system was initially used by both, but during the next year and a half, MDCI developed a completely new system based on the same technology, but directed to the legal profession. As a result, the Data Central capabilities became the purview of Data Corporation.

NOTE

From the Fall of 1971 thru the present, I have no information concerning the activities of MDCI.

23. Oct 72 - Data Corporation is renamed Mead Technology Laboratories (MTL). Contractual service continued.

24. Early 74 - Online edit capabilities introduced into the Data Central system. This included the immediate update of the source material with the inverted file to be updated with the necessary transactions on the next regularly scheduled batch update.

25. 1974-1975 - Support continued for the Washington, DC client base as well as for the Avionics Central Client base.

26. Spring 76 - The BOSTON GLOBE contracted with MTL for a pilot/feasibility program to automate their news clipping library. The data was to come from the GLOBE's automated typesetting process. All work necessary for the interface was to be on the part of Data Central, no interference in the producing of the newspaper was to be tolerated. Principle participants included Mr. George Collins, The Globe's librarian, Ms Jennifer Chao, ass't librarian and Mr. Giering from MTL.

27. Oct 76 - The Globe's automated library began operation. Manual clipping of the Globe (newspaper) by the GLOBE library was terminated in early 1977.

28. Summer 1977 - The Philadelphia Newspapers (the INQUIRER and DAILY NEWS) contracted with MTL to install a Data Central capability on their inhouse computer for their library. Mr. Joseph DeMarino is the INQUIRER librarian.

29. Dec 77 - Mr Giering left MTL to begin INFOTEX ASSOCIATES.

30. Subsequent to my leaving, Mead transferred the Newspaper business back to MDCI. The contract with Philadelphia has been terminated and they had to return the Data Central system to Mead. The Boston Globe contract was terminated as of the end

of 1979.

I hope the above is helpful in your project; it is my best recollection of the events.

Now to bring you up to date on my activities since leaving Mead. I formed Infotex Associates (a partnership between myself and a Massachusetts firm), immediately upon leaving Mead. We currently have six additional employees. As soon as the firm was formed we began development of a data base system for a mini-computer. It has been in operation at two sites for some time: At the Chicago Sun-Times (Mr. E. Perez, librarian) and at the Boston Globe (Ms. Chao, Librarian - Mr. Collins has been promoted). The GLOBE's data base is just under a billion source characters, containing 250,000 stories for the period Oct 76 thru the present. As soon as the developmental process is fully completed we will begin marketing efforts, selling complete turnkey systems: hardware, software, training, etc., for a wide variety of data base applications.

If I can be of further assistance in your project, please don't hesitate to contact me.

SINCERELY

RICHARD H GIERING
Managing Partner

11 re memo

INFOTEX, INC.
1476 Miamisburg-Centerville Road, Route 725
Dayton, Ohio 45459

(513) 435-8852

September 29, 1983

Mr. Charles P. Bourne
DIALOG Information Services
3460 Hillview Avenue
Palo Alto, California 94304

Dear Charlie:

I'm sorry for the delay in getting back to you with respect to the major events concerning the history of the online industry from my perspective.

While I am not completely positive of the actual dates, the attached is my best recollection of the events with which I was party while at Data Corporation, both before and after it became part of The Mead Corporation. I believe that it dove-tails with your current information; if not please call and we can straighten out the differences.

I hope this is helpful.

SINCERELY

Dick

RICHARD H GIERING
President

Major events in the History of Data Central

The following is an attempt to define the major events in the life of Data Central:

1. ?/64-65 - Experimentation (by Data Corporation under Air Force Contract) and breadboard development started of full-text (text fields) management of project resume's and related information. Goal: to prove feasibility of using text (as opposed to codes) in a management environment. Limited data base, one terminal to a dedicated IBM 360 with limited data cell use. Participants: Gorog, Crouch and Roalof (latter two are AF).

2. Late/66 - Feasibility model completed and feasibility proved. Limited services began as RECON CENTRAL. Planning for development of production system started.

3. Spring/67 - RECON CENTRAL used with CIRC to evaluate and compare full-text search against keyword only search.

4. Oct/67 - AF funding (for development of production system) depleted, planning effort terminated. RECON CENTRAL continues service with limited breadboard system. Data Corporation begins an in-house development effort to generate a generalized full text DBMS to be called DATA CENTRAL. The effort directed by R. H. Giering. OBAR formed and contracted with

Data Corportation to test the feasibility of using the Data Central technology in legal document access. "INFORMATION PROCESSING AND THE DATA SPECTRUM" by R. H. Giering published. Main participants: Bill Gorog, R H Giering (both of Data), James Preston, William Harrington (both of OBAR).

5. Mid/68 - Beta test of initial production Data Central system - Union Carbide. First recursive search full-text system. First full-text service bureau opened by Data Corporation in Washington, DC.

6. Oct/68 - First public demonstration of a commercial full-text system (version one of Data Central) conducted at the ASIS convention, Columbus, Ohio. Simultaneous multiple terminal access (TTY, IBM 1050) against multiple large volume data bases - each with multiple files - using data cells and/or IBM 2311 disk drives. Mead acquired Data Corporation. Initial OBAR data base (3 files) at 50 million characters.

7. Latter half/68 - Data Central operation filmed for COSATI movie. Biological Effects of Electromagnetic Radiation (BEER) project feasibility started at RECON CENTRAL. RECON CENTRAL continues to use limited breadboard system.

8. Spring/69 - Legal research feasibility proved. Development of enhanced Data Central continued with distance (proximity) searching implemented. OBAR begins loading additional re-

revised again for chapt 8 + 10 LEXIS

trospective legal material.

9. Mid/69 - Federal Water Pollution ENVIRON file loaded at the Data Corporation service bureau in Washington. Legal Research Business feasibility study by Arthur D. Little for Data Corporation begun. Main participants: H. Donald Wilson and J. Rubin. OBAR starts evaluation of full-text legal research with a limited number of TTY terminals in law offices.

10. Fall/69 - BEER data base moved from RECON CENTRAL (under an Air Force laboratory) to the Data Corporation service bureau in Washington (under Walter Reed Army Hospital).

TO CB

I can not comment on the KIDD report. I only know that we were doing work for them about that time.

11. Late/69 - Data Central support of IBM 2740 and 2741 terminals began. Support for CCI CRT at 1200 baud implemented. Phych abstract application loaded for APA.

12. Feb/70 - Legal business feasibility shown. Mead Data Central, Inc (MDCI) formed with H. Donald Wilson as president.

Vice-Presidents included R. H. Giering, P. J Vann, R Welch (all from Data Corporation) and J. Rubin. The Information Systems Division of Data Corporation was the nucleus of MDCI. Other CRT service by Data Central announced. OBAR data base purchased by MDCI. Its size at about 550 million characters.

13. Jun/70 - Color (as well as monochrome) CRT service begins from MDCI. Replacement of TTY and TTY compatibile terminals in law firms begins. Extensive selling of Ohio law begins. Believed to be the first full-text information sales activity.

14. Sep/70 - KWIC release - Capabilities of Data Central increased to support the KeyWord In Context (KWIC) display with color highlighting.

15. Oct/70 - The Air Force contracts with MDCI to obtain a copy of DATA CENTRAL to replace the limited breadboard system servicing RECON CENTRAL. RECON CENTRAL begins dial-up service and CRT support.

16. Jan/71 - Negotiations with state bar associations outside of Ohio begins. An audio-visual application is loaded for HEW.

17. Mid/71 - Personnel application (for Mead) implemented.

18. Sep/71 - MDCI makes the decision to concentrate its efforts

on the selling of legal information. J. Rubin becomes President and H D Wilson becomes Vice-Chairman of the Board. Welch resigns. Giering and Vann return to Data Corporation with the non-legal business. Vann subsequently resigns. MDCI begins extensive development (based on the same technology) of a system (to become known as LEXIS) dedicated to legal research under Ed Gottsmann. The Data Central system becomes then two systems: the LEXIS system at MDCI and a general purpose system (called Data Central) at Data Corporation. Continued development continues independently one from the other.

To CB

Subsequent activities about LEXIS/MDCI can not be made since I was not involved.

19. Mid/73 - Data Corporation changed its name to Mead Technology Labratories (MTL).

20. Mid/74 - On-line editing and direct update of the serial file (not of the inverted file) implemented. Washington service bureau closed with service to customers continuing from the Dayton office.

21. Mid/75 - Weighted retrieval (based on the nbr of terms hit)

implemented as part of the sort facility. Data recursiveness (ability to use data retrieved as a result of a given search as the parameters for a subsequent search) implemented. Many other terminal types implemented. Transaction billing for Data Central at RECON CENTRAL (Now known as Avionics Central and later as INFOCEN) started. Ability to support input from terminals having cartridges (or other storage devices) implemented. Special universal characters (numeric only and alpha only) supported.

22. June/76 - Initial contract with the Boston Globe for feasibility test for news library application began.

23. Sep/76 - Boston Globe contracts for first newspaper production data base.

24. Jul/77 - The Boston Globe stops manual clipping of the newspaper. The automated records are the only records available for research. Support for remote terminals with attached printers (without manual intervention) implemented.

25. Sep/77 - Negotiations for implementing a system at the Philadelphia Inquirer began.

26. Nov/77 - Mead Technology Labs forms NEWSLIB to sell news data bases. Initial version of Data Central implemented at Philadelphia Inquirer.

27. Dec/77 - Giering resigns. Leaves to form INFOTEX.

28. Feb/78 - System enhanced for Phila Inquirer (IBM 3270 terminal support plus scrooling).

29. Mid/78 - NEWSLIB and all non-government data base business transfered from MTL back to MDCI. NEXIS is announced. Phila contract cancelled.

30. Late/1980 - MDCI halts support of Data Central business. Data Central system scrapped.

Cy, STRICS

INFOTEX ASSOCIATES
3101 Kettering Blvd.
Dayton, Ohio 45439

(513) 293-4173

May 9, 1980

Mr. Charles P. Bourne
C/O Lockheed Information Services
3460 Hillview Avenue
Palo Alto, California 94304

Dear Charles:

Thank you for your letter on the history of online services and please accept my apologies for the delay in responding. I am glad to be included. The following information is the best I can recall; as I am no longer associated with Mead, I don't have access to the records so I don't have the total facts. I am afraid that some of the information differs from that you have. First a question however: In reference to your first set of information dealing with SDC in ?/60 - It was my impression that, while the full-text could be retrieved for display in that demonstration, the search was on manually assigned index terms. Please correct me if I'm wrong. Now for the Data Central history:

1. Pre-1964 - A manual library (primarily using KEYDEX) was set up and was contractor operated in support of the Reconnaissance Laboratory (Recon Lab) operations. It (the library) was named RECON CENTRAL.

2. Mid/64 - The contractor operating the RECON CENTRAL facility was Data Corporation of Dayton. The contract was expanded to include the determination of the feasibility of using full-text automated techniques to aid in the operation of RECON CENTRAL. People involved: Bob Roalof and Len Crouch (Recon Lab) and William Gorog (Data Corp). There were, of course, others involved, but I'm not sure I know who they were.

3. 1965 - As theoretical feasibility was determined, the contract was expanded to include the construction of a feasibility (breadboard) model. This was the standard way in which the laboratory worked in the development of Recon-oriented hardware.

4. Mid-1966 - Feasibility software demonstrated in-house at Data Corporation. Scope was expanded to prepare for on-site feasibility testing. Computer involved was an IBM 360-40, one bin of an IBM 2321 data cell (40 million bytes) two 2311 disks (7 megabytes each) and the online console typewriter.

5. Mar/67 - On-site demonstration of the RECON CENTRAL software (the feasibility software). Three files (one for the RD status forms - DD Form 1498 -, one for research reports - textual documents - and one for hardware specification records) of varying complexity from textual documents thru combinations of text and numeric fields to primarily numeric fields were used. The system used an inverted file from both the text and the structured fields. Boolian search capabilities were demonstrated, but the support was limited to the online console typewriter.

6. Later (about spring of 67) the capability to support either the console typewriter OR an IBM 1050 hardcopy terminal was introduced. Note that this was an OR condition, the system would support one or the other but not both simultaneously.

7. Late in 1967, funding for the development of the full capabilities was denied. Recon Central capabilities remained in operation using the capabilities in hand only.

8. Fall 1967 - Ohio State Bar Association formed OBAR (Ohio Bar Automated Research) to work with Data Corporation in determining the feasibility of using the Recon Central techniques in the full-text search of legal documents. Main participants included Wm. Harrington and James Preston from OBAR and Wm Gorog, Eugene Bold and R H Giering for Data Corporation.

9. Dec 67 - Mr. Giering published a report on the feasibility of using full-text techniques in data base processing of all kinds of data bases. It is titled "INFORMATION PROCESSING AND THE DATA SPECTRUM".

10. Spring 1968 - Mr Giering assigned responsibility to develop a commercially viable system to process all types of data bases from those containing structured fields to textual fields and-or documents. The project was fully in-house funded by Data Corporation.

11. Summer 68 - A time-sharing capability was developed and tested using the Recon Central package. Simultaneously, work continued on the OBAR project. Feasibility was determined and funding for the data conversion was sought.

12. Aug 68 - Data Corporation acquired by The Mead Corporation.

13. Fall 68 (You can determine the exact month from your records, I think it was September.) Permission was obtained from the Recon Lab to allow the public demonstration of the marriage of the first part of the commercial system (the time-sharing capability) with the Recon Central package. The commercial system, when completed, was to be called Data Central. The demonstration, using multiple IBM 1050 terminals and model 33 teletypes was during the ASIS convention in Columbus, Ohio. Files demonstrated included some OBAR legal texts, some COSA-

TI sponsored Form 1498's (RD status reports), a demonstration personnel file, etc. It is believed that this was the first public demonstration of a system specifically designed for full-text search.

14. late 68 - Union Carbide started feasibility testing using chemical compound files. Feasibility resulted in a purchase of this breadboard model by Union Carbide. The system was installed in their Charleston, West Virginia plant.

15. Late 68 - The OBAR data base was tested in law firms using teletype terminals. Development of the commercial system continued.

16. Early 1969 (about Feb.) Data Corporation contracted with A. D. Little for a market research survey concerning the viability of a data base business. Main participants were H. Donald Wilson, Jerome Rubin for ADL and Wm Gorog, R H Giering for Data Corporation.

17. Summer 69 - Version one of the commercial system was demonstrated and used for both the legal service from Dayton, Ohio and for a number of government contracts from Data Corporation's data center in Arlington, Virginia. Contracts included (but were not limited to):

    1. COSATI (Committee on Scientific and Technical Information).

    2. BEER (Biological Effects from Electromagnetic Radiation).

    3. EARS (Epilepsy Abstracts Retrieval System).

    4. PADAT (Psychological Abstracts Direct Access Terminal).

    Version one of the system included high speed (1200 baud) black and white CRT terminals in addition to a wide variety of other terminal types. Boolian searching was available including the ability to indicate distance search specifications, universal character and truncated words. The system also had arithmetic search capabilities on structured fields.

