DCE 18-MAY-71 12:18 6990

Comments on Harvey's DEX design notes

Introduction: 1

Our approach to developing deferred-execution features has had a peculiar history that goes back long before you (Harvey) inherited the unenviable and slightly ambiguous role that you now have. The general goal of deferred-execution features is to increase the utility of our computer aids by in most cases reducing the support cost of computer-aided text manipulation, and in some cases providing more service value to the user than he would obtain from immediate-execution processes. 1A

There should be a smooth spectrum of features applicable to different situations of service level, terminal device, information context and type of task, priority of task, etc. 1A1

Users should find complete consistency and continuity in concepts, nomenclature, and operating skills required for operating effectively in these different situations. 1A2

Users should be able to switch from one level to another in degree of interaction while at an on-line terminal. 1A3

At any degree of computer service and terminal sophistication, the user should have means for deriving maximum utility toward his working goals. 1A4

My undestanding of our current DEX-development state is that we are trying to specify a DEX-1 system that we can implement quickly, and yet that will be reasonbly consistent with the expected trends in later states of DEX-system evolution, in both user features and mode of implementation. 1B

The general trend of the design evolution evidenced in your memos is very good. The pages of detail below don't signify disapproval -- I want to get my head working again on DEX matters, and also I'm trying to get in the swing of this recorded-dialogue working mode that I have been pushing onto everyone else. 1C

I think we're ready soon to produce a proposed user's guide that will give interested parties a look at the user features. Then, relative to an implementation plan, I'd like to see a spec on the proposed process-data organization (Command Table, ring-element flags/pointers, etc.) and the sequence of operations. 1D

cf (6939, 4c2) -- For repeated LNs, I recommend that a

non-numbered instance of an LN designation (in a deferred command) refer to first occurrence rather than last. 2

Consideration -- if LN-alone refers to most recent occurrence 2A

then, when have repeats, 2A1

the same address found in different places in input text may refer to different statements. (which isn't "neat") 2A1A

while, if LN-alone refers to first, the address will be unique to statement referenced no matter where found in input text. 2A2

Cf (6939, 4D1A) -- after LN have GP rather than SP 3

Cf (6939, 4D1A) -- I would like to see us use symbolic names for functional characters -- their assignment is an act separate from specifyng the syntax. e.g.; 4

STABORT rather than '$ (Statement Abort) 4A

STDELIM rather than '! (Statement Delimit) 4B

Note: I feel sort of uneasy about calling this function a "command accept." That may be an unfounded and purely thalamic attitude -- what do you think about it? 4B1

CHDEL rather than '< 4C

WDDEL rather than '> 4D

cf (6936, 4D5) -- I think a LNDEL (line-delete) character would be worth including. It would cause deletion back through first preceding LINEGAP -- i.e., through first preceding GAP containing a LF or CR. 5

cf (6936, 4D5F) 6

If a command statement is to be indicated by the general means you propose, then I suggest, "COMSPEC = '. when '. is first PT in LIT (rather than if it follows first SP)." User would sometimes have ambiguous concern when reading back over text if he had to be sure that it was one and only one SP between LN and '. otherwise. 6A

But I would propose an alternate syntactic convention for a command statement wherein the "command-statement indicator" is

embedded in the PSTRING of the LN, rather than after the following GAP. 6B

The back-delete operations on the LN string will probably occur at a different time than on the literal following the LN -- at least for insertion statements. 6B1

Suppose a user had typed other than a PERIOD as his first PCHAR literal, after the LN of what he intended to be a command statement: 6B2

He would be tempted to do a back-delete to remove the unwanted PCHAR(s) and to leave a PERIOD as the first PCHAR of the remaining text. 6B2A

This would be an awkward provision to implement, since the DEX system would have to do a complete back-delete execution of each insertion statement to determine whether it indeed might end up (in its delete-executed form) being a command statement. 6B2B

Consider the following syntatic convention: a command statemnt is indicated by having a COLON (or else a SEMICOLON, or SLASH, or etc.) be the terminating PCHAR of the LN. Here are some possible command statments: 6C

2a3:d 6C1

Delete Branch 2a3. 6C1A

4b:ud 6C2

"Undelete" Branch 4b 6C2A

2.5:m 4c 6C3

Move Branch 4c to become Branch 2.5 6C3A

Note that we could have eiher of two conventions here -- the other being that Branch 2.5 is to be moved to become Branch 4c. I'd pick the first-stated convention for consistency's sake, so that the user may scan up his text and see for each Copied or Moved branch a left-margin LN indicating where the new entry will be -- which is consistent with the way his newly inserted statements are marked. 6C3B

4c:um 6C4

"Unmove" Branch 4c -- i.e. wherever it may have been told to be moved to in preceding commands, bring it back to this position. 6C4A

Note that we may want also to let the user accomplish this by writing: 6C4B

4c:m 2.5 6C4B1

2.5:c 4c 6C5

A copy of Branch 4c is to become Branch 2.5 6C5A

3b7:a Once the bit is set, ... 6C6

Append to Statement 3b7 the string " Once the bit is set, ..." 6C6A

An alternative convention could be to write simply: 6C6B

3b7: Once the bit is set, ... 6C6B1

There is a bit of a problem in the user knowing just how to control the number of intervening SPACEs he may want between the last PCHAR of the referenced statement and the first PCHAR of his appending literal. 6C6C

One convention to consider is that all of the text string following the ":" be appended. 6C6D

But this makes it hard to proofread back over your input text, if for instance you wanted to have three intervening Spaces, because it is often hard to tell the difference (for sure) between a three- and a four-Space gap. 6C6D1

Another convention to consider is that all of the text string following the ":" be appended. 6C6E

And, for instance, what if he wants no intervening spaces, he'd like for the convention to be that the GCHAR before his LIT be omitted; 6C6E1

But, suppose his appended text began "m 3a2 was not ..." 6C6E1A

Then the command parser would think that the LN:m

signified a move command rather than an Append Command. 6C6E1B

But, if he wants several Spaces, then he'd like his leading GAP to be appended literally. 6C6E2

Could get around thdilemma by using a different character to indicate an Append Command; For instance:. 6C6F

3b7/ Once the bit is set, ... 6C6F1

(Where, unambiguously, everything following the slash is appended immediately following the last PCHAR of Statement 3b7) 6C6F1A

The command parser can operate still upon only the first Visible String, editing it and then "compiling the statement." 6C6F1B

cf (6949, 1B) The only consistent interpretation of (6936, 4D5C) would be that when executing a delete cluster (i.e., after parsing it, when actually moving left and gobbling up text), recognition of all special characters is suspended and they are all treated just as any other Printing Character is. (See below) 7

We probably want to allow delete characters in the LN of either an insert or a command statement, but otherwise make the LN of a statement invulnerable to delete action by any succeeding cluster 7A

cf (6936, 4D5B) and 4D5C, regarding special conventions for executing a delete cluster 8

In the first place, the user must assume that the computer scans the string from right to left, so that the latest cluster is analyzed and executed first -- he wants always to be able to correct what he has typed if at all possible. 8A

During execution of a cluster, by deleting back (left) through visible strings, printing character (PTs), and invisible strings, we should have all PTs treated alike. 8B

So then any special character (', !, $, <, >, +, /) will be gobbled up just as any other PT, and a delete cluster would be just another (or part of another) visible string. 8B1

Thus the note in (6949, 1B) is appropriate. 8B1A

We will need a convention for how far back up the input text string a back-delete directivee buried in a LN can be effective. 8C

A user will have a terrible temptation at times to try deleting through the intervening gap to the end of the preceding statement, throuh the STDELIM (CA) or ABORT symbols, in order to append something he just remembered about. 8C1

We will also need a convention for handling any visible String(s) that isn't a LN but is found between a statement-termination point and a succeeding LN. 8D

General comment: 9

It is generally wiser to minimize the number of characters which are "special" and which must be preceded by the LITFLAG if they are to be treated literally. 9A

To remove the "$" from this special category, why not treat the DOLLAR symbol as an ABORTFLAG only when it immediately precedes the STDELIM (CA) at the end of a statement? 9A1

cf (6965, 2b) -- note 7c, 7d and 8 above. 10

cf (6965, 2c2) 11

I think that, in later versions of DEX, just after delimiting a statement, we'd probably want to do: 11A

If a command statement, do backup-delete operation and then compile the appropriate entries in the Command Table and (if relevant) add appropriate data to the Ring Structure. 11A1

For an Append Command, don't do the backup-delete operation over the appending literal at this time. 11A1A

If an Insert or Append Command, save the literal strings until after the appends have all been done, and the composite statements are each bundled together -- then do the backup-delete operation over the whole composite statement. 11A2

There is good reason to consider a special operation at this point: scanning this input text and replacing all

special flag characters with internally coded characters which are unique and won't be found in any "old text" which this input material may be appended to. 11A2A

This is in case one appends "new" text to an "old-text" statement in an "old file," where such symbols as "<" and ">" won't be protected by LITFLAGS, but where an attractive value would be obtained if the embedded edit directives in the new text were effective over the old text. 11A2A1

We'd either have to use this approach, of converting the characters during input processing, or else find some way of differentiating between old and new text during the subsequent statement-editing pass. 11A2A2

Changing to standard internal codes during the input process has an advantage in accommodating individualized user choices for the different flag characters; they would all be converted to a standard internal set of codes which subsequently would be processed in the same way no matter what set (or sets) of flag characters had been used in the input text. 11A2A3

Note: I think we'll want other editing operations, beside the backup deletion, to be effected by embedded directives. I assume that all such embedded directives are isolated and executed during one single backward scan over the text. (The syntax and functions of such directives will warrant a serious study.) 11B

cf (6965, 2g) Command Tables 12

I have made trial designs where all of the multi-pass bookeeping was to be done within an extended-space arrangement in the ring elements, so I know that a DEX system can be implemented without using a Command Table. 12A

But there is one particular consideration that leads me currently to favor using at least a minimal amount of central, process-control information (a Command Table of sorts) for a DEX processor: 12B

I would like to see the DEX processor be able to handle a deferred-execution job efficiently whether it contains only one command or contains several days worth of stacked-up input. 12B1

If all of the execution-control information is embedded within the ring structure, none of the various ways of doing the multi-pass operation (tantamount to compile and execute) avoids scanning over the whole ring structure during one or two of the later passes, tidying up the structure, throwing ot dummy statements, editing statement strings, etc. 12B2

When doing a DEX process on a small number of deferred commands on a big file, this would involve a lot of extra buzzing around (calling needlessly on many pages) in the ring structure. At least keeping a table of the ring elements that need processing would seem efficient, which I guess could be called a start of a Command Table. 12B3

Then, if there is a table entry for each relevant ring element, an entry of only transient existence, it would seems sensible to store therein any special, transient information in the way of flags, identifier codes, pointers, etc. that would otherwise require special reconfiguration of the ring elements to contain. 12C

Some data, though, may have good cause to be put into the ring elements anyway; for instance, a flag indicating whether or not an element corressponds to an interpolative location (to save referencing the commannd table for every element in the path to the given LN's position). The interpolative field can well enough be stored in the table, because it would be less-often accessed than the flag, and because it would be two or three bytes in size, rather than one bit. 12C1

Besides the Interpolative-Elment Flag, might want another "DEX bit" stored in the ring element, indicating that this element is listed in the DEX Command Table. Perhaps also want a pointer back to the corresponding table entry. 12C1A

I know that one can soon get very involved with flags to say that this branch has been moved or deleted, or is to be filled by a move from elsewhere, or is part of a compound-move chain, or is a dummy statement, or its text is to be appended to from elsewhere, or its text is to be appended to another, or it is part of a chain of appended strings, or etc. Also, special pointers are involved in almost any implementation scheme of any fexibility. For some data elements and some execution-sequence approaches one is led toward storing the

data in the ring elements (as above), and sometimes in the table entries. 12D

So, without getting involved beyond the useful point in specific implementation plans, my comment about the approach for DEX-1 is that I'd guess a mix between ring-storage and table-storage of intermediate, transient DEX-execution data would be sensible. 12E

cf (6965, 2g3) Error printouts: 13

We usually found that users didn't make much use of detailed run-time error printouts, and that it did't take very much in the way of clues in their resulting printout for them to guess at what went wrong. A little more help would be nice. We had just begun to have th system append garbagy things (that it couldn't interpret) at the end of the file where they could easily be thrown out on a subsequent DEX cycle. This was very useful, and I'd like to see further evolution in this direction. 13A

cf (6965, 2h) Execution-sequence precedence: 14

In my past experience, this has proven to be quite sensitive to the configuration of details in user-feature design. 14A

For instance, if you want to edit old text by appending new text to it and then executing edit directives in the new text, this dictates deferring the text-editing pass until all of the appending/merging is done. 14A1

Or, if you are allowing the user to continue referencing a branch by its old LN after it has been (directed to be) moved to a new location, which is a useful feature, then the sequence of operations during execution is constrained -- you can't throw away the corresspondence data until after all of the commands have been at least initially processed.. 14A2

I expect that you'll be juggling this sequence as long as you keep experimenting with different twists of user feature and process-data storage. 14B

Sequence numbering for Command Statements 15

I have at various times considered the possiibility that the user gives each command a unique "Command Number" so that: 15A

The command could be explicitly referenced for subsequent, direct alteration (instead of a sort of "indirect" alteration by "undoing" it); 15A1

And the execution sequence could be controlled by the sequence numbering rather than the order of input, which gives a good deal of explicit flexibility for a user's altering his "text-editing program" after many commands have already been written. 15A2

There would also be the benefit that his command list could be printed out in clean form and order, so the user could either get a better intermediate view during his spooling stage, or so that he could have a debugging check after he gets his new-text output. 15A3

This causes him to type in extra information, and I have never gotten so far into the analysis of value derived to resolve in my own mind whether it is worthwhile. 15B

DCE 18-MAY-71 12:18 6990

Comments on Harvey's DEX design notes

<JOURNAL>6990.NLS;1, 18-MAY-71 12:19 DCE ; (Expedite) Title: Author(s): Douglas C. Engelbart/DCE; Distribution: James C. Norton, Richard W. Watson, William H. Paxton, Charles H. Irby, William S. Duvall, Harvey G. Lehtman/JCN RWW WHP CHI WSD HGL; Clerk: DCE;
Origin: <ENGELBART>NBDEX.NLS;14, 18-MAY-71 12:14 DCE ;

BLP 18-MAY-71 14:05 6991

Comments on 6356 -- Baseline Record System

Comments on 6356 -- Baseline Record System

This is dialog on the Baseline Record System. It is in reply to (Journal, 6356,) by Bill Duvall which was in reply to (Journal, 6349) by me (Bruce Parsley). 1

re branch 1A1: on identification of people 2

I agree that we ought to have just one system for identifying people and initials looks OK. I just trying to fight depersonalization by THE MACHINES as long as I can -- like until there are some processes that manipulate the BRS. 2A

re branch 1A2: on Buyers and ARC goals 3

I know that list was incomplete. I keep hoping that we'll make a more complete list someday. 3A

re branch 1A3: on Sub-Contractors 4

I've always assumed that Buyers and Sub-Contractors would be linked by back links. 4A

re branch 1A4: on Requirements 5

I disagree. Functional specs mean specification of what functions something is to perform -- not how it is to perform them. How is design specs. Thus Requirements are functional specs. 5A

re branch 1A5: on Costs 6

I agree we ought to start working in this area soon. 6A

re branch 1A6: on Sub-Tasks 7

See (Journal, 6979, 2:gw) for some thoughts of mine on Sub-Contracts and Subtasks. 7A

re branch 1A7: on Task names 8

There currentsly exist well over 100 tasks. I shouldn't think we would want all those groups, especially since they change so fast. 8A

The 28 letters has to do with the tab in fornt of the begin-week doing the right thing. 8B

re branch 1A8: on Dates 9

Comments on 6356 -- Baseline Record System

Dates in the BSR are in there as month and week because it is my contention that nobody can plan more than a week ahead with greater than a week acccuracy. So use the units that fit best. I'm sorry it doesn't fit with the rest of the world's dates. 9A

re branch 1A9: on <who 10

The < is in there to make it look pretty. 10A

BLP 18-MAY-71 14:05 6991

Comments on 6356 -- Baseline Record System

<JOURNAL>6991.NLS;1, 18-MAY-71 14:05 BLP ;Title: Author(s): Bruce L. Parsley/BLP; Distribution: William S. Duvall, James C. Norton, Douglas C. Engelbart/WSD JCN DCE; Keywords: baseline record system; Clerk: BLP;

DCE 18-MAY-71 17:29 6992

Note about pre-implementation DEX-1 specifications

Harvey: A note about (Journal, 6990, 1d). I intended that to indicate what I'd like to see done before implementing DEX-1. For the user-guide part, you might as well write it up as well now as you eventually would do anyway for DEX-1 users. For the implementation approach, perhaps make an outline that we both assume is just an outline (like what you'd make for yourself before you started coding), and then let's talk that over (like with unrecorded verbal dialogue even) towards agreeing upon what a "recorded planning document" might be like in this case. 1

DCE 18-MAY-71 17:29 6992

Note about pre-implementation DEX-1 specifications

(J6992) 18-MAY-71 17:29; (Expedite) Title: Author(s): Douglas C. Engelbart/DCE; Distribution: Harvey G. Lehtman, James C. Norton/HGL JCN; Clerk: DCE;

DCE 18-MAY-71 17:52 6993

Note re technical custodianship of NIC's Catalog System

Dick: While I am exercising the Dialogue System, and trying to do my bit toward experimenting with recorded dialogue, I'll send this (perhaps superfluous) note:

Regarding the Catalog Management System of programs, format conventions, programming techniques, people procedures, etc. that Jim is ready to take leave of -- I would like for you to take it off his hands firmly and completely by making yourself fully acquainted with all the technical and procedural details. This request is supported by three reasons: (1) I want to be sure that Jim can leave it and be done with it, so I'd like to see it transferred to a "permanent" NIC staffer who is technically qualified (of which there aren't many candidates); (2) it represents such a basic component of NIC that I want to feel that you are in full control of its further improvements with respect to NIC's needs; and (3) I think that this represents a very effective way fo you to increase your familiarity with our augmentation techniques (and problems).