18. Feb 1969 - The Information Systems Division of Data Corporation (Mr Giering and Mr Vann were co-directors of the division) became the nucleus of a new Mead corporate subsidiary: Mead Data Central Incorporated (MDCI). Mr Wilson, formerly of ADL was named President, Mr J. Rubin (also formerly of ADL), Mr. Giering, Mr Vann and Mr. Welch (all from Data Corp.) were named Vice-Presidents. The mission of the new corporation was to exploit the new business(es) available because of the existence of the Data Central system. Development on added capabilities for Data Central continued under Mr. Giering's direction.

19. Mid 1970 - The name of Recon Central was changed to Avionics Central (the laboratory's name was changed). Avionics Central contracted with MTL for installation and use of Data Central. Additionally they began to serve outside clients from other organizations (rather than limiting the support to those within the laboratory). The first client was RPVSPO (Remotely Piloted Vehicle Special Projects Office) of Hq, Aeronautical Systems Division. The files included Reliability forms and trouble reports. Over the next few years Avionics Central grew with, among others, the following applications:

    1. AUGEN Central (For the Air Force Auditor General staff at Norton AFB, Calif.). Files included Reports of audit, Audit plans, Time (personnel time) studies, etc.

    2. Hq., Dept of the Navy, Pentagon. File was a composite of all Navy regulations.

    3. Defense Audio-Visual Agency (DAVA). Files included a cross-service inventory of Audio-Visual items (Slides, motion pictures, etc) and equipment (both for showing A-V items and for producing them), a personnel capabilities file, a facilities file and a budget file.

    4. Directorate of personnel, ASD. A personnel file.

    5. Headquarters, Systems Command. The file was made up of project information. Named MASIS, it was a Management And Scientific Information System.

    6. Naval Training Command. A file of lesson plans was maintained together with a personnel file for instructor capabilities.

    7. Electronic Systems Command, Hanscomb Field, Mass. The files included Preventative Maintenance instructions and reports of the actual PM's.

20. Fall 1970 - The KWIC release of the Data Central system was introduced. This version of the system allowed the online user to obtain a "personalized abstract" of the documents retrieved. This entailed the display of the occurrences of the terms used by the online researcher in the search and the surrounding context (KWIC = Keyword In Context) of those terms with the remaining text replaced by ellipses marks. At the same time, the full-text display included the use of color highlighting (KWIC also means KeyWord In Color).

21. The Legal data base continued to expand and the legal search service was expanded into many more law offices in Ohio. About this same time, the business expanded by contracting with the New York State Bar Association for support of the New York legal cases.

22. Fall 1971 - The businesses of MDCI were split. The legal search service remained with MDCI and the non-legal contracts were returned to Data Corporation. The Data Central system was initially used by both, but during the next year and a half, MDCI developed a completely new system based on the same technology, but directed to the legal profession. As a result, the Data Central capabilities became the purview of Data Corporation.

NOTE

From the Fall of 1971 thru the present, I have no information concerning the activities of MDCI.

23. Oct 72 - Data Corporation is renamed Mead Technology Laboratories (MTL). Contractual service continued.

24. Early 74 - Online edit capabilities introduced into the Data Central system. This included the immediate update of the source material with the inverted file to be updated with the necessary transactions on the next regularly scheduled batch update.

25. 1974-1975 - Support continued for the Washington, DC client base as well as for the Avionics Central Client base.

26. Spring 76 - The BOSTON GLOBE contracted with MTL for a pilot/feasibility program to automate their news clipping library. The data was to come from the GLOBE's automated typesetting process. All work necessary for the interface was to be on the part of Data Central, no interference in the producing of the newspaper was to be tolerated. Principle participants included Mr. George Collins, The Globe's librarian, Ms Jennifer Chao, ass't librarian and Mr. Giering from MTL.

27. Oct 76 - The Globe's automated library began operation. Manual clipping of the Globe (newspaper) by the GLOBE library was terminated in early 1977.

28. Summer 1977 - The Philadelphia Newspapers (the INQUIRER and DAILY NEWS) contracted with MTL to install a Data Central capability on their inhouse computer for their library. Mr. Joseph DeMarino is the INQUIRER librarian.

29. Dec 77 - Mr Giering left MTL to begin INFOTEX ASSOCIATES.

30. Subsequent to my leaving, Mead transferred the Newspaper business back to MDCI. The contract with Philadelphia has been terminated and they had to return the Data Central system to Mead. The Boston Globe contract was terminated as of the end

of 1979.

I hope the above is helpful in your project; it is my best recollection of the events.

Now to bring you up to date on my activities since leaving Mead. I formed Infotex Associates (a partnership between myself and a Massachusetts firm), immediately upon leaving Mead. We currently have six additional employees. As soon as the firm was formed we began development of a data base system for a mini-computer. It has been in operation at two sites for some time: At the Chicago Sun-Times (Mr. E. Perez, librarian) and at the Boston Globe (Ms. Chao, Librarian - Mr. Collins has been promoted). The GLOBE's data base is just under a billion source characters, containing 250,000 stories for the period Oct 76 thru the present. As soon as the developmental process is fully completed we will begin marketing efforts, selling complete turnkey systems: hardware, software, training, etc., for a wide variety of data base applications.

If I can be of further assistance in your project, please don't hesitate to contact me.

SINCERELY

RICHARD H GIERING
Managing Partner

copy to Trudi

+c MEAD

TO: HISTORY FILE

FROM: Charles Bourne

DATE: 17 June 83

RE: Notes of 11 June 83 meeting with Dick Giering at SLA New Orleans Meeting

I reviewed some of the early history with Dick, to obtain some information about early days of Mead. Dick confirmed that he was in fact the main technical leader and developer of the current Lexis system. He was working at Data Corporation at that time, 1967, when they began their development of an online system under an air force contract in Dayton Ohio. Their first commercial presentation of their online system was done at the 1968 ASIS Annual Meeting in Columbus Ohio. He recalls that there was no other full-text system on the market at that time. With air force project funding, they designed a multi-file, multi-user system from the beginning. Their RECON system breadboard model, required a complete and fully dedicated IBM 360 computer, and 30 megabytes of storage, to support 1 terminal and 1 file. He thought this was in 1966. There was no timesharing of the machine at that time, consequently the full machine was dedicated to that one terminal. The terminal could be a hardwired IBM 1050, or a terminal operating with a modem. This was their project breadboard to demonstrate the feasibility of the retrieval system. He remembers being in a great rush to get the system operable in order to demonstrate at the 1968 Columbus meeting. It was at that meeting that he was able to demonstrate his system that could accomodate 4 terminals easily. The air force project ran out of money, to their surprise, so the company decided to continue the development of the effort as their own commercial venture. This was prior to the project with the Ohio Bar Foundation. Their first real production model was in 1969. The 1968 Columbus ASIS demonstration not only had multiple users (4 terminals), but also multiple files. The user would choose the file by an equivalent of the BEGIN command.

The early system included the KWIC display, and proximity searching. The system was designed as a fulltext search system from the beginning, working primarily to reduce the set retrieved from the prior operation. He confirmed

that, to his knowledge, DIALOG was the first system to use set numbers, that would permit people to use all prior search products in different formulations.

He recalled the AFIPS meeting in Palo Alto, where many of the early designers described their systems, and remembered coming away from the meeting feeling terribly happy because his system seemed to be the best of the bunch at that time.

He continued with the development of the system as the company eventually became acquired by Mead, and expanded into the Lexis and Nexis activities.

CPB:kir

copy to Trudi

DIALOG
INFORMATION SERVICES, INC.
3460 Hillview Avenue
Palo Alto, CA 94304
(415) 858-2700 TELEX 334499

June 14, 1983

Richard H. Giering
President
Infotex, Incorporated
1476 Miamisburg-Centerville Road
Dayton, OH 45459

Dear Dick:

At the New Orleans SLA Meeting we talked about the early history of the online industry. I am trying to document and get agreement on the major milestones of that history, and would appreciate whatever help you can provide to set the record straight.

The enclosed chart represents the major events that I have been able to identify (& document in some cases). As someone who was in there from the beginning, your input would be very valuable to fill in many of the gaps. If you can spare the time, please annotate and return this chart with whatever information you can provide (along with supporting references where possible).

I look forward to hearing from you.

Best regards,

Charles P. Bourne
Director
Product Development

CPB:kir

Enclosure

RICHARD H. GIERING
PRESIDENT

Infotex, Incorporated

1476 MIAMISBURG-CENTERVILLE ROAD
DAYTON, OHIO 45459
513 435-8852

INFOTEX, INC.
1476 Miamisburg-Centerville Road, Route 725
Dayton, Ohio 45459

(513) 435-8852

September 29, 1983

Mr. Charles P. Bourne
DIALOG Information Services
3460 Hillview Avenue
Palo Alto, California 94304

Dear Charlie:

I'm sorry for the delay in getting back to you with respect to the major events concerning the history of the online industry from my perspective.

While I am not completely positive of the actual dates, the attached is my best recollection of the events with which I was party while at Data Corporation, both before and after it became part of The Mead Corporation. I believe that it dove-tails with your current information; if not please call and we can straighten out the differences.

I hope this is helpful.

SINCERELY

Dick

RICHARD H GIERING
President

Major events in the History of Data Central

The following is an attempt to define the major events in the life of Data Central:

1. ?/64-65 - Experimentation (by Data Corporation under Air Force Contract) and breadboard development started of full-text (text fields) management of project resume's and related information. Goal: to prove feasibility of using text (as opposed to codes) in a management environment. Limited data base, one terminal to a dedicated IBM 360 with limited data cell use. Participants: Gorog, Crouch and Roalof (latter two are AF).

2. Late/66 - Feasibility model completed and feasibility proved. Limited services began as RECON CENTRAL. Planning for development of production system started.

3. Spring/67 - RECON CENTRAL used with CIRC to evaluate and compare full-text search against keyword only search.

4. Oct/67 - AF funding (for development of production system) depleted, planning effort terminated. RECON CENTRAL continues service with limited breadboard system. Data Corporation begins an in-house development effort to generate a generalized full text DBMS to be called DATA CENTRAL. The effort directed by R. H. Giering. OBAR formed and contracted with

Data Corportation to test the feasibility of using the Data Central technology in legal document access. "INFORMATION PROCESSING AND THE DATA SPECTRUM" by R. H. Giering published. Main participants: Bill Gorog, R H Giering (both of Data), James Preston, William Harrington (both of OBAR).

5. Mid/68 - Beta test of initial production Data Central system - Union Carbide. First recursive search full-text system. First full-text service bureau opened by Data Corporation in Washington, DC.

6. Oct/68 - First public demonstration of a commercial full-text system (version one of Data Central) conducted at the ASIS convention, Columbus, Ohio. Simultaneous multiple terminal access (TTY, IBM 1050) against multiple large volume data bases - each with multiple files - using data cells and/or IBM 2311 disk drives. Mead acquired Data Corporation. Initial OBAR data base (3 files) at 50 million characters.

7. Latter half/68 - Data Central operation filmed for COSATI movie. Biological Effects of Electromagnetic Radiation (BEER) project feasibility started at RECON CENTRAL. RECON CENTRAL continues to use limited breadboard system.

8. Spring/69 - Legal research feasibility proved. Development of enhanced Data Central continued with distance (proximity) searching implemented. OBAR begins loading additional re-

trospective legal material.

9. Mid/69 - Federal Water Pollution ENVIRON file loaded at the Data Corporation service bureau in Washington. Legal Research Business feasibility study by Arthur D. Little for Data Corporation begun. Main participants: H. Donald Wilson and J. Rubin. OBAR starts evaluation of full-text legal research with a limited number of TTY terminals in law offices.

10. Fall/69 - BEER data base moved from RECON CENTRAL (under an Air Force laboratory) to the Data Corporation service bureau in Washington (under Walter Reed Army Hospital).

TO CB

I can not comment on the KIDD report. I only know that we were doing work for them about that time.

11. Late/69 - Data Central support of IBM 2740 and 2741 terminals began. Support for CCI CRT at 1200 baud implemented. Phych abstract application loaded for APA.

12. Feb/70 - Legal business feasibility shown. Mead Data Central, Inc (MDCI) formed with H. Donald Wilson as president.

Vice-Presidents included R. H. Giering, P. J Vann, R Welch (all from Data Corporation) and J. Rubin. The Information Systems Division of Data Corporation was the nucleus of MDCI. Other CRT service by Data Central announced. OBAR data base purchased by MDCI. Its size at about 550 million characters.

13. Jun/70 - Color (as well as monochrome) CRT service begins from MDCI. Replacement of TTY and TTY compatibile terminals in law firms begins. Extensive selling of Ohio law begins. Believed to be the first full-text information sales activity.

14. Sep/70 - KWIC release - Capabilities of Data Central increased to support the KeyWord In Context (KWIC) display with color highlighting.

15. Oct/70 - The Air Force contracts with MDCI to obtain a copy of DATA CENTRAL to replace the limited breadboard system servicing RECON CENTRAL. RECON CENTRAL begins dial-up service and CRT support.

16. Jan/71 - Negotiations with state bar associations outside of Ohio begins. An audio-visual application is loaded for HEW.

17. Mid/71 - Personnel application (for Mead) implemented.

18. Sep/71 - MDCI makes the decision to concentrate its efforts

on the selling of legal information. J. Rubin becomes President and H D Wilson becomes Vice-Chairman of the Board. Welch resigns. Giering and Vann return to Data Corporation with the non-legal business. Vann subsequently resigns. MDCI begins extensive development (based on the same technology) of a system (to become known as LEXIS) dedicated to legal research under Ed Gottsmann. The Data Central system becomes then two systems: the LEXIS system at MDCI and a general purpose system (called Data Central) at Data Corporation. Continued development continues independently one from the other.

To CB

Subsequent activities about LEXIS/MDCI can not be made since I was not involved.

19. Mid/73 - Data Corporation changed its name to Mead Technology Labratories (MTL).

20. Mid/74 - On-line editing and direct update of the serial file (not of the inverted file) implemented. Washington service bureau closed with service to customers continuing from the Dayton office.

21. Mid/75 - Weighted retrieval (based on the nbr of terms hit)

implemented as part of the sort facility. Data recursiveness (ability to use data retrieved as a result of a given search as the parameters for a subsequent search) implemented. Many other terminal types implemented. Transaction billing for Data Central at RECON CENTRAL (Now known as Avionics Central and later as INFOCEN) started. Ability to support input from terminals having cartridges (or other storage devices) implemented. Special universal characters (numeric only and alpha only) supported.

22. June/76 - Initial contract with the Boston Globe for feasibility test for news library application began.

23. Sep/76 - Boston Globe contracts for first newspaper production data base.

24. Jul/77 - The Boston Globe stops manual clipping of the newspaper. The automated records are the only records available for research. Support for remote terminals with attached printers (without manual intervention) implemented.

25. Sep/77 - Negotiations for implementing a system at the Philadelphia Inquirer began.

26. Nov/77 - Mead Technology Labs forms NEWSLIB to sell news data bases. Initial version of Data Central implemented at Philadelphia Inquirer.

27. Dec/77 - Giering resigns. Leaves to form INFOTEX.

28. Feb/78 - System enhanced for Phila Inquirer (IBM 3270 terminal support plus scrooling).

29. Mid/78 - NEWSLIB and all non-government data base business transfered from MTL back to MDCI. NEXIS is announced. Phila contract cancelled.

30. Late/1980 - MDCI halts support of Data Central business. Data Central system scrapped.

copy to Trudi

RICS

DIALOG
INFORMATION SERVICES, INC.
3460 Hillview Avenue
Palo Alto, CA 94304
(415) 858-2700 TELEX 334499

April 11, 1984

RECEIVED
APR 18 1984
DIALOG GENERAL INFORMATION SERVICES

Richard H. Giering
President
Infotex, Inc.
1476 Miamisburg-Centerville Road
Dayton, OH 45459

Dear Dick:

Thanks very much for the very helpful information you sent with your September 29th letter. That kind of detail from someone who was on the scene at the time is extremely difficult to obtain.

My delay in responding to your letter is due to the press of regular work assignments, and certainly not a reflection of my interest in the data! I hope to be able to get back to this history hobby project soon, and I'm sure I'll have lots more specific questions for you at that time.

Lots of luck with your new venture.

Best regards,

Charles P. Bourne
Director, General Information Services

CPB:kir
84102DIS0015

CB - pls note comments & return. You should ask Rick C. & Mark Baylor their recollections, too

VZ

R.K. SUMMIT

INFOTEX, INC.
1476 Miamisburg-Centerville Road, Route 725
Dayton, Ohio 45459

(513) 435-8852

September 29, 1983

Mr. Charles P. Bourne
DIALOG Information Services
3460 Hillview Avenue
Palo Alto, California 94304

Dear Charlie:

I'm sorry for the delay in getting back to you with respect to the major events concerning the history of the online industry from my perspective.

While I am not completely positive of the actual dates, the attached is my best recollection of the events with which I was party while at Data Corporation, both before and after it became part of The Mead Corporation. I believe that it dove-tails with your current information; if not please call and we can straighten out the differences.

I hope this is helpful.

SINCERELY

Dick

RICHARD H GIERING
President

Major events in the History of Data Central

The following is an attempt to define the major events in the life of Data Central:

1. ?/64-65 - Experimentation (by Data Corporation under Air Force Contract) and breadboard development started of full-text (text fields) management of project resume's and related information. Goal: to prove feasibility of using text (as opposed to codes) in a management environment. Limited data base, one terminal to a dedicated IBM 360 with limited data cell use. Participants: Gorog, Crouch and Roalof (latter two are AF).

2. Late/66 - Feasibility model completed and feasibility proved. Limited services began as RECON CENTRAL. Planning for development of production system started. Database? — RKS comments 4/84

3. Spring/67 - RECON CENTRAL used with CIRC to evaluate and compare full-text search against keyword only search.

4. Oct/67 - AF funding (for development of production system) depleted, planning effort terminated. RECON CENTRAL continues service with limited breadboard system. Data Corporation begins an in-house development effort to generate a generalized full text DBMS to be called DATA CENTRAL. The effort directed by R. E. Gierlng. OBAR formed and contracted with

Data Corportation to test the feasibility of using the Data Central technology in legal document access. "INFORMATION PROCESSING AND THE DATA SPECTRUM" by R. H. Giering published. Main participants: Bill Gorog, R H Giering (both of Data), James Preston, William Harrington (both of OBAR).

5. Mid/68 - Beta test of initial production Data Central system - Union Carbide. First recursive search full-text system. First full-text service bureau opened by Data Corporation in Washington, DC.

6. Oct/68 - First public demonstration of a commercial full-text system (version one of Data Central) conducted at the ASIS convention, Columbus, Ohio. Simultaneous multiple terminal access (TTY, IBM 1050) against multiple large volume data bases - each with multiple files - using data cells and/or IBM 2311 disk drives. Mead acquired Data Corporation. Initial OBAR data base (3 files) at 50 million characters.

7. Latter half/68 - Data Central operation filmed for COSATI movie. Biological Effects of Electromagnetic Radiation (BEER) project feasibility started at RECON CENTRAL. RECON CENTRAL continues to use limited breadboard system.

8. Spring/69 - Legal research feasibility proved. Development of enhanced Data Central continued with distance (proximity) searching implemented. OBAR begins loading additional re-

trospective legal material.

9. Mid/69 - Federal Water Pollution ENVIRON file loaded at the Data Corporation service bureau in Washington. Legal Research Business feasibility study by Arthur D. Little for Data Corporation begun. Main participants: H. Donald Wilson and J. Rubin. OBAR starts evaluation of full-text legal research with a limited number of TTY terminals in law offices.

10. Fall/69 - BEER data base moved from RECON CENTRAL (under an Air Force laboratory) to the Data Corporation service bureau in Washington (under Walter Reed Army Hospital).

TO CB

I can not comment on the KIDD report. I only know that we were doing work for them about that time.

11. Late/69 - Data Central support of IBM 2740 and 2741 terminals began. Support for CCI CRT at 1200 baud implemented. Phych abstract application loaded for APA.

12. Feb/70 - Legal business feasibility shown. Mead Data Central, Inc (MDCI) formed with H. Donald Wilson as president.

Vice-Presidents included R. H. Giering, P. J Vann, R Welch (all from Data Corporation) and J. Rubin. The Information Systems Division of Data Corporation was the nucleus of MDCI. Other CRT service by Data Central announced. OBAR data base purchased by MDCI. Its size at about 550 million characters.

13. Jun/70 - Color (as well as monochrome) CRT service begins from MDCI. Replacement of TTY and TTY compatibile terminals in law firms begins. Extensive selling of Ohio law begins. Believed to be the first full-text information sales activity.

14. Sep/70 - KWIC release - Capabilities of Data Central increased to support the KeyWord In Context (KWIC) display with color highlighting.

15. Oct/70 - The Air Force contracts with MDCI to obtain a copy of DATA CENTRAL to replace the limited breadboard system servicing RECON CENTRAL. RECON CENTRAL begins dial-up service and CRT support.

16. Jan/71 - Negotiations with state bar associations outside of Ohio begins. An audio-visual application is loaded for HEW.

17. Mid/71 - Personnel application (for Mead) implemented.

18. Sep/71 - MDCI makes the decision to concentrate its efforts

on the selling of legal information. J. Rubin becomes President and H D Wilson becomes Vice-Chairman of the Board. Welch resigns. Giering and Vann return to Data Corporation with the non-legal business. Vann subsequently resigns. MDCI begins extensive development (based on the same technology) of a system (to become known as LEXIS) dedicated to legal research under Ed Gottsmann. The Data Central system becomes then two systems: the LEXIS system at MDCI and a general purpose system (called Data Central) at Data Corporation. Continued development continues independently one from the other.

9/71

To CB

Subsequent activities about LEXIS/MDCI can not be made since I was not involved.

19. Mid/73 - Data Corporation changed its name to Mead Technology Labratories (MTL).

20. Mid/74 - On-line editing and direct update of the serial file (not of the inverted file) implemented. Washington service bureau closed with service to customers continuing from the Dayton office.

21. Mid/75 - Weighted retrieval (based on the nbr of terms hit)

implemented as part of the sort facility. Data recursiveness (ability to use data retrieved as a result of a given search as the parameters for a subsequent search) implemented. Many other terminal types implemented. Transaction billing for Data Central at RECON CENTRAL (Now known as Avionics Central and later as INFOCEN) started. Ability to support input from terminals having cartridges (or other storage devices) implemented. Special universal characters (numeric only and alpha only) supported.

22. June/76 - Initial contract with the Boston Globe for feasibility test for news library application began.

23. Sep/76 - Boston Globe contracts for first newspaper production data base.

24. Jul/77 - The Boston Globe stops manual clipping of the newspaper. The automated records are the only records available for research. Support for remote terminals with attached printers (without manual intervention) implemented.

25. Sep/77 - Negotiations for implementing a system at the Philadelphia Inquirer began.

26. Nov/77 - Mead Technology Labs forms NEWSLIB to sell news data bases. Initial version of Data Central implemented at Philadelphia Inquirer.

27. Dec/77 - Giering resigns. Leaves to form INFOTEX.

28. Feb/78 - System enhanced for Phila Inquirer (IBM 3270 terminal support plus scrooling).

29. Mid/78 - NEWSLIB and all non-government data base business transfered from MTL back to MDCI. NEXIS is announced. Phila contract cancelled.

30. Late/1980 - MDCI halts support of Data Central business. Data Central system scrapped.

Returned - Address Unknown.

March 21, 1995

Richard H. Giering, President
Infotex, Inc.
1476 Miamesburg - Centerville Road
Dayton, OH 45459

Dear Dick:

Now that I've retired from DIALOG, I am able to spend more time working with Trudi Bellardo, formerly of Catholic University and SLA, to write a book for Academic Press on the early (pre-1976) history of the online search services. You may remember that I discussed that topic with you several times in the past..

The Mead and Data Corporation story will be a part of that text, and at this time we are actively reviewing the final text of that story. Because you were so closely involved with the those activities during the pre-1976 time period, we'd appreciate whatever help you can provide with our final reviews. We've gone about as far as we can go from the published material that we've been able to get our hands on, and from earlier information that you've provided. Now we need to have the current draft checked by the people who were on the scene at that time -- to correct the factual mistakes, fill in some of the missing pieces, and to provide additional comments as appropriate. We'd also appreciate any stories or anecdotes that we can repeat for our readers.

With that introduction, I invite you to review and annotate the attached draft text of the Mead, Data Corporation, and related activities for 1964-72. The 1973-76 activities will be covered in a later chapter.

You'll notice that some of the text is in boldface. That's just a temporary artifice to permit me to keep track of my own text, so that I can keep track of where things came from. You'll also see some notes passed between Trudi and me as part of the dynamic text-building and review process.

415/322-7101

If you have any questions or comments, please give me a call. Just annotate and return the draft if that's easier for you. I look forward to hearing from you.

Best regards,

Chuck

Charles P. Bourne
1619 SANTA CRUZ AVE.
MENLO PARK, CA 94025

Enclosure - Chapter 8 (3/11/95 edition)

P.S. Could you give me an approximate date for the enclosed announcement of the availability of the Data Central Software?

xc: Trudi Bellardo (letter only)

MEAD TECHNOLOGY LABORATORIES

Through our masking techniques, a single terminal command provides any operator with the information necessary to use any transaction; if that operator and terminal have security clearance. Our Data Base Analyzer allows reporting on any field, or combination of fields, within the system. Demographics and file statistics can now be used for production of marketing or "snapshot" reports. Pre-conversion analysis and model modification for each bank give excellent conversion results. The system interfaces with your present application balance information, on-line or batch, and even non-automated applications may be incorporated with ease. And yes, we can interface KOMPOZIT+ with your present monitor, if you so choose.

IBM 360/370, DOS/OS/VS - BAL, ANSI COBOL
Mr. Jim Greenwood . . . . . PRICE UPON
President REQUEST
Automated Financial Systems, Inc.
One Decker Square, Suite 420
Bala Cynwyd, PA 19004 Tele. 215-667-1000

Circle **86** *on Reader Service Card*

VANDEX INFORMATION RETRIEVAL SYSTEM

VANDEX is a compact and powerful data base system for managing large quantities of data and easily retrieving specific data by simple categorical requests. The system contains three principal files — a dictionary file, an inverted file, and a master file, and also the modules to maintain them and retrieve data from them. The system operates on the well-known principles of coordinate indexing as developed by Dr. Calvin Mooers and implemented by Dr. Warheit in the Combined File Search System. Advances in computer technology along with superior computer software have now made it possible to perform the same work at many times greater speed and lower cost. VANDEX can scan millions of documents and find the information in fractions of a second. All three of the files used in this system are maintained on magnetic disks in a randomly accessible mode. This organization permits the loading of a file on an integral number of cylinders with variable size overflow and index areas. The updating operation is very quick and simple. There is no requirement for any order at all in the sequence of the cards in an update job since these are processed on a random basis.

IBM 360, 12K, DOS/OS - COBOL
Mr. John Boulavko . . . . . PRICE UPON
President REQUEST
Vanguard Information Systems, Inc. SALE
P.O. Box 330 or LEASE
Peter Stuyvesant Station
New York, NY 10009 Tele. 212-477-2034
or
256 S. Robertson Boulevard
Beverly Hills, CA 90211 Tele. 213-659-4210

Circle **87** *on Reader Service Card*

RAMIS ®

A complete program designed to perform all of the tasks needed to set up and operate an information system from the smallest to very largest. Particular emphasis is given to information systems operated in a time-sharing environment such as TSO or VM-370. RAMIS contains: a) Hierarchical file structures for naturally describing the relationship of data fields. b) A transaction processing language which permits new records to be added or deleted from files and existing records updated, without writing computer programs. c) A powerful report generator which allows a user to summarize and sort data and perform calculations by simply describing what is desired in ordinary English supplemented with some special rules of grammar. No forms or coding sheets are required as all communication to RAMIS is with free-form English sentences. Graphical as well as tabular reports can be produced. d) A RUN EXECUTIVE which permits a designer of an application to catalog not only the often repeated activities but also the logical decisions the system is to make in response to actual operating circumstances, e.g., if too many new records are rejected, then don't print the usual reports but some other diagnostic report. e) A host language interface which permits programs coded in COBOL, FORTRAN or PL-1 to access and change records in a RAMIS data base and effectively integrate the RAMIS files with some ongoing but non-RAMIS system. RAMIS is an ideal system for use by non-EDP personnel in such areas as Personnel, Financial Planning and Market Research. When EDP assistance is available, large and sophisticated systems can be implemented in a fraction of the time it would take to write programs. There are over 400 RAMIS applications, about half of which use communication terminals. Options are available for preparing financial-type reports based on models, e.g., cash flow balance sheets and for using the RAMIS report writer to read IBM/IMS and other special file systems.