Yours truly, Doug

1

DCE 18-MAY-71 17:52 6993

Note re technical custodianship of NIC's Catalog System

(J6993) 18-MAY-71 17:52;Title: Author(s): Douglas C. Engelbart/DCE; Distribution: Richard W. Watson, James C. Norton, William S. Duvall. Jeanne B. North/RWW JCN WSD JBN; Clerk: DCE;

NIC Catalog Procedures
19 MAY 71 2:38PM

NIC Catalog Procedures
19 MAY 71 2:38PM

THE NIC CATALOG 1

Items of information relevant to the Network appear in many forms, including technical reports, RFC's, brief network memos, journal articles, and letters. Reference to these is simplified by assignment of an ARC catalog number to each. To record the item to which this number refers, a description of each item, using a set of standard data elements, i.e, author, title, etc., is coded by ARC and entered as as statement in an NLS file. 1A

An example of a statement with typical coded data elements: 1B

*a1 James E. White #2 org *b2 University of California at Santa Barbara #3 Computer Research Laboratory #5 Santa Barbara, California *c1 An NCP for the ARPA Network #6 142p. *d1 21 December 1970 *f1 r *f2 o *r1 UCSB CRL 12 *s1 ARPA #6 AF 19628-70-C-0314 *w2 3-11-71 *y1 Describes program designed and implemented at Santa Barbara node of ARPA Network, written in assembly language and implemented on 360/75. Discusses interface with hardware, software, and operator. *y3 Host-Interface protocol; Host-IMP protocol; User-NCP protocol; Host-Host (A5480) protocol; Host-IMP messages, IMP-Host messages; *z1 all *z2 NIC *z3 new * 1B1

The group of files at ARC containing these statements of data about NIC items and other informational items is the ARC Master Catalog. The term NIC Catalog refers to the NLS file created by collecting the statements coded *z2 NIC in the Master Catalog. 1C

NIC CATALOG LISTINGS AND INDICES 2

Programs have been written at ARC to collect, sort, analyze and reformat the statements and the data elements in the statements to produce catalogs and indexes such as those in the Current Catalog of the NIC Collection, NIC (5145.). The Current Catalog of the NIC Collection is a functional document. It has as its contents, at any time, the current issue of bibliographies (called listings) and author and titleword indexes. 2A

Examples of entries in Catalog Listings and Indices: 2B

Catalog Listing by Author: 2B1

An NCP for the ARPA Network 5480 White

James E. White (University of
California at Santa Barbara)

21 December 1970

Describes program designed and implemented
at node of ARPA Network, written in
assembly language and implemented on
360/75. Discusses interface with
hardware, software, and operator. 2B1A

Catalog Listing by NIC number: 2B2

An NCP for the ARPA Network 5480 ←

James E. White (University of
California at Santa Barbara)

21 December 1970

Describes program designed and implemented
at node of ARPA Network, written in
assembly language and implemented on
360/75. Discusses interface with
hardware, software, and operator. 2B2A

Author Index: 2B3

Subject: Education Status, memo to	5465	Westlund
NWG/RFC 78 (NCP Status Report	5199	White
An NCP for the ARPA Network	5480	White

2B3A

Titleword Index: 2B4

Natural Communication with Computers	5639	Natural
Proposal for a Network Interchange	4752	Network
An NCP for the ARPA Network	5480	Network

2B4A

Index by Number: 2B5

NWG/RFC 79 (Logger Protocol Error), Memo	5601 ←
NWG/RFC 77 (Network Meeting Report), Memo	5604 ←
/re:replacement of Sher survey response/	5605
Short-Range Services and Experiments Using	5606 ←

2B5A

NIC Catalog Procedures
19 MAY 71 2:38PM

NIC CATALOG PROCEDURES 3

Files and programs: 3A

ARC master catalog files: 3A1

Citations are stored in groups of 1000 statements under user CATALOG: (catalog,m3xxx,:Dxnnz) thru (catalog,m6xxx,:Dxnhz) 3A1A

Many of the earlier citations are in an older or partly-converted format, although all NIC citations have been converted to the present data-element format. 3A1A1

NIC subcollection file(s): 3A2

These are stored under user NIC: (nic,nic1s1,:Dxnhz) 3A2A

NIC analyzing, sorting, formatting programs: 3A3

Currently, all NIC programs are contained in: (nic,prog,:Dzxn) 3A3A

It contains first level branches (PROGRAMS) such as: 3A3B

PROGRAM getnic % gets NIC citations from Mnxxx file set% 3A3B1

PROGRAM getrfc % by RWW: gets RFC citations from Mnxxx file set % 3A3B2

PROGRAM fmtrfc % By RWW: this program formats the rfc file into oneline citations% 3A3B3

PROGRAM getifall % gets citations from Mnxxx file set sent to all cites% 3A3B4

PROGRAM authorkey % authorkey with numberkey % 3A3B5

PROGRAM numberkey % simple document numberkey % 3A3B6

PROGRAM titleword %title/subtitlewordkey and numberkey% 3A3B7

PROGRAM formatter % Listing format - by number % 3A3B8

PROGRAM formatter % Listing format- by author + number % 3A3B9

PROGRAM formatter % Index format - by number + author or title word % 3A3B10

PROGRAM formatter % Index format - by number with some title% 3A3B11

Steps to produce the NIC subcollection , listings and indices:

3B

To produce NIC subcollection file(s): 3B1

Enter TNLS, connect to NIC directory. 3B1A

Load file PROG.nls;#, 3B1B

Position CM at :1 (getnic?). 3B1C

Viewspecs: Diw, 3B1D

Go to L10, to compile the NIC citation finding program. 3B1E

Execute collector sorter: 'e 'x CA 3B1F

Provide the input file names: 'f CR <catalog>m3xxx CR <catalog>m4xxx CR <catalog>m5xxx CR <catalog>m6xxx CA, 3B1F1

Provide an output prefix: 'o nicls CA 3B1F2

Change viewspecs: 'v i CA 3B1F3

Command Go: 'g CA 3B1F4

Upon completion, you will get an *, and there will be a newly created nic subcollection file: nicls1 3B1G

Change viewspecs: 'v j CA 3B1G1

To produce author, titleword, number sorts, load file PROG.nls;#, select the appropriate branch, open viewspecs (DW), compile, and use the collector sorter to produce the

file(s) desired, adding commands Sort Yes: 's 'y and Delete Keys Yes: 'd 'y before Go. 3B2

Author listings, for example, are made by compiling the authorkey program first, using the collector-sorter to create interim author-sorted citation files, then the author listing formatting program on those files with the collector sorter again. When this is done, the only remaining tasks are assimilation of statement:0 from a previous author listing file (nic,5802,), appending its directives to this file's statement:0, and outputting to the printer. 3B2A

The titleword sort program returns many trivial words that must be deleted by hand before formatting and output to the printer. 3B2B

We hope to use a trivial-word table for such exclusions soon, greatly aiding the index-building process. 3B2B1

Future automatic processes needed: 3C

We need a subsystem for the clerical process of creating the official NIC Listings and Indices. 3C1

It might have commands such as: 3C1A

Create NIC subcollection: 'n 's CA 3C1A1

Create Author Index: 'a 'i CA 3C1A2

Create Author Listing: 'a 'l CA 3C1A3

Create Number Index: 'n 'i CA 3C1A4

Create Number Listing: 'n 'l CA 3C1A5

Create Titleword Index: 't 'i CA (uses trivial word table) 3C1A6

Create all NIC Listings and Indices: 'a 'n CA 3C1A7

The process would be greatly aided if it could run as a background job with some status feedback at reasonable intervals. 3C1B

The end result of the process would be: on-line NIC files and printer-formatted files ready for on-line viewing or printing. 3C1C

The above subsystem might be easily made into a more general system for producing Journal/AFIPS/NAS and other such catalog listings and indices. The reason for having a NIC-oriented clerical-aid subsystem is to simplify and thereby control the NIC production process as soon as possible. 3C2

JCN 19-MAY-71 11:46 6995

NIC Catalog Procedures
19 MAY 71 2:38PM

<JOURNAL>6995.NLS;1, 19-MAY-71 11:47 JCN ; (Expedite) Title: Author(s): James C. Norton/JCN; Distribution: Richard W. Watson, Douglas C. Engelbart, Jeanne B. North, Marilyn F. Auerbach/RWW (This should be considered as only a start. I intend to make a more complete guide soon) DCE (For your information. This is a start, but needs more detail and clarity, I think) JBN MFA; Keywords: ; Clerk: JCN;
Origin: <NORTON>NICPROC.NLS;6, 19-MAY-71 11:35 JCN ;

DVN 19-MAY-71 14:06 6996

baseline record of ARPA proposal

n

DVN 19-MAY-71 14:06 6996

baseline record of ARPA proposal

Following chats with JCN on 5/18 and 5/19, I have entered a brach in the baseline record describing the work need for the proposal to ARPA due July 1. 1

DVN 19-MAY-71 14:06 6996

baseline record of ARPA proposal

(J6996) 19-MAY-71 14:06; (Expedite) Title: Author(s): Dirk H. Van Nouhuys/DVN; Distribution: James C. Norton, Douglas C. Engelbart/JCN DCE; Keywords: proposal baseline; Clerk: DVN;

DVN 19-MAY-71 14:56 6997

NLS Status File and Rewriting thereof

This replaces the copy of my initial file you received earlier

DVN 19-MAY-71 14:56 6997

NLS Status File and Rewriting thereof

WHP and I had an offline conversation following our dalogue (journal,6971,)(journal,6973,)(journal,6977,) /If two people writing in the same space about the same thing without regrd to one another may be called a dailoge/ aboout letting poeple know what works in NLS from day to day. We agreed that he would keep up the announcement file (NLS,Status) as described in his (journal,6973,), that I would read it from time to time, and when that I felt an entry to be comprehensible only to a trivial fraction of those who need the infomation, rewrite I would it. I will add a branch a branch beginning: "In other words". 1