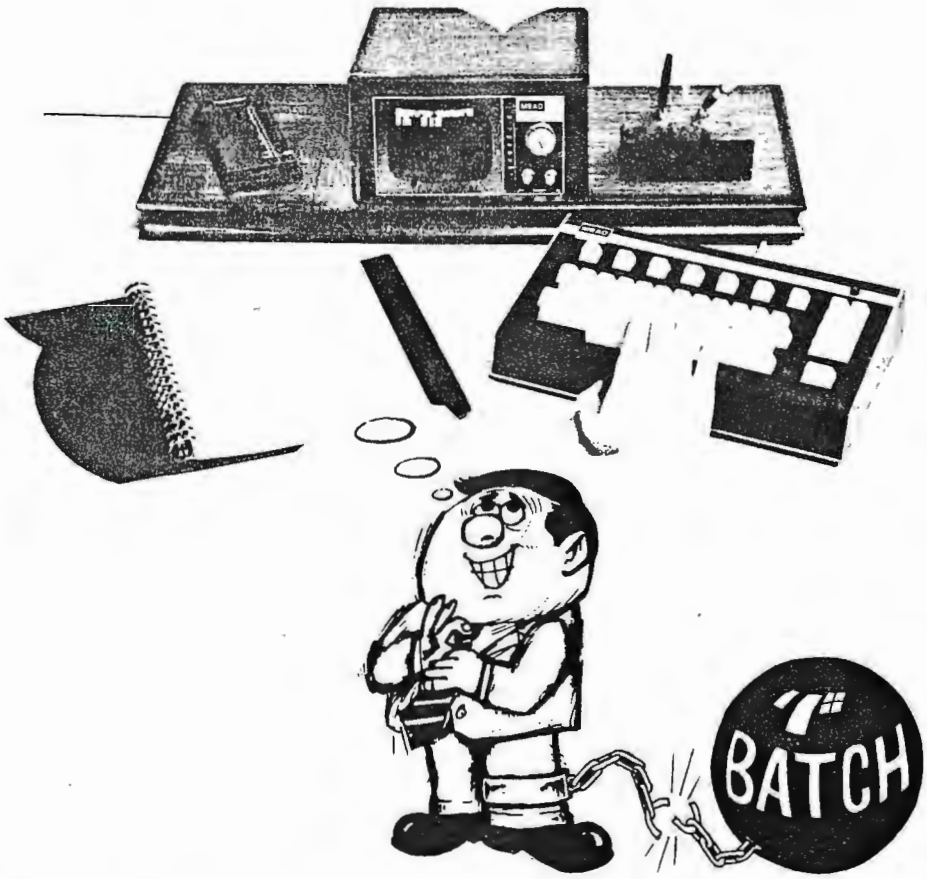
IBM 360/370, OS, Batch Version (also VS-1, VS-2), TSO, CP/CMS and VM/370 Conversational Version BAL, FORTRAN
Mr. Gerald Cohen . . . . . $28,000.00 - $55,000.00
MATHEMATICA, Inc. or LEASE
Princeton Station Office Park $840.00/MO.
P.O. Box 2392
Princeton, NJ 08540 Tele. 609-799-2600
Or
Mr. Frank Fish
MATHEMATICA London
Roxburghe House
Regent Street
London W1, England Tele. 629-2822

Circle **88** *on Reader Service Card*

(DATA/CENTRAL) INFORMATION SYSTEM (4 Programs & Subsystems)

(DATA/CENTRAL) is a general purpose terminal oriented flexible information processing research system. With respect to on-line access and display, update, and data base definition, it is a self-contained fully operational system. With respect to extensive complex input (including on-line input and editing) and output formatting, it is a host language extendable system. (DATA/CENTRAL) executes in its roll-out/roll-in virtual environment allowing large numbers of simultaneous terminals without a large core memory requirement. Access is highly interactive command structured and relatively free-form English. All information in the data base is searchable with automatic unit of measure conversion on arithmetic material and automatic keyword conversion on textual material; all textual material is searchable at the word level. The (DATA/CENTRAL) data base definition is an extension of the DBTG definition which allows for, in addition to standard file format description, the establishment of other user oriented

the GREAT ESCAPE !



it's not a dream... it can be reality !

Now you can enjoy the benefits on your computer of the versatile data system called (Data/Central).

Proven—The system has been in operation for over five years. Its customers can attest to its capabilities.

Fast—Its unique inverted file concept (at the word/value level) allows questions to be answered in seconds.

Reliable—It recovers all errors, and only the terminal causing the error is restarted.

Inexpensive—Operating as its own roll-out/roll-in executive allows for large multiterminal operation in only modest amounts of core.

Multipurpose—Indexing on any value and/or word also allows (Data/Central) to process text (both search and retrieval).

Existing Languages—(Data/Central) interfaces to existing languages (including COBOL), so your programmers do not have to learn a new language to generate report programs into the (Data/Central) library.

Use your existing files and start planning your great escape now.

1972+

DATA CENTRAL

Contact Dick Giering. . .

Mead Technology Laboratories

3481 Dayton-Xenia Rd., Dayton, Ohio 45432 • (513) 426-3111

Circle **89** *on Reader Service Card*

parameters including a fully operational and changeable on-line tutorial for non-ADP users. Multiple tutorials for multiple levels of non-ADP oriented users can be generated with different syntax. Data bases are considered to be made up of up to 256 separate files, all searched separately or by sets of files. Each file may contain up to 60,000 fields. Terminal support includes all IBM hard-copy, teletype and teletype-compatible and color CRT's. Data base definition compilation flexibility allows the external view of the (DATA/CENTRAL) data base to be modified without it being reloaded.

IBM 360/40 & Up, 370/145 & Up, 2311, 2314, 3330, 2321, Core: 100K & Up, OS, DOS, MVT, MFT, VS1, VS2 - Interfaces with any Programming Language, Including COBOL. Its DML is Via CALL.

Mr. Richard H. Giering ..... $1,800.00/MO.
Mead Technology Laboratories LEASE
Research Park Dep. on Application
Dayton, OH 45432 & Usage at
Tele. 513-426-3111 Service Center

Circle **90** *on Reader Service Card*

SCORE III/IV FILE MANAGEMENT/REPORT GENERATOR

The SCORE System is a useful tool that greatly simplifies the preparation of file management and reporting applications. The System accepts non-procedural request forms filled out by the user and generates custom-tailored COBOL which can then be compiled and executed to perform the requested function. User entry capability allows for easy insertion of 'free form' COBOL own code along with user specifications. This permits one step compilation of SCORE generated and user written COBOL statements. Current releases allow for use of standard COBOL Source Statement libraries as data definitions, eliminating the need to redefine data files already in use. Interface to non-standard data bases such as DLI and TOTAL are also available. SCORE on-site training and installation is provided including work shop sessions. The basic features are quickly available to programmers and non-programmers as well. SCORE has been installed in well over three hundred (300) computer facilities and is a member of the ICP Million Dollar Software Club, as well as elected to the Data-Pro Software Products Honor Roll.

IBM 360/370, BURROUGHS B-2500, 3500, 5000, NCR CENTURY, UNIVAC 1106, 1108, 9000, CDC 3000, 6000, HONEYWELL H-200, RCA SPECTRA 70, SIEMENS

Mr. Edward Opengart ..... SCORE III $12,000.00
Programming Methods or LEASE $400.00/MO.
Division GTE SCORE IV $15,000.00
Information Systems, Inc. or LEASE
1301 Avenue of the Americas $480.00/MO.
New York, NY 10019 Tele. 212-489-7200

Circle **91** *on Reader Service Card*

INQUIRE

A self-contained, general purpose, information retrieval and data base management system designed to provide those functions normally required for the creation, maintenance, retrieval and administration of data files in a remote terminal, batch, remote batch, or time-sharing environment. Multi-terminal operation is available using TSO or similar OS-based time-sharing systems. INQUIRE is covered by U.S. Patent 3670310. Active INQUIRE applications currently operational include grant management information, bibliographic literature searching, accident data reporting, manpower analysis and reporting, equipment inspection and analysis pharmaceutical patent searching, screening of biological test data, clinical data analysis and reporting, financial analysis and planning, sales reporting, and many others. Retrieval is accomplished using a free-format, English oriented Command language which involves either random or sequential processing of the file. The user identifies the information using full Boolean logical connectors such as AND, OR, AND NOT, and a unique operator, LINK, which insures retrieval based upon commonality of levels of data within repeating groups. Additionally, the searching of text and testing of field values or ranges of values is performed using CONTAINS, EXCLUDES, IS, TO, GT, LT, GE, LE, or EQ. INQUIRE includes a complete report writing capability. Multi-file link (MFL) available for inquery and reporting of up to 31 data bases simultaneously.

IBM 360/370, OS/VS, MVT-MFT, HASP, TSO, IMS/DC, (IMS/DB avail. in 1975), RJE, 1DA device, 130K real. Terminals include 2780, 2741, TTY, 2260, 3270 etc.

Mr. Robert S. Stahl ..... MO. RENTAL
Commercial Marketing Manager /LEASE Avail.
or
Mr. Edward H. Carlson
Manager, Government Systems
Infodata Systems Inc.
5205 Leesburg Pike, Suite 701
Falls Church, VA 22041 Tele. 703-578-3430

Circle **92** *on Reader Service Card*

TUMS – THE TOTAL UTILITY MAINTENANCE SYSTEM

TUMS is a general purpose data base maintenance utility program containing six basic functions for TOTAL data bases. These are: 1) dumping a TOTAL data set to a backup tape, 2) reloading a TOTAL data set from a backup or source tape, 3) printing a TOTAL data set in vertical hexadecimal-character format (no TOTAL control records, no blank records – just data), 4) building a TOTAL data set from card input, 5) deleting records from a TOTAL data set, and 6) validating a TOTAL data set. TUMS can select TOTAL data according to volume and key values. TUMS can be used for a multiplicity of applications. TUMS gives the Data Base Administrator the ability to create the original data base, add new records, delete old records, backup/recover, generate test data bases from live data, and repair data sets with broken linkages. A very important capability of TUMS is reorganizing TOTAL data bases when changes in device type, record size, block size, new relationships (linkage paths), expanded file size, or new data sets are required. The purpose of TUMS is to eliminate expensive custom written programs. You code only four (4) easy-to-use control cards which contain the same information as you might code in a TOTAL program. No compilation, or linkediting --- all binding is at execution time. The elapsed time from request to guaranteed results is immediate. TUMS is shipped ready-to-catalog with no local customization necessary. It installs in less than five minutes. And you will be able to use it within a 1/2 hour with the users manual.

CHARLES BOURNE AND ASSOCIATES
1619 SANTA CRUZ AVENUE
MENLO PARK, CALIFORNIA 94025
TEL. (415) 322-7101

August 27, 1995

Richard H. Giering
5873 Westhaven Drive
Fort Worth, TX 76132

817/263-7259

Dear Dick:

Now that I've retired from DIALOG, I am able to spend more time working with Trudi Bellardo, formerly of Catholic University and SLA, to write a book for Academic Press on the early (pre-1977) history of the online search services. You may remember that I discussed that topic with you several times in the past..

The Mead and Data Corporation story will be a part of that text, and at this time we are actively reviewing the final text of that story. Because you were so closely involved with the those activities during the pre-1976 time period, we'd appreciate whatever help you can provide with our final reviews. We've gone about as far as we can go from the published material that we've been able to get our hands on, and from earlier information that you've provided. Now we need to have the current draft checked by the people who were on the scene at that time -- to correct the factual mistakes, fill in some of the missing pieces, and to provide additional comments as appropriate. We'd also appreciate any stories or anecdotes that we can repeat for our readers.

With that introduction, I invite you to review and annotate the attached draft text of the Mead, Data Corporation, and related activities for 1964-77. I hope you'll also be able to review the COSATI activities that resulted in your appearance in the Battelle/COSATI movie.

You'll notice that some of the text is in boldface. That's just a temporary artifice to permit me to keep track of my own text, so that I can keep track of where things came from. You'll also see some notes passed between Trudi and me as part of the dynamic text-building and review process.

If you have any questions or comments, please give me a call. Just annotate and return the draft if that's easier for you. I had a difficult time tracking you down. I look forward to hearing from you.

Best regards,

Chas

Charles P. Bourne

Enclosure - COSATI text and supporting cites from Chapter 5 (10/9/94 edition)
Chapter 8 (3/11/95 edition)
Mead text and supporting cites from Chapter 10 (5/20/95 edition)

P.S. Could you give me an approximate date for the enclosed announcement of the availability of the Data Central Software?

xc: Trudi Bellardo (letter only)

To: History File

From: Charles Bourne

Re: Notes of 19 September 95 Call from Dick Giering (home & office: 817/263-7259)

Dick called in response to my recent letter to him. He'll review the draft and send a response in a week or so. He's retired now, and has time to do it.

Regarding the CIRC vs. Data Central issue at Wright-Patterson. He was at DIA and participated in a review. It was very political. He'll tell me the story, but not for the record.

He'll be in San Francisco next month, and could stop in on his way to Monterey.

He may have the Data Spectrum article.

5873 Westhaven Drive
Fort Worth, Texas 76132

(817) 263-7259

September 23, 1995

Charles Bourne and Associates
1619 Santa Cruz Avenue
Menlo Park, California 94025-5761

ATTN: Charles P. Bourne

Dear Charlie:

I am gratified that someone is putting together an unbiased account of the "History of Online Search Services". Most articles seem to be biased toward one activity or another: Battelle, Lockheed, Mead, SDC, etc. Your inclusion of each viewpoint lends a level of authenticity not found other places.

In hind sight, these activities - together with others such as the ARPA spinoffs (e.g. Packet switching that evolved into what is now Sprint) can be looked upon as the "birth" of the Information Superhighway.

Not seeing the overview of the document (e.g. the Table Of Contents), it is difficult to know what information is relevant and what is not. Some comments may not be proper for the actual book.

It is important to note that most of the technical details in the chapters are minute modifications of the same technology. The state-of-the-art movement in commercial enterprise was and is Online interactive access.

I am enclosing copies of five documents I wrote over that period of time. Some of which you asked for in your text; others asked for I have not been able to locate.

One of the enclosed documents deals with the Data Spectrum: I felt then and I feel just as strongly now that having two different classes of systems (each with its own protocol and user interaction), one for structured material and one for document handling is and was superfluous. Codes are words - albeit non-natural-language words, but words non the less - and if document handling systems expanded to handle numeric fields, the document handling systems, with proper Data Base Administration, could easily handle what is currently thought of as MIS or data base requirements. Sadly, the rest of the world doesn't agree.

I look forward to seeing you and discussing your most intriguing project. Our current plans for our San Francisco trip include Friday, October 20 as a day to the Monterey area. Could you meet us somewhere along 101 for a cup of coffee or something. Please let us know.

Sincerely,

Dick

Richard H Giering

Enclosures:

Notes and commentary on dispatched material
Analysis of Existing and Proposed Data Handling Systems
Information Processing and the Data Spectrum
This is Data Central (1972 Technical Specifications
Presentation to Conference on Large Data Bases, National Academy of Sciences
TTAM A Time-Shared Teleprocessing Access Method

\family\bourne95.not: September 17, 1995, 11:44 PM
Modified: September 17, 1995, 11:44 PM
Printed: September 25, 1995, 4:40 PM

1 COSATI text and supporting cites

1.1 First paragraph: "disseminating scientific and technical information ..." dealt primarily with the use of form 1498 (DD form, NASA form and other departmental forms of similar or identical nature).

1.2 Second paragraph: sentence containing Landau reference: As I remember it, Battelle was also invited to participate in the evaluation. They declined. Note they made their own movie.

1.3 Second paragraph - even though the information was published in the 1969 ASIS proceedings, the work was accomplished - as I remember it - some time (maybe the year) before. This goes along with the RKS notes referenced on page five.

1.4 It should be noted that prior to 1969, the names DIALOG and Data Central were not in general use. Only the corporate names - Lockheed and Data Corporation were used.

1.5 Page 74, "... two of the systems (_____ and ____) ...". If my memory serves me correctly, the two were Lockheed and Data. In answer to your question in the CB Note: I believe the demo was later, in 1969 or 1970.

1.6 Next paragraph, question note: I believe the Noreen Welch/Mitre report was that issued officially in August, 1968, but it was unofficially distributed in spring, 68 to the COSATI members and to the involved members for comment. After issue, I believe it was available from NTIS.

1.7 Page 75, bottom paragraph: I believe CCA also participated. Battelle partially participated by using their own data base of library collections (see previous paragraph) and some dictionary definitions.

1.8 Page 76, top: I don't know when George Tressel did all of the filming, but do remember that the film had four vendors: Battelle, CCA, Data and Lockheed.

2 Chapter Eight - Data Corporation ...

2.1 Page 6, bottom: The initial computerized system for RECON CENTRAL

was non-traditional batch. Requesters would give the requirements to an operator who, using the computer console (similar to an IBM 1050 terminal) to interactively obtain the requested data. True online interactive capability was subsequently integrated (circa 1967), but limited to the use of only one IBM 1050 terminal instead of the computer console. At this point in time the data (both new and replacement records) were only entered into the system in true batch mode. Charlie: I would think that this identifies, at least for the Data Corporation system, the transition from classic batch to what is the current concept of online interactive operation.

2.2 Page 7: Note questions at top of page: Data Corporation was actively involved in highly sensitive (at that time) super-high resolution photography - both aerial and satellite photo reconnaissance. This involved research and development in both camera equipment and processing technology. The processing technology was used by Data Corporation - under contract to NASA - in the Lunar Ranger and other moon-related programs. In this photo area, work was being done (research primarily) in high resolution photo display which culminated in the multi-color ink-jet activity that later became the other Mead spin-off - Mead Dijit - that was later sold to Kodak. Data was also involved in sensitive visual and electronic targeting activities. Charlie: Is this detailed enough or do you need more?

With respect to CIRC and CIRCOL, both activities were of a different WPAFB organization (FTD), there was no relationship. SDC had a contract for support at the one organization and Data had the contract at the other. I'll let you draw your own conclusions. In about 1966, when DIA heard about what Data Corporation was doing at the Recon Laboratory (later renamed the AVIONICS Laboratory) while they were funding FTD's activity in CIRCOL, they requested and got a copy of unclassified CIRC data and loaded it at RECON CENTRAL. Technically, the comparison favored the Data Corporation approach, but the report was buried. On the next page, the year is given as 1967. If that is what I reported in my earlier correspondence, I was wrong. The study activity started in late 1966, but for a number of reasons the data did not actually move until early 1967.

Question about "of an online system" in the next paragraph: In the sense of an online system as stated above with a single terminal connected to a dedicated computer.

Charlie: your note in that same paragraph is indicative of several non-communicating activities all moving in the same direction

toward an online activity for document access
Next question note (at Data or WP): The initial computer installation was in a vault at the Recon Laboratory at WPAFB. Data Corporation's computer was installed later (circa early 1967 or late 66). Next note is correct. The final note in that paragraph is correct, but with the amplification that the feasibility was the breadboard model. Except for the date, the note on the bottom of page 7 is correct.

2.3 Page 8, top: I might add that the implication of "numeric information" should include arithmetic (e.g. gtr, lss, etc.) searching.

Middle paragraph, page 8: While we in Data Corporation were aware, as was DIA (see above), of the FTD work with CIRC, the classified nature (and "need to know") requirements within FTD prohibited any detailed knowledge of it. I was not aware that "funding shifted", but only that the funding to proceed to a full production system was curtailed, but I believe that was in '68 rather than '67.

With respect to the "hotbed of invention" comment - The Recon Laboratory had sponsored a research contract to Northwestern University sometime earlier on the same "full-text" topic. This resulted in a set of programs (not a system) that was batch oriented for the IBM 7094 and used FORTRAN logical variables with Boolean logic for full-text searching. It was called BIDAP (Bibliographic Data Processor) and was employed by DIA in the late 1966 early 1967 time frame.

Question in last sentence of middle paragraph: See above for an answer as to why (FTD vs. Recon Lab). I can verbally give more information as I was involved at DIA, but won't put it in writing.

Question on the bottom of page 8: "Full capabilities" included, among other things, time-sharing for multiple simultaneous terminals, other terminal types, more generalized data base definition (including cross-field and cross-file access), more and deeper recursiveness, optional (personalized) synonymy and SDI, KWIC display, columnar display with sub and final totaling, more generalized input and output sub-routines (now called filters), etc.

The library continued using the computer with the "breadboard" system. Data Corporation (Bill Gorog) obtained entrepreneurial funding and began the in-house development of the "full system". Recon Lab, having received as a deliverable, the breadboard system, continued to use it until the 1969/1970 time frame when

Data Corporation/Mead Data Central sold/leased them the rights to the then production version. From then on the full Data Corporation/Mead system (programs) were used by the Air Force at RECON CENTRAL, the Avionics Central and finally AVCEN, all at WPAFB. This just adds dates to that which is elaborated later in your text.

2.4 Page 9 - middle of the page, question about dates: As I remember the sequence, the late 1967 date, I was given the responsibility for the design and implementation of the time-sharing functionality being the first step in the expansion to a "production capability". Design, programming and testing saw the full implementation of the capability in early 1968. Up until that time Bill Gorog and others of the Data Corporation management were unsure that a production capability could be obtained within the planned (and budgeted) dollars and time frame. The implementation convinced them and I was tapped to manage (up to that time I was merely the design manager and programmer) the complete conversion and business including the data centers (Washington, DC, Dayton and WPAFB - Data still had the personnel support contract with Recon Lab, etc.) It depends on which window one looks through to determine which of the two dates is valid for what.

Charlie, see below for information on my background up to 1970

Comments with respect to "tested with the RECON CENTRAL package". While the time-sharing capability was being developed, significant internal (no functionality changes) modifications of the breadboard system was accomplished to give greater reliability, etc. The test "with the RECON CENTRAL package" (as opposed to the now upgraded system) was an attempt to get the Recon Lab people interested enough to buy into the time-shared capability.

Except for the dates that are more or less accurately stated in the earlier part of the chapter (the part on OBAR), the comments dealing with OBAR feasibility relate to the first part of the on-going OBAR activity. The first step in the project was an attempt, by Data Corporation, to prove that the Data Central (this is about the time-frame when that name began to be used) capability could support legal research. A small (about 50 million character) data base was loaded for testing; I don't remember if OBAR funded that effort or if Data funded it internally or if a combination. One of the results of that effort was the realization of the magnitude of the data conversation effort. It was determined that neither Data nor OBAR would have the funds for the effort.

Comment on the note on bottom of page 9: whether it was "recursive" or "semi-recursive" sounds like the type of semantic commentary used by the various vendors of the day in their attempt to one-upmanship their competition. Roger's comment "single answer set" is accurate, but recursiveness is a little like being pregnant; can one be semi-pregnant? I don't think that, thirty years later it really matters. The concept of recursiveness was being implemented by both vendors; the mode of implementation matters not nearly as much as the fact that it was being implemented. He is absolutely accurate in that Data Central, of that vintage, had no index display capability. This item was on the list of functionalities that were to be implemented. Each vendor had their own ideas of which items had what priority. For example, while Lockheed had index display as a relatively high priority, Data Corporation thought that KWIC and color display were their high priority. I believe this difference is the direct result of the philosophic background of the two companies - manual indexing vs. full text. But is this really germane to the question of the advance of the state-of-the-art?

2.5 Page 10 - Statement ahead of the CB Note at the bottom: The key to the statement is the phrase "on the market". Other systems were available, but were not "on the market"; for example, LEADERMART from Lehigh University was available, but was not commercial or saleable.

I look forward to your research into the SDC-SATIRE and 1963 SRI work. As far as ASPEN is concerned, I have no idea even if they ever went online commercially. A number of years later I ran into Larry Berul (I think this is the correct spelling) in Washington; he was heading up ASPEN at that time, and they had switched from being a provider to performing primarily data entry work for Government and for litigation support. Horty left shortly after OBAR confirmed its decision to go full-text.

2.6 Page 11, top: Subsequent use of stop word lists have generated questions, especially with respect to proximity searching: are stop words counted or not, etc.

Note on "OBAR ... tested in law firms": I remember that we had some OBAR material at the ASIS demonstration in fall, 1968, but I'm not sure that we had the full 50 million characters available at that time. It was this 50 million character data base that was tested in law firms and resulted in feasibility being accepted.

Note in the middle of the page: As related in the previous

paragraph, the late 1967 contract was for a feasibility test and required the loading of the 50 million character data base. Subsequent amendments to the contract (subsequent to feasibility being established) called for the online service, etc.

Charlie: you might want to consider contacting Bill Harrington (LEXIS and/or Don Wilson might give you information on how to contact him) to get OBAR's view of the picture at that time. This might include the answer to what other firms made presentations to OBAR at that time.

2.7 Page 12, top (and bottom of page 11): OBAR: a salient point you might wish to include, although it may not be germane, is that in Ohio and many other states (at least at that time) the copyright to the printed court decision publications was in the hands of the state bar association.

Note on top paragraph: The modifications "deemed necessary" were all in the original functional-item list discussed earlier concerning the movement from the RECON CENTRAL breadboard to the "full production" system. They were all in process of being implemented at that time (time of the subsequent modifications to the original 1967 contract - after feasibility). The modifications were to be finished prior to the contract amendments taking affect. An aside: "changing the stop word list" required that the generalized capability for having different stop word lists for different applications be finalized (see above for comment about Generalized Data Base Definition.

Note as middle paragraph: While the contract was substantially as stated in the paragraph, subsequent amendments and modifications essentially nullified all of the elements. This is especially true about exclusivity, OBAR paying for and owning the data base, and the division of revenue.

Last sentence and note - bottom of page 12: While it is true that GOROG sought out the Ohio Bar Association, it has been the OBAR contention that, since they had a "comparative study" (especially vs. Horty's ASPEN) they sought out Data Corporation. Whether Preston's opinion is valid or not is moot. Whether the subsequent ADL study would have been so emphatically in favor of the legal research market or not, or indeed if an ADL study could have been afforded or not are all open questions. I agree that the sentence needs to be amplified to clarify the apparent conflict. There were a number of other applications available (as is seen later in the text) that were available to Data Corporation as it was seeking customers for its retrieval software.

2.8 Page 13, top: In continuation of the previous comment, Preston's quotation must be taken with "a grain of salt". The time frame being discussed is 1967/68 and 69 while the OBAR contract was operable with Data Corporation. Mead Data Central (MDCI) was not formed until Feb. 1970 and he refers only to MDCI.

Middle paragraphs: In the 1968/69 time frames, things (with respect to legal research) were moving very rapidly and specific dates are difficult to remember Pricing definitely was experimental as no one else was "in the marketplace" so there was no guidance as to the price sensitivity of the product. A great deal of concern was voiced about the price sensitivity vis-a-vis volume of data available, type of terminal and line speed in use, etc. The $75/hour figure is probably a good medium or average, but discussions - both in house at Data Corporation and with lawyers around the state - ranged from much lower numbers to amounts more than double that figure.

My recollection is that initial testing in law offices did not begin until the 1969 time frame (this was while the ADL study was in process). Full beta testing around Ohio (in law firms) did not occur until 1970. As far as who arranged the testing, it was a combination of Data Corporation sales people and OBAR representatives.

As far as how the data base was constructed - very little if any change has occurred in the data base structure for case law records in the intervening period. The opinion was broken into the Majority Opinion, Concurring Opinion, Dissenting Opinion, etc. Other fields included Syllabus, head notes, case name, Judges, date decided, and others. It is interesting to note that if one queries the LEXIS data base of OHIO SUPREME COURT cases, there will be some found in upper case only (as opposed to the vast majority being in upper-lower case). These are cases from the original 50 million character test data base (they were keypunched).

2.9 Page 14, top: While the paragraph (starting on the bottom of page 13) is correct, the implication of Mead's (financial) participation at that time is incorrect. Except for funding the ADL study, very few Mead funds were expended on any Data Central application until the following year with the founding of MDCI.

Next paragraph note: The first sentence is absolutely correct. OBAR was one of a number of clients for the Data Corporation Service bureau. With respect to Diana McCabe's article: one

should remember that she was an employee of OBAR (she is - or at least was - Bill Harrington's wife) and wrote from its viewpoint for the benefit of other lawyers. In only one respect - in my opinion - do you short-change OBAR. OBAR took a lot of flak in that time frame from many and sundry organizations - both legal and non-legal - concerning this "new fangled" idea of full-text storage and retrieval. One must remember that the accepted principle of legal research at the time was the West Key Number system which was, effectively, a manually assigned keyword (actually key numbers, each of which represented a key phrase) in a somewhat controlled vocabulary. They stood their ground and "took the arrows in their back" as we all did at the time. The last comment by Trudi is correct (see above).

Middle of the page: Since MDCI was not formed until the 69/70 time frame (and this section (by the Table of Contents) deals with 68/69, the title should be Data Corporation. In the early days Mead Data Central, Inc was referred to as MDCI. Years later when it was more fully incorporated into the Mead family and lost its separate incorporated status, it became Mead Data Central (MDC).

Date of acquisition: You can (and should) verify the date of acquisition from Mead Corporation and/or other stock market sources (e.g. Value Line or Dow Jones). I specifically remember that the employees of Data Corporation were told of the acquisition in August, 1968. The deal was made about that time, but the date of official close of the sale is at least then or maybe later.