We hope others will try to keep up that file. 2

3

4

DVN 19-MAY-71 14:56 6997

NLS Status File and Rewriting thereof

<JOURNAL>6997.NLS;1, 19-MAY-71 14:58 DVN ;Title: Author(s): Dirk H. Van Nouhuys/DVN; Distribution: Marilyn F. Auerbach, Barbara E. Row, Beauregard A. Hardeman, Fred P. Hocker, Ed K. Van De Riet, James C. Norton, Jeanne B. North, James C. Norton, Kenneth E. Victor, Bruce L. Parsley, Richard W. Watson, Douglas C. Engelbart, Mil Jernigan, J. D. Hopper, Walter L. Bass, Dirk H. Van Nouhuys/MFA BER BAH FPH EKV JCN JBN JCN KEV BLP RWW DCE MEJ JDH WLB DVN; Keywords: ; Clerk: DVN; Origin: <VANNOUHUYS>JOURPAXDRAFT.NLS;3, 19-MAY-71 10:00 DVN ;

Discussion Log: DCE with HGL on DEX-1 design

Topic of imple. DEX, w.r. to possibility for altering the ring-element design. 1

Considerations through NLS architecture, matters of efficiency, etc. make this a good example of business for role of NLS-Architecture Coordinator -- principles and practices of NLS architecture. 1A

Q: Is current role structure contemplated for software people satisfactoryfor handling this particular case? 1B

How do considerations of efficiency come into these role responsibilities? 1C

User's want value with balanced costs -- the architecture should respond completely to this user need, but in an over-all sense where in a balanced way such factors as reliablity, implementation cost (which the user must bear), maintenance cost (ditto) are all brought into the picture. 1C1

Complexity of "logic" in the DEX system, and in the multi-level considerations involved in the user's end of the design, raised another kind of a role picture. 2

I hypothesize that this is the most sophisticated and complicated subsystem design we have in ARC, with respect to user-interface and "design-logic" issues. 2A

ARC's designers are generally oriented to simpler subsystems (simpler in these issues of design). 2B

Need very much to have role of User-feature-architect Coordinator at work on these issues. 2C

DEX-feature agreements: 3

Do the append command "2ab:a LIT", where the first space after the LN belongs to the LN string, and all the rest is part of the LIT. Agreed that the user wishing for more positive surety can do such as "2ab:a <<, and if ..." where he wanted to be sure to repace a trailing PERIOD with a COMMA SP TEXTSTRING. 3A

I.e., establishes a specific gap in the lead of his LIT, then explicitly deletes that (with one CHDEL) and the period (with a second CHDEL) so he knows where his remaining LIT will start. 3A1

Discussion Log: DCE with HGL on DEX-1 design

Follow an all-Command-Table implementation approach for DEX-1, avoiding any use of ring-element DEX-processing data until there can be more discussion with Charles. 3B

Seems reasonable perhaps to propose having some "scratch area" in the NLS ring element, for a process to be free use temporarily, which for no process is assumed as "safe storage" between the times that it (the process) is active on a file. For instance, would seem that not-insignificant savings in page-swapping, when processing large files, could be had if the DEX processor could store in the ring elements a one-bit Interpolation-Element Flag, and an n-bit Command-table Pointer (pointing into a corressponding location in the Command Table if it is involved in a certain way in the DEX operation). 3B1

Discussion Log: DCE with HGL on DEX-1 design

<JOURNAL>6998.NLS;1, 19-MAY-71 20:02 DCE ;Title: Author(s): Douglas C. Engelbart/DCE; Distribution: Harvey G. Lehtman/HGL; Clerk: DCE;

DCE 19-MAY-71 20:20 7000

Phone log: RWW to DCE from NWG meeting at SJCC in Atlantic City

IPT needs our next-period proposal by July 1 (at least the money-budgeting part, I guess) 1

Some relevant comments by Larry Roberts: 2

Role of NIC in Net plans of the future -- "depends upon how much value (whether?) it seems to provide" 2A

LR assumes that 50% of our computer resources are dedicatable to NIC support. (Some confusion between percentage of ARC activity and percentage of computer service) 2B

LR has big pressure for achieving significant (visible?) Net activity by July, and NIC is important part of that. 3

Many sites apparently plan to be up by July. 4

Agreed wth Dick that we'll bend all efforts toward getting our resources aligned to maximize (in practical sense) early, visible progress in NIC (which, I assume, is strongly oriented to its on-line services). 5

NWG participants seem on the whole to appreciate NIC developmnts and activities. Some disgruntlement about delay in updating the catalog/index, and some flap when he commented that ultimately the special numbering of RFC's would probably be redundant (he assured them that RFC numbers wouldn't be discontinued unlss it became apparent to all concerned that NIC had other, better ways to handle special sub-collections). 6

DCE 19-MAY-71 20:20 7000

Phone log: RWW to DCE from NWG meeting at SJCC in Atlantic City

<JOURNAL>7000.NLS;1, 19-MAY-71 20:20 DCE ; (Expedite) Title: Author(s): Douglas C. Engelbart/DCE; Distribution: Richard W. Watson, John T. Melvin, Charles H. Irby, James C. Norton, Bruce L. Parsley, William H. Paxton, Ed K. Van De Riet/RWW JTM CHI JCN BLP WHP EKV; Clerk: DCE;

DCE 19-MAY-71 20:48 7001

Visitor log: Brian Scott, Australia, Dave Evan's boss, re. augmentation-system possibilities, Network, etc. there.

Dr. Brian W. Scott, Director, W. D. Scott & Co. Pty. Ltd,, 100 Pacific Highway, N. Sydney, Australia 2060; 1

Telephone 929-0033, Telex AA 21883 1A

Management Consultants 1B

They are currently trying to asses potential and approach regarding Aug-Sys introduction and utlization in Australia. Giving DAE some time to scout out possibilities, along with (also) possibilities for activity in Australia (and in Scott Co.) on Technological Policy activities. 2

Brian is director of the the company's foreign activities (New Zealand, Southeast Asia); did his specialty work in long-range forecasting and planning strategies (MBA at Stanford, PhD at Harvard). 3

Main possibility for mutual work would seem to lie in a confluence of: 4

Network connection seems an important eventual component, they'd like to learn more of ARPA's overseas Net links. 4A

Possibility of CSIRO (? Australia's government supported scientific-research organizaton) getting involved, too, like also being Net participants, and handling the sevice-system part for Scott's participation (and other Australians). 4B

Australian move toward aug-sys exploration, like in actual support of an intelligence system on computer-systems (and/or associated managemet systems, or user systems) area(s) of the Technological Policy work. 4C

They (Scott) would very much like to get involved in dialogue about BC type of formulation, planning, assessment, launching, etc. 5

Brian outlined one possibility for significant utilizaton of Aug Sys: in supporting the studies, analyses, dissemination etc. of development NP and plans for Southeast Asian countries. 6

I told him that such a possibility corresponded closely to long-range dreams of mine and provided fundamental motivation for me -- but, that my current strategy strongly leaned toward refusing a direct involvement in such an application endeavor. Instead, I'd argue strongly that those who had

Visitor log: Brian Scott, Australia, Dave Evan's boss, re. augmentation-system possibilities, Network, etc. there.

resource-allocation interest in such directions would be better off in the long run to invest resouces first in BC, to provide development and shakedown experience of techniques that would be largely applicable, and simultaneously be training many levels of systems people who could in turn attack the problems of implementing such systems. 6A

DCE 19-MAY-71 20:48 7001

Visitor log: Brian Scott, Australia, Dave Evan's boss, re. augmentation-system possibilities, Network, etc. there.

<JOURNAL>7001.NLS;1, 19-MAY-71 20:48 DCE ; Title: Author(s): Douglas C. Engelbart/DCE; Clerk: DCE;

DCE 19-MAY-71 20:51 7002

Visitor log: Dick Reed

Richard Reed, 1605 Castro, S.F. 94114; 824-2007 1

Back in the area to live. Sort of re-aligning his personal professional endeavor. Currently excited by the "bio-feedback" possibliies as applied in group conferencing. No special prospects discussed. 2

My basic reaction was that he could make a maximum contribution to the effectiveness of group conferencing by beginning within the current "marketplace" in association with people already established effectively and dealing with organizational problems. Fun to think of future collaboration in some of these areas if BC materializes as I hope it might. 3

He visited with Brian Scott, too, during Brian's second stint here of the day. 4

Dick and I to keep in touch. 5

DCE 19-MAY-71 20:51 7002

Visitor log: Dick Reed

<JOURNAL>7002.NLS;1, 19-MAY-71 20:51 DCE ;Title: Author(s): Douglas C. Engelbart/DCE; Clerk: DCE;

Position Notes for proposed FJCC 'Interactive Search' Panel,
D. C. Engelbart, Stanford Research Institutue, 28 April 1971

7003

Mailed to John Bennett, IBM San Jose, on 28 Apr 71

Position Notes for proposed FJCC 'Interactive Search' Panel,
D. C. Engelbart, Stanford Research Institutue, 28 April 1971

The knowledge worker of the future will find many valuable applications of interactive computer "tools". It should be assumed that many tools, for many types of application, will have a chance over the coming years to evolve, prove out, and perhaps make it into the integrated "interactive knowledge workshop" where he will do his serious work. Searching over bibliographic data is but one such candidate application. 1

My position, with respect to the Panel's topic, stems from twenty years of belief in this and of dedication toward developing an "augmented workshop" for knowledge workers. My statements of ten years ago (see Attachment A) are still largely representative of my view, and my work since has been entirely consistent. The outgrowth of this subsequent work is the Augmentation Research Center at Stanford Research Institute (see Attachment B). 2