Mark Bayer's "legend" is very factual, with one difference and one incorrect implication. Data Corporation was jointly and equally owned by Bill Gorog and Lysle (I'm not sure of this spelling) Cahill. Bill was the driving force behind cameras and computer technology, while Lysle was more interested in photography (film, etc) and other reconnaissance applications. As I remember the story, it was Lysle (not Bill Gorog) that was Jim McSweeny's neighbor. While it is my understanding that OBAR was mentioned in the backyard discussion, it counted for no more or less than a number of other activities such as the high resolution photo processing and ink-jet work. That so-called legend was common knowledge among Data's employees at the time.

More on the acquisition topic: In my opinion, Harrington's note has a measure of creditability. Regardless of the nature of the backyard discussions, I do remember the following - I was director of the Information Systems Division (ISD) of Data at the time and was responsible for all of the computer oriented business (which

included not only information retrieval activities such as OBAR and RECON CENTRAL, but other non-informational retrieval activities in which Data was engaged. In about Feb., 1969 when members of the Mead Staff came to the Data Corp location to "find out what they had gotten" and how it could fit into their corporate family, it was quite a shock for them to find that the ISD had such a significant on-going business while simultaneously pursuing developmental activities toward even bigger business opportunities in the future. OBAR and its implications appeared to be one of the more significant future opportunities.

2.10 Page 16, top: Leadermart - independent of the dates - did indeed use word positioning. As I remember it, however, they used it, not in the direct inquiry sense as is the current accepted definition of proximity searching, but only as ordering between the semantic search expression vs. the semantics of the text being searched.

By the summer of 1969, Data had also sold and installed a beta version (pre-version one) of the system at Union Carbide's Charleston, West Va. facility for use in their in-house tracking of research in the chemical field. This is in addition to the two commercial (and one government - RECON CENTRAL) computer service bureau operations.

Response to Lancaster's "complex plural" question. The answer is "Yes, complex plurals were supported." The problem with this statement is that the support was not algorithmic, as was the support for regular formed plurals, but rather non-regular plurals were handled as synonyms where they each had to be listed in a list.

2.11 Page 17, top paragraph, note: My recollection of the events leading up to DOE/RECON deal with COSMIC. NASA, having contracted with Lockheed, received a copy of the software for RECON. They in turn passed it on to COSMIC (the Government's holding facility for government owned software, located somewhere in Georgia). When DOE decided they needed a system, getting the software from COSMIC and contracting with Lockheed for modifications was deemed to be cheaper than any other route - no open source bid was solicited.

Second sentence under OBAR: This is one of the modifications to the Data Corporation/OBAR contract I related above.

BEER: There was no additional Wright-Patt contract for the load onto RECON CENTRAL; it was part of the ongoing RECON CENTRAL Effort. When it was moved to Arlington, it became a different

contract.

TIMPS/ENVIRON: I don't believe it started as early as 1966, as the service bureaus were not in operation until late 1967 or early 1968.

2.12 Page 18, PADAT: PADAT was operational on the Arlington, Va. Data Corp Service Bureau computer. APA made it available to their subscribers (I am not aware of their rules for access, Data served only as a service). While a significant amount of their data was converted to machine readable form by Data's service bureau, some of the data may well have been constructed by Inforonics; I only know we converted (both at our in-house facility and via off-shore contract) some of their data.

EARS: no comment or disagreement with contents or notes of page 18 EARS information.

2.13 Page 19, Note after chart: There was indeed a lost opportunity for MDCI. During the Spring and Summer of 1971, MDCI agonized over a business direction decision. One faction wanted to concentrate on legal research while another wanted to exploit the technology across the marketplace including legal research. During this period, many opportunities in the non-legal marketplace were lost. If you want to investigate that story, be my guest, but be aware there are a lot of "worms in it".

NINDS: The National Institutes of Health Neurology Institute used, as I remember it, the EARS contract to experiment with neurology abstracts. No additional contract resulted.

2.14 Page 20, note at top of page (starts on bottom of previous page): The universal character was a significant feature of early Data Central Capability. It may have been (and possibly was) a feature of the original RECON CENTRAL breadboard.

Note about looking for publications: I am reminded that Martha Williams, who was active in ASIS was, at that time with the University of Illinois. She might be able to help locate some material.

COSATI: I agree with your note; Data Central, as a system, did not have the capacity to "agree" on anything.

2.15 Page 21, HEAP: Our Arlington facility had data relating to pesticides, but from where, I don't remember. My recollection is that it was part of our EPA activity - see below. It dealt

primarily with antidotes and treatments to pesticide poisons.

A.D. Little Consultation: As I remember it, the contract with ADL was with Data Corporation, but was directed and funded by Mead. Negotiations began early (not as early as Feb.) 1969. The work was done, as I remember it, during the summer months and culminated in a report in the Aug./Sep./Oct. time frame. These dates can be clarified (I will stand corrected) by Don Wilson. You should contact him on the question.

The original contract (with ADL) called for a study of how the Data Central technology could be exploited - from a business standpoint - without being restricted to a given marketplace. After initial evaluation, the non-legal market was considered too small (to put another way, there was no market identified outside of the legal research business that was considered big enough to warrant considering.) As a result of this initial evaluation, the contract was changed to that as specified in your text.

Comment on the latter part of the note: Wilson is also a lawyer, but non-practicing.

2.16 Page 22, top: The date of the report (Feb., 1970) is not correct. The formation of MDCI began in late November, early December of 1969 and was completed, including incorporation, by Feb., 1970.

Indented material: "Rubin served as President and Wilson as Vice-President" (of RW Development Corporation). See also comments above as to name: MDCI vs. MDC.

A note on the contents of the middle paragraph: The LEXIS system contains/contained very little in new functional features. It relies on the same developed technology of inverted files with the original functionality. Its primary claim to fame was a bias in interaction to legal research, extremely high simultaneous terminal use with no degradation and high reliability. In fact, as late as the late 1980's, the non-interactive (e.g.update, etc.) functionality was unchanged from that of the late 60's and early '70's.

Note the dichotomy between the lower paragraph on page 22 and that of the lower middle paragraph ("In September 1971 ..."). The original formation of MDCI (text on page 22) spun all information science technology business from Data Corporation to MDCI. This included the Arlington facility (which was primarily in support of the governmental and commercial (e.g. PADAT) services). The Information Science activities of the RECON CENTRAL contract were

also transferred as was support for the Union Carbide contract. See below for information on the Sept., 1971 split. While legal research (e.g. OBAR) was included in the spin off, it was not the only activity. The referenced (46 and 49) material was prepared for law journals and was therefore biased. Vann is Peter J. Vann. Bob Bennett was hired during 1970 or early 1971.

2.17 Page 23: Paragraph dealing with Bob Bennett ("many Data Corporation officers ...") should indicate the officers mentioned in the previous paragraph ran MDCI. Bob Bennett assumed operational control after the Sept., 1971 split; as you state on page 30, Bob Bennett was Director of Training in the 1970 time frame. Also Mead Technology Laboratories, at that time did not exist.

The next paragraph deals with a mixture of the pre- and post-Sept., 1971 period. The main thrust of the paragraph deals with the post Sept., 1971 split. The middle sentence ("The mission of the new corporation ... Data Central system.)") relates to the original mission of MDCI. This mission was changed in the post-1971 period to concentrate on the Legal research business.

The next paragraph is correct, but I believe the phrase "... new system based on the same technology ..." needs to be emphasized if the book is to relate to milestones in technologic development; no new technology came about with this development - see above.

The last paragraph should indicate that Welch resigned in 1971 and P J Vann resigned in about (date may not be exact) late 1972. It is this "non-legal business" for which the newly re-constituted Information System Division was responsible.

2.18 Page 24, question on MTL date: I don't remember the exact date of the name change. One of the enclosed documents I wrote (the one dealing with the 1972 specifications of the system) was written in March, 1972 and the entity was still Data Corporation then.

You indicate the "alternative" meaning for KWIC - I assume you, somewhere earlier in the book, define KWIC as "KeyWord In Context". The term "personalized abstract" was really a "personalized extract". I thought that was the way it was advertised, but I may be wrong. A KWIC display was the keywords used by the online user in the search as found in the resulting documents together with their (the keywords) surrounding context. Each of these contexts in the document was separated from the next and/or previous context with an ellipsis mark. This display, especially when in hardcopy, was considered an extract

personalized because it was constructed by virtue of the terms used by the online user at the time.

Response to the CB Note and the TB question in this paragraph: If the Mead section deals with this version of Data Central (prior to the LEXIS version), the answer to the question should be yes, otherwise it would depend on whether LEXIS had the capabilities.

SSIE: Previous comments on applications included commentary as to whether the experiment continued, if it didn't, why not, and so forth. Should this section be similarly expanded?

2.19 Page 25, AVIONICS CENTRAL (starts on page 24): The contract was with Mead Data Central, Inc as Data Corporation (MTL did not exist until after March, 1972) had turned over the system to MDCI when it (MDCI) was formed. Also it was AVIONICS CENTRAL (rather than RECON CENTRAL) that thus began the dial up service. This activity of AVIONICS CENTRAL ("private-file" service becoming a commercial service bureau in competition with Data Corporation) led to some major disagreements in subsequent years. More on this if you like when and if we meet.

HEW: This data base contained files on audio-visual equipment, both hardware and software (software being slides, films,etc.) as well as A-V project information. Mr. Tom Joyce of HEW was the project manager. The application was patterned after the DOD Audio-visual application indicated in the previous paragraph on AVIONICS CENTRAL. The HEW application continued into the 1973 time frame during which time a concerted government wide effort was underway to integrate all audio-visual files into one application. I don't know the outcome of the effort.

Mead Personnel: A complete classic personnel application was implemented. This included the automatic generation of the various required govermental (e.g. EEO) and other management oriented reports. Full text fields, in addition to the structured fields of data, included resume', job description, annual evaluation, labor oriented reports such as grievences, negotiations, etc.

EPA (Charlie: why the special additional heading?): This data base contained, in addition to the legal and management data as Mark Bayer indicated, data on pesticides, toxic spills, research and chemical company information. Air and water quality files were also contained in this data base. I believe it started before MDCI was formed, continued through MDCI's start-up and returned to Data Corporation following the Sept., 1971 split.

2.20 Page 26, First milestone: OBAR and LEADER (really LeaderMart) were contemporaries. The difference was that while OBAR was a commercial venture (the initial 1967/68 discussions between Data Corporation and OBAR dealt with the question of splitting revenues), the Lehigh University activity was limited to testing of research and development activities. I might add that I was quite impressed with their semantic analysis work, but it appears to have gone nowhere at that time. It would be interesting to see if the work (apparently similar in nature) that has recently been accomplished at CONQUEST, Inc of Columbia, Md (Don Wilson is heavily involved with this) followed the LEADER activities or whether the CONQUEST activity is altogether new.

First paragraph, "in 1969, MDC" should read "in 1969, Data Corporation (MDCI was not formed until late '69 and/or early '70). By the same token, the title is incorrect.

My recollection is that Harrington's subsequent statement concerning Squire, Sanders ... in more correct. While the Ohio Attorney General used the system, I believe we in Data Corporation furnished it free of charge for some time and that Squire, Sanders and possibly Jones, Day subscribed before the Attorney General did. Charlie: I agree with your suggestion - ask Bill Harrington to review this text. He may be able to refresh my memory (or vice-versa) and add to the authenticity of the effort.

With respect to the question as to what computers were being used, the computer configuration was as described at the lower middle of page 8.

2.21 Page 27, Harrington's findings (started on bottom paragraph of previous page). I have referred to (as I believe others have) the 1970 and early 1971 time frame as the MDCI beta test period. There were three major factors that forced us to conclude that the test was something less than a full success: lack of reliability, wrong terminal equipment (very slow - 10 characters a second - and noisy (clattering) teletypes have no place in the dignified surroundings of a law office) and too little data (while the Ohio case law data went back to about 1903, the lawyers wanted it all, regardless of the "cost-effectiveness" - from a business standpoint - of the old data.

Jerry Rubin's comments must be taken in context of the times. The other "competing" automated system was Horty's Aspen (the date at which Horty's operation became known as Aspen, is an open question). Horty would respond to lawyers calling his operation

(in Pittsburg) where his staff of lawyers would digest the calling lawyer's statement of the problem to be researched. The Morty lawyer would then somehow (I believe not via full-text) get material selected from the data base and send it to the requesting lawyer.

2.22 Page 28, top (starting at last paragraph on previous page): I believe the study referred to was performed by Jim Carlisle. He was later associated with the Annenberg (spelling) School at USC and I believe he received an advanced degree there. Where he is now, I have no idea.

What is the difference between the two milestones under the first "1970 milestones"? Why two identical headings?

First base paragraph - KWIC was also used (feasible) on hardcopy terminals such as the teletype, 1050, 2740 and 41, etc. Keywords were highlighted on these terminals by underscoring. Your other statements dealing with manners of highlighting are also correct.

In the second set of "1970 milestones", how does the second milestone differ from the first milestone of the first set?

I question your use of the "judgmental" term (who says they were) "gaudy". Also, citations were in yellow ONLY for display formats of citations only. In full text display, yellow was used to define the context (a la KWIC) of the keywords. For example, if the red keyword was to appear on the top of the next page (screen full), the bottom of the current page would be yellow warning the online user that the context was starting. Remember that, in those days terminals had no internal memory and the re-display of a previous page to read context required the re-transmission of the previous page. While at 120 cps this was faster than the teletype, it still was a delay of about seven seconds.

2.23 Page 29, top paragraph: The term "personalized extract" and "OBARizing" (Note the similarity to the term "Shepardizing".) were synonyms.

Second paragraph note: what reconciliation is desired, if I can help, please call.

1970 milestone (again): This paging feature was implemented shortly following the implementation of support for CRT terminals. This would be late 1969 or early 1970.

2.24 Page 30, note about my AFIPS meeting attendance: I believe that

was the Stanford University meeting as referenced in number 62 (see above and below for more).

Note on bottom of page: As president of OBAR, Asman's were for lawyer's consumption (especially lawyers that had put money into the OBAR bonds.

2.25 Page 31, Terminal communication starting message (as started on previous page): There was a lot of contention, at that time, about the generalized verbage of communication and this message was at the crux of the disagreement. There were many customers of the Mead Data Central Service using the Data/Central computer system; OBAR was only one, albeit a large one. This generalized message orientation was a significant factor that led to the Sept., 1971 split. OBAR wanted this message (and others like it) to be "legal oriented" such as "YOU HAVE REACHED THE OBAR LEGAL RESEARCH SERVICE".