Our central push is toward augmenting our 25-person system-development team. We already have many tools and methods affecting the individual working in his own problem domain, and our current new-development emphasis is on augmenting the collaborative dialogue among the team members. 3

In this we are treating each communique as a library would treat a document -- only here we have much smaller parameters of communique size and of time taken for cataloging, indexing, circulating, retrieving, etc. Also, we provide for explicit citation right down to a specific character in a given communique, and for automatic, on-line, full-text access to any communique (or to an embedded entity that is cited). It is expected that people engaged in collaborative dialogue will constantly be generating, managing, and studying special sets of communiques (or of passages therefrom). 3A

Immersed in this "Dialogue Support System" (DSS) will be a rich set of tools and techniques for querying the recorded-dialogue data base. This "query system" must be designed in a sensitive, coordinated way with the rest of the tool-technqiue workshop system in which it is embedded. All the tools and techniques are applied within a "working methodology" that differs markedly from that of a non-automated intellectual workshop. 3B

Besides the abovementioned activity, the Augmentation Research Center is also a participant in the experimental ARPA Computer Network. For the community of Network participants, we play a special role by operating a Network Information Center. This is currently an off-line service to distributed sites, where our computer aids help us manage the data base, publish some of the documents, generate shelf lists and special indexes, etc. 4

Position Notes for proposed FJCC 'Interactive Search' Panel,
D. C. Engelbart, Stanford Research Institutue, 28 April 1971

As sites become increasingly "on line" through the Network, we will be providing interactive search service, as well as a message system whose recorded messages will form part of the NIC information base. The message system will evolve towards the equivalent of our Dialogue System as that system stabilizes through local use. 4A

My general "position" on Interactive Search is strongly oriented towards its being but one important tool in an integrated intellectual workshop. For instance, I feel that: most of the significant future interactive searching will take place at consoles designed for general interactive "knowledge work"; the searchers will be skilled at other interactive processes, with which the search process will be fully compatible; the cost of terminals, computers, etc., will be shared among these users ("interactive search" doesn't by itself have to justify these costs); the cost/value considerations related to training interactive users in special knowledge, procedural skills, and operating skills will also be weighed over the whole set of coordinated workshop operations. In general, it won't be up to the interactive-search system developer either to carry the whole ineractive-system development burden or to set the interactive-environment stage (console design, skill levels, etc.). 5

Position Notes for proposed FJCC 'Interactive Search' Panel,
D. C. Engelbart, Stanford Research Institutue, 28 April 1971

<JOURNAL>7003.NLS;1, 20-MAY-71 10:08 DCE ; Title: Author(s): Douglas C. Engelbart/DCE; Clerk: DCE;
Origin: <ENGELBART>POSITION.PAPER;5, 28-APR-71 18:19 DCE ;
.DIR=1; .HED= "Position Notes for proposed FJCC 'Interactive Search' Panel,
D. C. Engelbart, Stanford Research Institutue, 28 April 1971"

.PEL; .PGN=PGN-1; .GCR;

Received at ARC Week Ending 21 May 1971

Books 1

Software Engineering, Report on a Conference Sponsored by the NATO Science Committee, Garmisch, Germany, 7-11 October 1968. January 1969. 1A

Periodicals 2

IBM Computing Report April 1971 2A

NAS News Report April 1971 2B

Datamation May 15, 1971
Contains: Minicomputer Survey, by D. J. Theis and L.C. Hobbs. p.24-34
Also: The Rise and Fall of Viatron, Part 1, by W. David Gardner, p.38-40. 2C

Graphic Science May 1971
Contains: Microfilm and Computer in Engineering, by George H. Harmon, p.14-15, 23
Also: A System for Micro-graphic Reports, by Thomas G. Doran, p.21-23. 2D

Tekscope May 1971 2E

Reproductions Review May 1971
Contains: Microfilm in the 70's: Where Is It Going? by Howard A. Floyd. p.30-31, 67.
Also: A Glossary of Microfilm Terms, prepared by NMA, p.31, 34
Also: Developing an Effective Microfilm Retrieval System,
p.32-33, 51, describing system at North Dakota Blue Cross. 2F

Innovation Newsletter No. 18 May 1971
Contains: News of Technology Management Seminars 2G

Received at ARC Week Ending 21 May 1971

Reports 4

MAC

RFC 134 Network Graphics Meeting NIC 6711
A. Vezza 29 April 1971

RFC 148 Comments on RFC 123 NIC 6751
A. Bhushan 7 May 1971 4A

ILL

RFC 156 Status of the Illinois Site NIC 6761
J. Bouknight 26 April 1971 4B

UCLA

RFC 157 Invitation to Second Symposium on Problems in the Optimization of Data Communication Systems NIC 6762
V. Cerf 12 May 1971 4C

WASHU

Macromodular Computer Systems 6861
Tech Report 4 20 June 1967
Reprints of 7 papers from SJCC, 1967

Analysis of Implementation Errors in Digital Computing Systems 6862
Robert M. Keller and Donald F. Wann
Tech Report 6 March 1968

A Unified Approach to the Design and Restructurable Computer Systems: The Meta Macromodule Machine 6864
Robert A. Ellis Tech Rpt. 7 June 1968

Control of Concurrent Operations in Asynchronous Digital Processes 6863
Fred U. Rosenberger Tech Rpt. 14
July 1970 4D

JBN 20-MAY-71 11:53 7004

Received at ARC Week Ending 21 May 1971

<JOURNAL>7004.NLS;1, 20-MAY-71 11:53 BER ;Title: Author(s): Jeanne B. North/JBN; Distribution: Douglas C. Engelbart, Little Black Book, ARC Black Board, James C. Norton/DCE LBB ABB JCN; Clerk: BER;
Origin: <ROW>NEWARRIVALS.NLS;1, 20-MAY-71 11:06 BER ;

Re-Turning programs over to operations (Journal, 6946,)

Dick... 1

I basically agree with all of the points you made in the memo (Journal, 6946,). 1A

I feel there are some problems which are overlooked, however, and in our paticular environment (which may well change) make it difficult to design and bring up a system in the manner described. 1B

Operations. 1C

In the past, there has not been a firm concept of computer operations. 1C1

Consequently, no distinction has been made between testing systems (with user guinea pigs), and turning over a finished product to operations. 1C2

In order for this distinction to occurr, it is necessary to define an operations group (as is currently being done), and, moreover, firmly establish that: 1C3

(a) Programs turned over to operations are well tested as described. 1C3A

(b) Operations has the subsequent responsibility for maintaining and coordinating changes, improvements, etc. 1C3B

(c) Definite interfaces be established between the service system contractors and operations to allow the development of adequate and unambiguous operating specs. 1C3C

Environment 1D

On order to satisfactorily design a system for a particular operating environment, some description of that environment must be available to the designer. 1D1

The ARC operating environment has been sufficiently nebulous in the past to defy any long range description. 1D2

Moreover, the design of a system which will change the operating environment (which was the case with the journal system, for example) is complicated by second order effects of the designed change in the environment. 1D3

This is, I feel, a sticky problem because I at least am

somewhat hesitant to give up a constantly changing, flexible environment. 1D4

Presumably we may plan to intercept the changing environment in a relevant manner when our planning technique becomes accomplished. 1D4A

Research and continuous evolution 1E

A strict, classical, operations approach to bringing up systems makes continuous evolution difficult. 1E1

The research nature of our work implies that we will wish to bring up systems with acknowledged faults in order to test out certain concepts. 1E2

Such is the case with the response in the Journal system. 1E2A

It was implemented at a high level in order to test and develop NLS machinery for writing Higher Level Processors. 1E2B

That it would be slow was not at all unexpected, but, in the long run, it was felt to be a necessary evil which was to be encountered in the evolution of Higher Level Processes. 1E2C

WSD 20-MAY-71 13:15 7005

Re-Turning programs over to operations (Journal, 6946,)

<JOURNAL>7005.NLS;1, 20-MAY-71 13:15 WSD ; Title: Author(s): William S. Duvall/WSD; Distribution: Richard W. Watson/RWW; Keywords: Program Operation; Clerk: WSD;

JCN 21-MAY-71 16:26 7009

Initial On-line Message Transmission Between RADC-ARC

Initial On-line Message Transmission Between RADC-ARC

To: Jim Bair and Tom Lawrence (RADC) 1

From: Jim Norton (ARC) 2

As we discussed, your introduction to TNLS and the coming NIC Dialog Support System (NICDSS) will be much more effective with your participation in the NIC TNLS course to be held at SRI in Menlo Park, June 16-17. 3

Until the course occurs, you may wish to try sending some messages to us, using commands we have just put into operation. 4

There are presently two message-sending systems: 4A

The Executive subsystem: SNDMSG 4A1

This system permits text messages to be sent to other users (RADC, NORTON, ENGELBART) by creation of or appending to a special file named MESSAGE.TXT;# in the directories of those users to whom messages are sent. 4A1A

To send a message: 4A1B

At the Exec level, type: 's 'n 'd ALT CR 4A1B1

The system will echo: MSG.SAV;3 4A1B1A

On the next line the system will print: 4A1B1B

TYPE LIST OF USERS 4A1B1B1

The user should reply with: 4A1B1C

<username> $(',/SP <username>) CR 4A1B1C1

On the next line the system will print: 4A1B1D

TYPE MESSAGE.EDIT WITH CONTROL-A,Q,R,X
END WITH CONTROL-Z.
INSERT A FILE WITH CONTROL-B 4A1B1D1

The user should reply with: 4A1B1E

<message> CONTROL-Z or 4A1B1E1

<filename> CONTROL-B. 4A1B1E2

(SP=SPACE; CR=Carriage Return; ALT= Control q on the Model 37 TTy) 4A1B2

To receive a message: 4A1C

When you have a message, the system will type "YOU HAVE A MESSAGE" immediately after you login. 4A1C1

To see the message, type: type SP message ALT CR 4A1C2

Unless you want to save the message, you may find it best to delete the file MESSAGE.TXT after viewing. If you do not delete the file, the next message you get will be appended and you will have to print out all prior messages before you get to the new one. 4A1C3

For a while, this message system may be useful for notifying others of the filename of a more substantial message. When journal messages are delivered on-line in a month or two, this system will probably be of less use to us. 4A1D

The ARC Journal, a subsystem of NLS: 4A2

Overview 4A2A

The Journal system is intended to provide a convenient method of entering documents and messages into the ARC Journal, while providing for automatic distribution and cataloging of these documents. 4A2A1

It may be accessed directly from NLS, and appears in the form of a 'submode' with commands. 4A2A2

The term message applies to an existing single statement or literal entry submitted to the Journal. 4A2A3

A branch, group, plex, or file is taken to be a document. 4A2A4

From the sender's viewpoint, messages and documents are handled in the same manner. 4A2A5

The system, however, takes advantage of the distinction in certain effeciency considerations. 4A2A5A

User Information 4A2B

NLS commands 4A2B1

The user enters the Journal submode by typing: 4A2B1A

'e 'j
submit (('s/'b/'g/'p/'f) <address>/(SP/'m)
<literal>) CA
number (CA/<number> CA Assigned to:
<identification> CA) 4A2B1A1

Semantics: 4A2B1A2

Execute Journal Submit 4A2B1A2A

When the user types 'e 'j, NLS will echo "execute journal" and "submit" requesting a structure type of the entity to be submitted. 4A2B1A2A1

In the event that the structure is a statement, branch, group, or plex the user must specify its address. 4A2B1A2A2

In the event that the type is a file (indicated by typing 'f CA), any statement part of an address will be ignored. 4A2B1A2A3

The file you wish to send, or send part of, must be the one you currently have loaded in NLS. 4A2B1A2A3A

In the event that the type is a literal message (indicated by a SP or 'm), a carriage return will be typed, and the user will enter the desired literal (<literal>) using the same control characters he would in insert statement (i.e. BC, BW, etc.) 4A2B1A2A4

Number (CA/<number>CA Assigned to:
<identification> CA) 4A2B1A2B

When the document or message to be entered has been specified, the user is asked for a number. 4A2B1A2B1

If the user has not a pre-assigned number

for the document (which will be the normal case), he types a CA and a number is assigned to the document by the system, and typed out. 4A2B1A2B2

In the event that the user has a preassigned number, it should be entered followed by a CA. In this case, the system asks to whom the number has been assigned. 4A2B1A2B3

Once this procedure is completed, the user is at the Journal sub mode command level indicated by the system typing the herald character '&. 4A2B1B

In the Journal submode, the following commands are allowed: 4A2B1C

'i: Interrogate CA 4A2B1C1

This causes the Journal system to enter a non-passive command mode, whereby the user is interrogated for header information. 4A2B1C1A

The information which will be requested by the system will be: 4A2B1C1B

Author(s): 4A2B1C1B1

Title: 4A2B1C1B2

Distribution 4A2B1C1B3

Expedite 4A2B1C1B4

Keywords: 4A2B1C1B5

Comments: 4A2B1C1B6

Go: 4A2B1C1B7

The user should respond to the prompts as follows: 4A2B1C1C

Author(s): <identlist> CA 4A2B1C1C1

This allows the user to specify the

contents of the suthor field in the header statement. 4A2B1C1C1A

The default value for this field is the identification of the current person using NLS. 4A2B1C1C1B

The normal response is CA 4A2B1C1C1C

<identlist> is defined in (5a2b1d2). 4A2B1C1C1D

Title: <literal CA> 4A2B1C1C2

This allows the user to enter a title which will be used in the hard copy printouts of the document. 4A2B1C1C2A

<literal is defined in (5a2b1d1). 4A2B1C1C2B

Note that the title will be overidden by any .HED directives within the text of the document. 4A2B1C1C2C

Distribution: <identlist> CA 4A2B1C1C3

This command allows the user to specify a list of persons who will receive copies of the document being entered. 4A2B1C1C3A

<identlist> is defined in (5a2b1d2). 4A2B1C1C3B

Expedite ('y/CA/'n/SP/CD) 4A2B1C1C4

This tells the system that distribution of the document is to be expedited. Currently, this implies hardcopy distribution at ARC or mailing within three or four hours. 4A2B1C1C4A

Keywords: $<words> CA 4A2B1C1C5

This allows the user to specify key words which he thinks typify the nature of the document. 4A2B1C1C5A

These words will be included in a special keyword field in the header

which may be used for later content searches of documents. 4A2B1C1C5B

Comments: <literal> CA 4A2B1C1C6

This allows a user to enter a comment about the document being entered. 4A2B1C1C6A

The text of the comment will be the last item in the header statement, and will be preceded by 2 carriage returns. 4A2B1C1C6B

Comments are printed on an unnumbered page preceeding page one of the hardcopy printout. 4A2B1C1C6C

<literal> is defined in (5a2b1d1). 4A2B1C1C6D

Go CA 4A2B1C1C7

This causes the Journal entry process to continue after the user has specified any header information he wishes. To use the Status command, type a CD here, rather than a CA. After status is accepted, type 'g CA to start the process. 4A2B1C1C7A

If he does not wish to respond to a certain command, he may proceed to the next by typing a CD 4A2B1C1D

If he does not wish to 'GO' following the interrogation, a CD will return him to the Journal System command level for further specification. 4A2B1C1E

Instead of using the Interrogate command to cause the system to prompt to the user for each entry item, the user may enter each item independently using the following commands. 4A2B1C1F

'a <identlist> CA 4A2B1C1F1

When the user types "a", the system echoes "Author(s):". The user should

respond as to the Author(s): prompt
under the Interrogate command. ↓A2B1C1F1A

't <literal> CA ↓A2B1C1F2

When the user types "t", the system
echoes "Title:". The user should
respond as to the Title: prompt under
the Interrogate command. ↓A2B1C1F2A

Note that the title will be overidden
by any .HED directives within the text
of the document. ↓A2B1C1F2B

'd <identlist> CA ↓A2B1C1F3

When the user types "d", the system
echoes "Distribution:". The user
should respond as to the Distribution:
prompt under the Interrogate command.
↓A2B1C1F3A

'e ('y/CA/'n/SP/CD) ↓A2B1C1F↓

When the user types "e", the system
echoes "Expedite:". The user should
respond as to the Expedite: prompt
under the Interrogate command. ↓A2B1C1F↓A

'k $<words> CA ↓A2B1C1F5

When the user types "k", the system
echoes "Keywords:". The user should
respond as to the Keywords: prompt
under the Interrogate command. ↓A2B1C1F5A

These words will be included in a
special keyword field in the header
which may be used for later content
searches of documents. ↓A2B1C1F5B

'c <literal> CA ↓A2B1C1F6

When the user types "c", the system
echoes "Comments:". The user should
respond as to the Comments: prompt
under the Interrogate command. ↓A2B1C1F6A

The text of the comment will be the

last thing in the header statement, and will be preceded by 2 carriage returns. 4A2B1C1F6B

'g CA 4A2B1C1F7

When the user types "g", the system echoes "Go". This causes the Journal entry process to continue after the user has specified any header information he wishes. To use the Status command, type a CD here, rather than a CA. After status is accepted, type 'g CA to start the process. 4A2B1C1F7A

's CA 4A2B1C2

When the user types "s", the system echoes "Status". As or after the user enters this header information, he may review it by causing the system to type out the values currently in the header statement. 4A2B1C2A

The fields typed out by status are: 4A2B1C2B

Document number 4A2B1C2B1

Author 4A2B1C2B2

Title 4A2B1C2B3

Distribution 4A2B1C2B4

Expedite 4A2B1C2B5

Keywords 4A2B1C2B6

Clerk 4A2B1C2B7

Comments 4A2B1C2B8

Any fields which are empty are omitted from the typeout. 4A2B1C2C

At this time, the user may change or add to the items in this list by re-entering any command. 4A2B1C2D

'o <identification> CA 4A2B1C4

When the user types "o", the system echoes "operator". This allows the user to specify the identification of the person who is currently doing the entry of the document. 4A2B1C4A

It is assumed (by default) to be the NLS user. 4A2B1C4B

When the Status command is executed, the entry supplied to the Operator command appears as "clerk" is the status list. 4A2B1C4C

'q CA 4A2B1C5

When the user types "q", the system echoes "quit". This command may be used to return to the NLS command mode.
It should not be necessary except in cases where there has been a system error. 4A2B1C5A

4A2B1C5B

Relevant metalinguistic variables 4A2B1D

<literal> ::= $CH CA 4A2B1D1

<identlist> ::= <identification> $((',/SP) <identification>) 4A2B1D2

<identification> ::= (L $LD/CR <newident>)('($CH ')/<empty>)/<empty> 4A2B1D3

An identification is a letter digit sequence assigned to each NLS/Journal/NIC/etc. user. 4A2B1D3A