Mark Bayer's comments are quite accurate. In addition, when CRT service was established (and into the subsequent Data Corporation non-legal period), we (including yours truly) used to carry two big boxes (in excess of seventy pounds each and checked as luggage on airline trips) to demonstrate the system. It reminded one of Willie Loman in "Death of a Salesman" with his big sample case; our sales people had two of them. That was no picnic!

To the next paragraph (concerning early missionaries), I can only say "AMEN".

The note on the bottom half of the page: It should be noted that the copyright (to the print version) to the state's case law was, for the most part, with the state bar association. Some of these rights were restricted while others were unrestricted - it varied state by state.

2.26 Page 32, continuation of previous note: The "partnership" question was indeed a business/marketing question; it had nothing to do with the technical nature of the service.

By early 1971, support (technical as well as management) to the non-legal marketing effort was reduced to an almost non-existent state. As a result many opportunities in the non-legal (especially governmental) world were not followed-up upon. The fact that fewer and fewer non-legal opportunities became contracts, lent creditability to the belief on the part of some management people that the non-legal business was doomed to failure.

McCabe's article: I remember many discussions between Diane (or Diana, my notes show both forms of her name) and members of MDCI's management concerning the affordability of the service. The difference between being a profitable business and being a philanthropic enterprise continues - with respect to many parts of the information highway/superstructure today.

2.27 Page 33, Asman's quote: many of those early users did not have the patience and endurance necessary to carry on. The early successes of LEXIS were in New York and other states. Many of the Ohio lawyers had a "bad taste in their mouths" as a result of those early days and they did not get back into it for a long time.

Outstanding Issues - HEAP: As stated above (comments for page 21 and 25) the Data Central system was used for EPA's Pesticide data base. I will try to get additional information if necessary.

BIO Data: See below - if more is needed, contact me.

Unsuccessful bidder: We bid on many RFP's in those days including NTIS as early as 1968, Library of Congress (about the same time frame), National Farmer's Home Association (part of Dept of Agriculture, at a later time), etc. I can't remember all of them.

Publications: Five of the six stated publications deal with Data Corporation. I am enclosing a copy of five documents I have in my possession; three of the enclosed five are listed as desired. I do not have either of the 1968 documents.

2.28 NEXIS - Today's operation is known as LEXIS/NEXIS. I believe it might well be germane to your discussion to indicate from where this very significant part of the business came. The part the Boston Globe and the Philadelphia Inquirer played in originating newspaper data bases is given short shrift in chapter ten (see comments on it below). Suggest that a) reference be made here to that section and b) that story be told in greater detail. It is significant, I believe, to note that the original mission of MDCI (all business, both legal and non-legal) is at least partially the case today (what with newspaper data, patent data, medical data, etc) even though the Sept., 1971 split got the company out of any business not directly related to legal research.

3 Chapter ten Info ...

3.1 Page 31, note in the middle of the page: They may have claimed to have no relationship to the Original Data Corporation or OBAR

system and, indeed the external interface (to the users) did not, but as we have seen above, the internal capabilities (full-text word search, KWIC display, etc. and associated functional features) were unchanged from the original. Some of the behind the scenes software was identical - the same software.

3.2 Page 33, "As a private company ..." in first main paragraph: You've hit the nail on the head! That sentence is probably the most telling, significant and free-enterprise related fact in the whole set of materials I have been sent - possibly in the whole book!

Next paragraph, prior to indented material: A syllabus (a separate field found in some case law material is a court generated abstract. Additionally, most published case decisions contain "head notes" which are court assigned index terms (generally uncontrolled or, at best, lightly controlled) for the case. In both instances, the words contained therein are equally searchable.

Next paragraph: I believe (I wasn't there, but this has been the topic of many off-the-record discussions) the original LEXIS system was intended to serve, as a minimum, about 200 simultaneous users WITHOUT DEGRADATION. At points of degradation, additional hardware would be added.

3.3 Page 34, Proprietary terminals: There was another pure-marketing factor involved in the decision of LEXIS to use a special terminal. Law firms that had this terminal could use it only for LEXIS (not WESTLAW, not DIALOG nor any other service that might come along) thereby locking those firms into LEXIS.

Bottom paragraph: The use of "subject terms" will normally imply - to information specialists that have been around awhile - manually assigned subject (index) terms, not full-text words. Also see below for the history of the reverse Boolean meaning. The term "conversational English" is critical here.

With respect to distance or proximity searching, the training included suggestions on how to obtain the equivalent of sentence and paragraph searching. Within 25 words would approximate a sentence while within 100 words was a paragraph.

3.4 Page 35, Rubin's indented material, the sentence "One is that the computer replaces the researcher; ...": the essence of the difference between the LEXIS/Data Central philosophy and end users and that of other online search services is embodied in this

statement! Other services assume and in fact perpetuate the concept of surrogate searchers (not researchers and not the person with the problem) or "para-professionals at the terminal. LEXIS and the original Data/Central philosophy employ the concept that only the "END-user" - the person with the actual problem - is able to know which material accurately answers the problem at hand.

Last paragraph, "four billion source characters": I believe that when the unit-of-measure of character count (rather than records) is used, LEXIS consistently, over the period from 1973 to the present, has had the single largest data holdings in the world. This is a significant milestone. See also the last full sentence on the page.

3.5 Page 36, middle paragraph: Suggest changing (unless it is a quote) "Mead purchased MDC for ..." to read "Mead purchased Data Corporation, the corporate platform on which the legal research application - LEXIS - was born for ...". Charlie: if we discuss this verbally at all, ask me about the rest of this paragraph.

Next paragraph: I believe the phrase "first commercial ... significant profits" identifies a milestone.

3.6 Aside comment: A very significant example of changing the "modus operandi of the legal profession" deals with the term "Shepardizing" (Charlie this term may be misspelled). Before the advent of OBAR/LEXIS, the Shepard cross-references had to be accepted by the legal profession as being absolutely complete or, to be completely sure of the completeness, prohibitive amounts of manual labor were necessary. With the advent of OBAR/LEXIS, the verification of the Sheperd's report could be, and was, verified giving credence to that part of the legal brief.

3.7 Page 37, first note: I concur - This should be a separate section.

Paragraph starting "In mid-1975 ...": By mid-1975, several additional features had been implemented (some/many even prior to 1975). These included recursive display (having selected a format for display of a given answer set, the user could request a change of the format for the same display, see some material in the new format, then change back, etc). The ability to jump between random pages (relevant to user) simulating actual hard-copy library research was implemented. One of the display formats included (assuming the privacy restrictions allowed a given user to invoke it) included the ability to edit material being viewed online. This also included direct update of the display material (but not of the inverted file). As a leasable software package it

also had its own DML to allow lease clients to enhance it with their own "filters".

INFOCEN was also known as AVCEN. They operated their own service center in a lease mode.

Next paragraph: The information in this paragraph (except for the last sentence) is extremely significant as it was the effort that led to the formation of the NEXIS portion of today's LEXIS/NEXIS operation. There was a significant amount of disagreement among the newspaper librarian profession as whether this was feasible and/or practical. One must remember that clipping and filing took up a significant amount of the staffing and budget of a news library. Would this cause cuts in either manpower or budget or both? Could it be accomplished without any disruption of service? We've always done it in the clipping way; dare we take this innovating step?. What is the impact of trying to sell access to this library outside of the newspaper? And so forth.

Once the Globe had become a production operation on the Data Central Service Bureau, certain people in MTL management wanted to set up an additional sales force, obtain rights to the news library data and begin to sell it. They saw the profits that LEXIS was beginning to generate and felt they could generate more of the same. Mead, on the other hand, saw the investment that had been poured into LEXIS and its predecessor MDCI, and had serious misgivings. Rather than stay with the proven and profitable (although with smaller than LEXIS margins) service business, MTL continued to pressure Mead to enter the information selling business. Finally, in 1968, the complete information science business was taken from MTL and transferred to MDCI; MTL gambled and lost the whole thing. While the GLOBE operated from LEXIS in a service bureau mode, the LEXIS management used it to approach other news organizations, obtained rights in data and NEXIS was born. Neither the GLOBE nor the Philadelphia (the last sentence of the paragraph) Inquirer lent rights in data to LEXIS. This opened an opportunity to Knight-Ridder (parent of the Inquirer) to start DATA-VU (Contact Joe DeMarino) and to the folks at the Oklahoma newspapers to start Data Times. NEXIS still is the predominate vendor in the news selling game. Charlie you might want to comment on the LEXIS/NEXIS service buying/incorporating the New York Time Info Bank and Dow Jones News-Retrieval Service. There are more details to this story if you want them.

3.8 Page 38, Psych Abs: This was the continuation - following the September, 1971 split - of PADAT. See above.

As far as NTIS is concerned, we did some work with them, but it was more of a test than an actual production service. I don't remember too much of it.

3.9 Page 158: I note that footnote 28 is an interview with "Caputo". If this is Rick Caputo, of DIALOG, you might question him about the APA/PADAT/Psych Abs activity. He worked for MTI, and was project manager (or whatever we called them at the time) for the APA application.

4 General comments and notes:

4.1 In the latter sixties or early seventies, there was a conference at Stanford University - this may,but I'm not sure, be the conference listed as reference 62 in chapter eight. I don't recall all of the participants, but the host was a handicapped bearded gentleman from Stanford (I believe he was a professor or Ass't Professor of Information or Library Science. People attending represented organizations of academia (Stanford, Northwestern and Lehigh), Government and commercial enterprises (Lockheed, Mead and SDC - Battelle may also have been represented). Discussions ranged far and wide concerning the then state-of-the-art, especialy the user's acceptance of what was available, but settled on the major topic of disagreement of the day: indexing vs. full-text (see below) and a very significant part of the business: relevancy vs. recall (especially how to achieve the "best" - whatever that means - response for the end-users) for the two technologies.

4.2 I hope you will be making reference to the major controversy of those early days. As I recall, there were two camps concerned with how data (in those days, primarily bibliographic information) was to be accessed and evaluated.

The advocates for one camp felt that complete material would not be directly available. The data would, of necessity, be bibliographic in nature. This meant that each "document" or "record" would have to be abstracted, extracted or otherwise reduced in size by some manual process. Access would be limited to the use of manually assigned index terms from either a controlled (e.g. the New York Times Thesaurus) or uncontrolled thesaurus. This group believed that the added expense of manually assigning indexing terms was minimal as the base expanse was invested in the manual abstracting/extracting. The cost of converting and storing the full-text was prohibitive.

Advocates for the second camp felt that automated indexing - now

know as full-text indexing - negated the manual costs of both indexing and abstracting. The cost of storage would come down to make it practical.

The first group included Lockheed, SDC and the majority of professional librarians (e.g. Martha Williams. The second group included Data Corporation, CCA, representatives from engineering schools (e.g. Lehigh and Northwestern) and some librarians (e.g. F. W. Lancaster.

Charlie, as you will remember, in those early days both the SDC and Lockheed systems were dedicated (and limited) to the use (for search purposes) of the manually assigned index terms. Data Corporation's system, LeaderMart, and CCA etc. were either full-text or quasi full text entries in the struggle.

4.3 Who was first? ... CCA and SDC dropped out of the information selling game early ... Mead first to sell full-text information as a service (service bureaus operating in both Washington and Dayton in 1968) ... Lockheed first to sell indexed information as data (ERIC in early 1969) ... both advocated online interactive access. Other "firsts" include: first to librarians, first to end-users; first at higher (e.g. 1200 baud) speed transmission; color screens; nationwide information sales force;

It doesn't matter who was first, all should get a measure of credit for advancing the state-of-the-art in online information retrieval - the birth of the information highway.

4.4 Charlie, I appreciate your subtlety on page 76 - I know that Roger and I have both aged in the intervening 27 years, but I would hope that we are both more than barely recognizable. In that same paragraph, if you do find a copy of the movie, I would like to buy a copy. An aside: as late as the mid-70's, we at Mead Technology Labs were still using the 1498 portion of the COSATI data (because of its non-proprietary nature) to demonstrate our capabilities.

4.5 Charlie: I get questions concerning terms and functions used in the industry today. If you're interested in this information, I'll be glad to discuss it when and if we get together. what is a segment? Why does LEXIS use Boolean precedence? Etc.

5 Giering's biographical information

5.1 Joined the Army in 1947 as a high-school dropout.

I graduated from the University of Arizona with a Bachelor of

Science Degree from the Systems Engineering Department in 1962. T taught various Data Processing courses for the University's Continuing Education Program while pursuing Masters level studies.

5.2 Among my military ADP experiences was as Chief of a Military Capabilities Section, ADP Systems Center, Defense Intelligence Agency. Developed and implemented an information system for military capabilities intelligence data. 1965-67. It was here that my interest in full text technology was started.

One of my assignments was to design a "ground (as opposed to air) military Capabilities System known as the Defense Ground Order of Battle System (DIGOBS). The design parameters were to make use of the Formatted File System (FFS). As a result, the design incorporated a large volume of codes and other highly structured data to be used in the system.

The Formatted File Systems (a number of different versions existed) use for DIGOBS was unsatisfactory as information arrives in the form of textual situation reports (sit-reps) and action reports from the front. FFS requires highly structured codes and numbers.

When the design was finished and published, I concluded that the DIGOBS was doomed to failure. Who would or could use this multi-inch thick conversion manual on the front lines to report in the structured manner? I began to look around to find what processing capabilities existed in the area of the manner GI's used in reporting information from the front lines. We found a Fortran oriented set of programs known as BIDAP that processed full text and the rest is history.

This lead me (through the Recon Lab) to Data Corporation, for whom I started working as an unpaid consultant in 1966. This also lead to my involvment with CIRC for DIA.

Charlie: what else in the way of Bio info do you desire.

GIERING

Dick – Thanks for your 9/23/95 speedy response to my letter.

I'd be pleased to meet with you on Oct. 20 on your way from S.F. to Monterey. Let me suggest that you join me at my home

1619 SANTA CRUZ AVE.
MENLO PARK

415/322-1101

which is about 45 minutes South of S.F. right along the way to Monterey. Directions are enclosed which may be helpful for your trip.

Let me know about when we can expect you (e.g. a.m., p.m., evening), so that we can have the coffee ready.

Best regards

Jim 20x95

To: History File

From: Charles Bourne

Re: Notes of 20 October 95 Menlo Park Meeting with Dick Giering

Pre-Data Corporation Activity. In 1965, Giering was an Army captain, working at the Defense Intelligence Agency (DIA) as Chief of a Military Capabilities Section, ADP Systems Center, with an assignment to develop a Defense Ground Order of Battle System for the Army (an inventory of enemy ground forces--numbers, strength, location--for a designated geographic area), making use of data processing equipment and the Formatted File System (FFS). He had, received an early training in data processing, including a degree from the Systems Engineering Department of the University of Arizona.