Current identifiations may be found in file: (journal,identfile,:xrnC) 4A2B1D3A1

In the event that the field is empty, and it is terminated by a CA, the identification of the current NLS user is indicated. 4A2B1D3B

If the field begins with a Carriage return (CR), the system assumes a new ident is being entered, and interrogates the user for the necessary information. 4A2B1D3C

A further discussion of identifications may be found in (journal,6203,0:gw)) 4A2B1D3D

If an identification field is terminated or immediately followed by a '(, the text between the '(and the next ') is taken as a note. 4A2B1D3E

This is useful for making notes about or to particular authors or recipients, e.g. WSD (Please see branch 1a2) 4A2B1D3E1

Noteworthy User Hazard 4A2B2

There is a user hazard regarding directives and the journal worth mentioning here: 4A2B2A

The Journal system does not tamper with any directives in your file other than those using .HED. 4A2B2A1

Caution should be exercised that there are no directives in a file which would contradict those added by the Journal in the header statement. It is recommended that the user delete or edit them so as not to detract from the format of the hard-copy of the Journal document being entered. 4A2B2A2

A convenient way of dispensing with any directives past the origin statement in a file is to include a directive such as .IGDir=1 (ignore subsequent directives) in the comment field of the header (using the 'Comment statement). 4A2B2A3

Distribution 4A2B3

All Journal messages and documents are sent only in hardcopy form at this time. We will add direct on-line delivery soon, probably through addition of a link (or a short message's text) to a "message" branch in the addressee's initial file (TFL, JHB, JCN, DCE named files). 4A2B3A

Since these initial files are automatically presented to each user as he enters NLS, people

will have an easy way to find messages sent to them. 4A2B3B

New messages will be added at the head of the message plex for quicker viewing of the most recent ones. 4A2B3C

When you are through viewing a message, you may delete it (or the statement with the link to it) for it will also be on file under directory JOURNAL in a file with a name such as 7006 (if it is a document) or as a branch in a file named JRNL1 (if it is a short message).

4A2B3D

We should try an initial message transmission soon. I would like to try on Wednesday morning, May 26th, if that is convenient for you. How about 11:00 your time (8:00 ours)? 5

JCN 21-MAY-71 16:26 7009

Initial On-line Message Transmission Between RADC-ARC

<JOURNAL>7009.NLS;1, 21-MAY-71 16:26 JCN ; (Expedite) Title: Author(s): James C. Norton/JCN; Distribution: James H. Bair, Thomas F. Lawrence, Douglas C. Engelbart, Richard W. Watson, Marilyn F. Auerbach, William S. Duvall/JHB TFL DCE RWW MFA WSD; Keywords: ; Clerk: JCN;
Origin: <NORTON>JBAIR.NLS;6, 21-MAY-71 16:04 JCN ; (:zxnD)

NAS/CLR Panel Report: Draft, RLW 2 April 71

7010

Draft developed during Panel session at SRI, 31 Mar and 1 Apr 71

Libraries and Information Technology, A National Systems Challenge

1

INTRODUCTION

2

The "National Information Problem" has been stated in many forms, among which are: 2A

Information storage and retrieval in support of science, technology, and medicine 2A1

Library automation 2A2

Information support for business and government decision making. 2A3

Most studies and plans relating to this problem, or some portion of it, have identified computer and communication technology as the basis for future systems to cope with increasing volumes of information, increasing complexity of interdisciplinary needs, and the quickened time response requirements for information in modern society. Many research studies, development projects, and pilot operations are working on various facets of the Information Problem and progress is being made. 2B

However, all too often the progress is slower than expected or needed, the problems encountered are more complex than originally understood, the costs for development and operations appear to be higher than can be afforded, and in some instances, the presently available technology has seemed to be deficient. 2C

This study represents cooperation between the Library and Computer Science and Engineering communities to examine the application of computers and related technologies to libraries and information systems. The results of this examination are intended to: 2D

assess the adequacy of technology to meet the needs of library and information systems at the national level; 2D1

provide guidance for the application of the results from existing exploratory and pilot projects; 2D2

identify weak areas in the understanding, design and use of computer systems as needed for effective and efficient information systems; 2D3

recommend national targets for improvements in the weak areas; 2D4

provide technical guidance for use by planners, system designers and operating officials concerned with information handling systems. 2D5

The study was performed by the Information Systems Panel of the Computer Science and Engineering Board of the National Academy of Sciences and was supported by the Council on Library Resources. Through site visits, published information, and the experience of the participants in the study, the insights and results from library automation projects, the computer and communication industries, academic information science research, science abstracting and indexing system development and operations, commercial computer science industry, and news processing and publication industry, have been brought into the study. They provide a diversification of the basis for this report. 2E

Some Limitations 2F

The authors of this report accept that it is to the National interest that access to information be provided, at least to the level of convenience and of completeness that has been provided historically by libraries. The need for considerable improvement beyond what has been historically provided has been proclaimed by many. Establishing the validity of the assertion is not undertaken within this report. 2G

(1) Cite reference to the Olsen review of the literature on the economis of information.) 2G1

(2) A "mixed bag" of examples are the National Park system, a Fine Arts museum, and even the Armed Forces of the United States. 2G2

The value to the Nation or to Society of providing access to information, along with the values of many other intangibles and services (2) is not measureable in dollars at the present state of knowledge (1). Consequently, quantitative cost/benefit studies are not part of this report, even though relative cost for equivalent service by different means may be examined, and costs themselves are of primary interest. In no way is the desireability of such measures of value disputed, but in their absence and until such are developed, we must proceed on the basis of qualitative judgement. 2H

(2) Examples 2H1

It is a legal requirement that in providing access to information, property rights and privacy rights of individuals be respected. Study of means of assuring that existent these rights not be contravened or alternative measuring recommended, has not been undertaken within the framework of the report. (3) 2I

(3) (Cite references to privacy studies (Weston) and to copyright discussions -- a recent paper by V. Clapp, others?)

2I1

The emphasis of this study is primarily on the intellectual and technical factors involved in the design and building of computer-based library and information systems and their implications. These factors are affected by many non-technological factors which must be acknowledged in this examination and taken into account in system design and operation; however, it is not the purpose of this study to propose missions and roles of public or private institutions or to propose technical policy outside of the areas of competence of the participants and sponsors. The results are intended to make known the system design principles which must be observed in order to make unified progress in solution of the National Information Problem, to relate the existing and expected state of the relevant technology for this purpose, and to identify key achievements that are necessary. 2J

THE PROBLEM ADDRESSED 3

General Nature of the Problem 3A

Entering the National Information Problem from the library centered view is useful because the library function is the "memory" of national information systems. The current embodiment of this memory is the wide range which includes the large national libraries, major academic and research libraries, federal libraries, industrial libraries, public libraries, and various information collections used in support of industry, government, and education. It has built up over many years, based on the traditions of centuries, oriented mostly toward the acquisition and management of storehouses of knowledge printed on paper, and based on manual processes which require extensive human interpretation of situations and judgment in handling them. 3A1

Thus, for purposes of functional characterization, present libraries will be regarded as the storehouses of recorded knowledge (or dialog) in graphic form to be used over a period of time (delayed information). 3A2

Examining this as a "memory function" provides explicit concentration on libraries problems and their solutions, but more important, it frees the investigation from considering only the mechanization of traditional methods, suggests that the library functions could be extended to cover new forms of information, and clarifies the requirements for the mechanized interfaces to the memory function. These interfaces supply information to the library as a general memory institution and deal with the access and delivery mechanisms which select and supply the information for use. 3A3

The intellectual problems thus must be shared between the traditional and the new technology disciplines. 3A4

Briefly stated, the overall problem is that: 3A5

There is more information than can be digested; 3A5A

Better intellectual* and physical access is needed to it; and 3A5B

* In the library world the methods used are called bibliographic access, the abstracting and indexing community emphasized indexing as an access route, and computer science practitioners think in terms of symbolic addressing as the name of the process for locating items in files. In this work we have frequently encountered varying terminology used by different specializations in discussing the same basic concepts. 3A5B1

The present methods for both have exponentially unfavorable cost and delay time with increasing volume of material handled. 3A5C

The "information technology" that is expected to be the opportunity for alleviation of this problem includes computer processes and compatible information storage media, input/output mechanisms, electrical communications, and various classes of information reproduction that collectively is termed reprography. Computers are the key element in this technology because they provide the control mechanism which can taken actions conditional on predetermined situations and on the representation of the information being handled. However, without the other facets of this technology to connect computers into networks and to provide human compatible interfaces, the full power of computer technology could not be as effectively applied. 3A6

This information technology, if properly applied, has characteristics than can replace repetitive labor intensive operations, freeing personnel for the necessary intellectual functions, and thus expedite the "digestion" rate of new information. It provides "economies of scale", as has been demonstrated in both the computer and communications industries, which can combat the faster than linear growth of library costs with increasing volume of material to handle. Further, the continuing rapid technological improvement and cost reductions per unit of performance in all aspects of electronic technology make it increasingly attractive for replacement of older methods of information transfer and handling, and library institutions must either keep pace with these developments or eventually face replacement by new institutions which provide the necessary services. 3A7

In terms of an automation project, certain characteristics of the problem need to be recognized. As a commodity to be managed, information is unusual. 3A8

Its quantity can be expressed in terms of characters or words or other "output" forms and its recording medium can be measured in terms of pages or geometrical volume or weight, but its value varies with timeliness, with format to aid comprehension and further processing, with knowledge already possessed by the recipient, and with the relevance of what is delivered to the recipient to his concerns at the time of his receiving it. 3A8A