The use of FFS required that a large number of codes would have to be used in the operation of the system (e.g. a Russian Tiger tank would have to be entered from an authority code list as a code such as V-127). When the design was finished and published, Giering concluded that such a system that operationally required large code books and authority files would not work in field use, and that what was needed was a means to search the text of situation reports ("sit-reps") and action reports from the front. He looked for, and located information about a research project at Northwestern University that was sponsored by the Recon Laboratory at WPAFB, that had developed a set of programs for the IBM 7094 to do batch searching of text that was stored on computer tape. He called the WPAFB project monitors requested more information about the project, and found that the project had been completed, and that the programs (BIDAP--Bibliographic Data Processor) had been filed away. He arranged to get a copy of the programs out of the warehouse and then started to experiment with them on an IBM 7094 computer at DIA.

Giering's first experiment was with airborne photo reconnaissance analysis reports. These reports were essentially narrative comments by a photo interpreter about a single photo image (e.g. "there is work underway to lengthen the runway."), and each photo image had its own separate printed report. Intelligence analysts regularly received large stacks of those printed reports for review. Giering proposed to replace that practice with on-demand searches of the narrative text itself. Using sample queries for topics of current interests, he demonstrated that approach to one of the Army intelligence analysts at DIA, and found immediate acceptance of the idea. And in one of those accidents of historical coincidence, the Air Force intelligence analyst at the adjacent desk overheard the conversation, jumped in, and arranged for this approach to be used with some current Air Force work at WPAFB. That work was the Recon Laboratory activity. While still on active duty with DIA, Giering worked to help Data Corporation install the BIDAP programs on the RECON CENTRAL computer facility for use by Air Force personnel. It turned out that DIA was coincidentally funding the CIRC/COLEX/CIRCOL activity at WPAFB, and in conjunction with that activity, had installed a computer terminal at DIA for use of the CIRCOL system. Giering made use of this CIRCOL terminal at DIA to work with the BIDAP programs that were now being operated at RECON CENTRAL.

The SDC CIRCOL system, and the beginning of the Data Corporation system crossed paths at this time, primarily because two people in different branches of the Armed Services happened to have their desks located together in one government office building in Washington. The Air Force and civilian project personnel at WPAFB associated with the SDC CIRCOL and Data Corporation DATA CENTRAL activity knew of each other's existence, but not the details.

Initial Data Corporation Activities

Giering retired from the Army in early 1967 and accepted a position with Data Corporation later that year, working for Peter Vann (Vice President, and Eastern Representative) with an assignment to build a regional office (computer facility, optics work) in the Washington, D.C. area. They located office space in Arlington, Virginia (across the Potamac River from Washington, D.C.) on the second floor above a large metalwork shop. This was a shop that worked with large metal pieces, and used large stamping presses and other heavy equipment. Giering had an arrangement whereby he received advance notice of any scheduled work to do heavy stamping, or cutting of I-beams, so that he could shut down all the computers to avoid problems due to the shaking and vibration of the building and everything in it!

Data Corporation had been running the RECON CENTRAL library as a contractor-run facility for several years before Giering had his first contact with them. Their library collection contained no image collections to begin with; those were added later. They did have large collections of specification and descriptions of cameras and associated equipment, used to help in Air Force make-versus-buy decisions about camera systems. The numeric search capability in this search system was desired for use in searching the specifications data (e.g. lens descriptions, film speeds), not for searching geographic coordinates of images. Gorog had been using some small IBM equipment at the RECON CENTRAL facility, and was talking with Air Force officials about a possible upgrade to one of the newer IBM 360 machines that were now starting to become available. He use the Northwestern project results and Giering's experiments with Northwestern's BIDEP text searching programs to argue their utility to the RECON CENTRAL activity. The Air Force officials agreed, and let Data Corporation acquire the IBM 360 as part of the existing facilities contract. Subsequent early work on the Data Corporation system then made use of the BIDAP programs from Northwestern, and the Air Force's IBM 360 computer being operated by Data Corporation as part of a facilities contract.

The BIDAP software was meant for batch processing, and when initially used by Giering it was used with a dedicated machine (i.e. fast batch, with no other jobs or users on the system), with the operator either at the computer console, or at a remote terminal (but not both at the same time). Thus it was online, but not interactive, and not with multiple (parallel) users.

Post-Mead. Giering left MTL in December 1977 to start his own company, Infotech. He installed his newspaper editorial text editing and retrieval system at the Globe during the Christmas season of 1980. This was the start of the major newspaper databases other than the NYTIB.

Don Wilson. CONQUEST (Contextual Query) is essentially an upgrade of the LEADER approach, and now a commercial venture.

Horty. Doesn't know what happened to him. When Aspen didn't do well (financially), the financial backers forced him out in 1971 or 1972, and Larry Berul came in and turned Aspen around. (Larry Berul was later associated with Amicus.)

Harrington. Don't know how to reach him. He separated from Diana McCabe, and was a consultant to Mead. He was also writing mystery novels under another name.

KWIC and Highlighting. At the same time that they introduced highlighting for color terminals, they also developed equivalent schemes for monochrome terminals (blinking characters, underlining, special characters before and after the term). For printing terminals that could not backspace for underlining, they used inequality signs as arrows (>TERM<) to show the search term.

Carrier Corporation. They sold the Data Corporation software to the Carrier Corporation in 1973 or 1974, but the installation was a failure because the operating instructions for the computer people (not the searchers) were not well documented.

In the small-world department, this was the same crew at Carrier Corporation that later provided the computer support to the initial BRS service.

Mead Financials (Chapter 10, page 36). LEXIS may have turned a profit in 1977 in 4 years after it started, but that was 7 years after the launch of the OBAR service. In 1969, when Mead made the investment in Data Corporation, profitability of the online service bureau activity and software sales was projected to be in 3-4 years (i.e. 1973-74).

Data Corporation Start. Don't know when Data Corporation started. He suggested that we ask Bill Gorog or Don Wilson. Gorog was with a magazine publisher 8-10 years ago.

Alternate Address. Giering is fully retired now. For part of the year, he stays at their condo in Florida:

Dick & Carol Giering
2866 N.E. 30th Street, Apt. 16
Ft. Lauderdale, FL 33306
(305/566-2238)

xc: Trudi

-1. Big Bang Beginning: What was the genesis of Recon Central computer/APP work? Who was the driving force for getting it started? Other contractors?

20 Oct 95 Giving Mtg. at [illegible]

1. What was the first computer system in Recon Central?
   - IBM ___ equip.
   - owned by ___ (A.F. vs. Data Corp.)
   - batch vs. online, using ___ record system.
   - developed by ___
   - R&D vs operational objective / use
   - when developed ___
   - when put into production use ___
   - index to an image collection vs. biblio records
   - sponsored by ___ WPAFB/DIA/___

2. When was Data Corp. founded? — Gilvy Bill Gorog. or Van Wilson
   was with a magazine publishing business 8-10 yr. ago.

3. Background of 1964 contract to explore [illegible] feasibility?
   - initiated by AF / Data Corp.
   - " " Gorog / AF person ___ / other ___.
   - What triggered the suggestion?
   - Project LITE (1962+ Univ. Pitt. / Horty / A.F. Fairview connection)

4. When did Data Corp. first operate as a service?
   Some pre-CRAR things. Late 67, early 68 as service.
   Rich Cooper worked for Giving on the APP project mgt.
   APP was production 1968 to 1971. They moved to DODAS.
   [illegible]

5. When did you stop using the name Data Corp? when is was named MTL.
(Was it changed to MDCI in Aug 68 on letterhead, etc?
started forming MDCI in Dec 1969
merger. finished 70.
Info this ended out in ~~1971~~

6. Were there 2 splits

1. Fall 70 (incorporation of MDCI)
- all ~~split~~ info existed between Data Corp & MDCI
(subsystem facility, info on activities of RECON CENTRAL
contracts, Union Carbide contract

2. Spl. 1971 split. - Bennett assumed duties.

7. What happened to Hardy?

8. Confirm that MTL name started in March 1972.
This was only a name change for Data Corp.?

Chpt 8. Pg. 36. Mead Personnel
Was this for Mead Corp (— employees) or MDCI (— employees)?

9. ATHOMICS CENTER — despite as a competitive service bureau.
Did they change the other Federal agencies? frequency timesharing.
" " name any non-govt. clients?

10. What is the CONQUEST Interlink query
upgrade of CENTER — now a commercial venture

Dirks attended the 1973 Stanford mtg.

10/20/95
~~10/95~~

Dick Gloring (& Carol)
2866 NE 30th St, Apt 16
Ft Lauderdale, Fl 33306
305-566-2238

Their address when they're not in Texas.

GIERWG

To: History File

From: Charles Bourne

Re: Notes of 17 August 95 call from Peggy Fischer (Office: 203/661-2287. Home: 203/661-5625)

Peg called in response to my recent letter to her.

1. Chester (Chet) Lewis. Chet Lewis was John Rothman's boss. Peg knew Lewis well from her time in New York, and feels that he should get more credit for the NYTIB development. When she was working at Time-Life, she felt that there were 3 really large and notable library reference services: 1) N.Y. Times; 2) Time-Life; 2) Congressional Reference Service; Chet was very involved with this activity, and had contacts with Peg of Time-Life and Content Peckham(Sp?) of LC/CRS? at that time.

   Chet retired from NYT and may still be alive. She'll try to find out.

2. John Rothman. Rothman was a funny guy, narrowly focused, almost secretive (in contrast to Lewis who had no secrets, would talk to anybody, and was an affable manager-type. Rothman had terrible problems with his users (the journalists at NYT) trying to get them to use the terminals. (Peggy talked to Rothman during these times, and contrasted it with her success at Time-Life by having the Chairman on her side, and giving frequent briefings.)

3. Peggy Fischer. Peg started INFORM, the first search brokerage firm. She was also involved in the first computer typesetting system. She left Time-Life in 1971.

   Peg will contact the Time-Life Alumni Assoc. to try to locate some other (Salzburger?) people.

4. Mark Bayer. He was suddenly released recently from the phone company (Ameritech?) in Chicago after Peg did some consulting work for him. His home phone number is 312/642-4442.

5. Mel Day. Mel's wife (Louise Day) still works at NTIS. Mel's home address is 4309 Chesapeake Street, N.W., Washington, DC 20016.

6. Bill Clabby. Is at:
   Wall Street Journal
   1 World Financial Center - 17th Floor
   200 Liberty Street
   New York, NY 10281

   work phone: 212/416-2415
   fax: 212/416-2637

7. Dick Giering. He was working about 3 1/2 years ago as head of programming for a CCH project in Chicago, but he's not there now.

8. Larry Berul. Don't know what happened to him (was with AMICUS). He just dropped out of sight.

xc: NYTIB
Mark Bayer
Mel Day
Bill Clabby
Dick Giering
Larry Berul
Trudi Bellardo

To: History File

From: Charles Bourne

Re: Notes of June 20, 1996 Call from Dick Giering

Dick called to check status on the book. He said that he was going to be in the Ft. Worth area for 2 weeks, and could review something during that time if it would be helpful. I reviewed the status with him, and said that I would send a copy of the most recent Mead chapter to him to review, to see if I incorporated his earlier comments correctly.

COSATI Movie. He hadn't been able to find a copy of the COSATI movie. He suggested that maybe the NTIS or NBS collections might have a copy.

DIA Staff. I asked if he had worked with Dave McCarn at DIA when both of them worked at DIA. He said he had, but only indirectly. He said that when McCarn left DIA to go to NLM, a Navy Captain also retired from DIA and took a job at NLM. Giering knew Ruth Davis at that time, but she worked at the Pentagon, and not at DIA.

IBM-FFS. Dick said that DIA was heavily into formatted file systems when he was there. I asked him about the DIA-IBM relationship with IBM's Formatted File System, and he said that there were really 2 systems. The first was an IBM 1410 system developed by the Navy in Jacksonville, Florida for the Fleet Intelligence Center for the Atlantic Fleet. When DIA was formed, the Navy jumped in quickly, then the Air Force, and finally the Army. The Navy people pushed DIA to take the Jacksonville system because it was already operational. That was the first FFS.

Because DIA had connected to an IBM 7094, a larger machine, they took what they could of the Navy software, and converted, on a piecemeal basis, what pieces they could. That's when they contracted with IBM to build a 7094 version of FFS. Thus the Navy and the Air Force really were responsible for the development of IBM-FFS. This was the beginning of the CCIS (Command and Control Information System) on an IBM 7094 in the basement of the Pentagon.

DIA was heavily into FFS until Dick brought a 7094 text program into DIA. Dick said that that move stepped on a lot of toes there, because text processing was seen by management as a threat to the IBM-FFS. It was after Dick left that they realized that they could have both.

xc: Trudi

DILLARD
FT. WORTH

5 July 96

Dick -

It was good to hear from you again. I'm going to take you up on your offer to review the most current editions of the history draft chapters that I'm guessing that you're most familiar with. You can see that they're still in draft form, awaiting typing and additional update information.

I look forward to hearing from you.

Best regards
Cho Bowen

Encl. SDC & IBM text & supporting cites from chpt. 3 (3/10/96 ed.)
DIA, IBM, & COSATI text " " " 5 (1/16/96 ed.)
chpt. 7 (6/3/96 ed.)
chpt. 8 (12/27/95 ed.)
Mead, QL, Westlaw, Dow Jones, Info Globe, LITE, JURIS, RADCOR & Aspen text & supporting cites from chpt. 10 (3/28/96 ed.)

GIERING
FT. WORTH

8/17/96

Dick –

I just got a card from Carol (give her my thanks) that said that you got my last package, but hadn't received it yet.

Let me make your task much easier. Throw away the Chapter 10 draft sections, & replace them with these cleaner (retyped) editions of the same stuff.

Thanks.

+c Chpt. 10 re
· Mead
· QL
· Wes
· Dow
· Info
· LIT
· JU
· RA
· As
· DDC/DROLS

GIERING

HISTORIC DEERFIELD, MASSACHUSETTS

Lilacs bloom in profusion in Deerfield dooryards in the month of May, when Spring brings green grass and full leaves to the village. The saltbox Allen House, shown here, was built in the mid-1700s and is open to the public on guided tour.

Couldn't find a post-card with the Indian House on it. Hope this is a good substitute. Deerfield is as beautiful as ever!

Dick received a package from you after we left Ft. Worth. Our neighbor has it and he (DICK) will take care of it after we return home around the 5th of Sept.

Hope all is well with you both.

Sincerely,
Carol Giering

Mr. + Mrs. Charles Bon
1619 Santa Cruz Ave
Menlo Park, CA