The extraction of information from mechanized storage forms by copying does not deplete the original resource, but the distribution of those copies may either increase or decrease its value depending upon whether the usefulness of the information is determined by universal or exclusive availability. 3A8B

The mechanical methods for handling information, whether computer forms, electronic communications, or microimage forms are such that the information is not directly perceivable by human senses and what is being handled and managed requires a high degree of ability to deal with abstractions and indirect perception. 3A8C

All this suggests that care must be exercised in looking to normal commodity production and routine automation projects for guidance in information automation projects. The economic and human factory have new and more complex aspects that must be considered as well as the normal concerns in such endeavors. 3A9

Dimensions* 3B

The dimensions of the overall "national memory" problem are difficult to pin down in accurate terms. Figures are not available on a consistent basis for all classes of libraries and those that are available are not necessarily for the same year. However, some gross figures are cited below to establish the order of magnitude of the resources employed in the great part of the memory institution in order to give some concept of the potential market for automated aids or the products of computer-based systems and the direct resources they can impact. 3B1

The American Library Directory 1970-1971 edition records 27,180 libraries in the United States categorized as public, college, special, armed forces, law, medical, or religious. 3B2

Too often, the feeling is that the library and information processing community is too small to have a significant impact on determining the characteristics of equipment to be marketed by the information technology industry through verbal market mechanisms. The result is that the approaches taken have been to take things produced for other purposes and "make do" or "adapt". That this need not be true is illustrated by the order of magnitude figures collected below. 3B3

The collection and summation of these figures is not meant to imply that the total is a measure of the dollar volume of equipment and services to be furnished by the information technology industry because it is a total budget basis, the majority of which is now labor costs, not equipment of materials. As pointed out elsewhere the initial savings sought through automation in overall unit processing cost (improved timeliness and completeness are other dimensions of improvement) occur by shifting the load from manual to machine activities, thereby making the target of costs to be impacted the majority costs of labor intensive operations, not the relatively minor costs of present equipment and materials. Once the balance has shifted, the subsequent attempts to achieve further cost savings must then be aimed at the equipment and materials segment of the overall costs. 3B4

Thus, the basis of an order of magnitude of total costs to be imparted the following illustrate that, if unified in view and demands, the library and information community has the economic potential as a market to influence products and services from the information technology industry. 3B5

Some earlier figures show that 2,370 college and university libraries in 1967-68 carried 305,000,000 volumes, served nearly 7 million registered borrowers and had operating expenses for the year of $509,000,000. The growth of these libraries is evident from a comparison with statistics from 1962: 1,985 libraries, with over 201,000,000 volumes, nearly 4 million registered borrowers and operating expenditures of $183,900,000. Later, in 1966 the data were 2,207 libraries, 6 million users, and $320,000,000 in expenses. Rough extrapolation based on these data indicate an estimated 1971 total operating expense, for college and university librariesy alone, of the order of $600 million, with some account taken of recent "plateauing" of budgets. 3B6

The 50 largest academic libraries in the United States in 1968-69 housed a total of 104,169,000 volumes, added 5,348,000 volumes during the period, had total operating expenditures of $152,448,000, expended $86,255,000 for salaries and wages and spent $48,975,000 for library materials (exclusive of binding costs). 3B7

By June 30, 1969, the total holding of the Library of Congress numbered more than 59,890,000 items including 526,000 reels and strips of microfilm. Appropriations for the operation of the Library of Congress in fiscal 1970 totalled $43,856,300, an increase of $2,143,400 over the previous year's appropriated fund total. (Bowker Annual 1971 p232) 3B8

In fiscal 1968 public libraries serving populated areas of 25,000 or more 135, shelved more than 186,000,000 volumes and had operating expenses of more than $421,000,000, and served more than 125,000,000 people. (OE publication 1968) 3B9

In addition, approximately 49,000 libraries serving public and private elementary and second schools in 1968 carried in excess of 209,000,000 volumes and expended over $85 million for library materials. 3B10

The number and budget for special libraries in industry are unknown. Projects to determine these figures by associations interested in them have been started but abandoned before completion. 3B11

Another view of a portion of the library or information world reveals that for fiscal year 1968 federal funds for scientific and technical information activities within various departments, such as Defense, Health, Education and Welfare, and Commerce, are recorded in excess of $534,000,000 with an additonal $295,000,000 reported by the

Department of State for foreign Dissemination and Research programs. Of these funds, over $155 million is accounted for in bibliographic and reference activity, with publications and distribution costs accounting for $172 million. (The M. Handerson paper.) 3B12

Even with the uncertainties involved in a mixed set of figures as given above, it is obvious that the nation is now spending well over two billion dollars per year on formal library and information services, a problem truly of national proportions. 3B13

Another dimension is visible in the estimated worldwide annual production of printed materials - 450,000 books, 200,000 periodicals, and 200,000 technical reports in 1966 and a rate of increase estimated at 8-10% a year with the increased cost of operating libraries estimated at 10 per cent a year. (Locke citation, also Special Libraries in Federal Government) 3B14

Staff salaries account for roughly approximately 50 per cent to 75 per cent of library operating costs. Book purchases are generally the next highest category of expenditure and may be roughly approximated in the thirty per cent or less category. Another clue bearing on the nature of library operations and the nature of the labor involved is afforded by Locke's calculation based on an analysis of a given university's library budget of approximately $2 million, about six-tenths of one per cent of the total budget was attributable to the time and effort expended by trained personnel in sophisticated reference search. Even if one were to challenge the approximation, the demonstrable fact remains that a considerable part of the workload in any library consists of carrying out repetitive tasks such as cataloging, searching files, shelving, preparing lists, checking on circulation and other routine labor-intensive functions. ("Computer Costs for Large Libraries", W. N. Locke) Clearly, the first target of automation is to take over repetitive tasks reducing labor costs and to improve performance factors of information transfer accuracy, cateloging timeliness, and access flexibility. 3B15

Looking "inward" at information activites themselves gives a very incomplete measure of the economic dimensions of importance of the information problem. These activites have some unknown leverage on the total conduct of industry, education, and government. The absence of adequate information services causes organizations to do much for themselves in inefficient and duplicative ways.

The lack of assured completeness, insufficient timeliness of availability, and unreliable access routes of existing recorded information cause organizations to expend resources to collect that information on their own and deny them the availability of useful information that already exists but cannot be found or delivered for use. Often, even if it is available and at hand, it is not in a form that is easily assimilated and organized or combinable with the information generated within the organization. If the formally recorded information were readily available through responsive library and information services, the internal resources of organizations could be diverted to the generation and organization of information from within, merging it with the information from outside, and making it available for use in decision making. 3B16

An introspective illustration of this situation is the task of finding the already existing information relevant to this study and then boiling down the relevant parts of what was found to provide a starting point for new intellectual contributions desired from this study. Continuing this recursive observation, perhaps the sequence of studies and plans relating to the "information problem" would not be so interminable if there were more timely and effective information support means to aid decision making and development effort in solving it. 3B17

Just as the transportation systems were the key to the development of commerce, modern society is becoming increasingly dependent on information transfer. The long term memory function in this transfer, the library function, can be a very important function if it is properly developed to be consistent with the modern environment of large volumes of information and quick response demand for it. However, trying to put a dollar value on that function, as a measure of it as a market for commercially supplied technology or as a commercial value sufficient to produce the revenue necessary to sustain it is extremely difficult. All we have at this time are rough estimates of the current magnitude of library activity and an unknown leverage of that amount of effort on the overall productivity of industry, education, and government. 3B18

SALIENT POINTS OF THE TECHNOLOGICAL AND DEVELOPMENT PROBLEM 4

As compared to other applications for computers and related technology, there have been requirements for library and information systems which have exceeded those for scientific and business application. The character set needed is much

larger than the 48 to 64 character sets which have been fully adequate for scientific calculations and business records. 4A

The printing industry has developed over many years the type fonts and page formats which are effective in presenting printed information to humans in compact forms for their rapid and accurate consumption. While there can be considerable argument over how much of the quality and stylistic variation are in fact necessary, there is no doubt that what is needed has far exceede conventional computer output capabilities. Library and information systems must also handle multiple alphabets of non-English language material and scientific notations. Even with elimination of stylistic variations of typography and tight restrictions on symbols from foreign alphabets, the minimum library character set for bibliographic information, as described by the Library of Congress includes 176 different characters, significantly in excess of the 95 graphics of the standard ANSI basic set (reference). 4B

Entities to be represented: bibliographic information, subject classification, free text, etc., are complex and difficult to standardize because of the large number of independent institutions involved. The organization and ordering rules necessary to sort the information for human lookup, so that related items are brought together, are much more complicated than simple alphabetic ordering. 4C

Whether dealing with bibliographic information or actual information content, the files involved are very large - resulting in very long computer runs with serial files on magnetic tape, or alternatively, requiring an amount of direct access storage that has been outside the bounds of economic practicality. Some of these technological barriers are falling, as will be noted in a later section of the report. 4D

Because the quantity of information is large - again whether the full content of original documents is involved or only bibliographic information and other document surrogates - the labor of original input into machine handleable form is a serious functional and economic problem. In principle, when the information can be captured at the original source, in a form suitable for later reuse, this would eliminate that problem. However, that is not yet possible in today's publishing environment to any significant extent. 4E

Looked at another way, the requirements of the library and information systems problems, as a combination of technological requirements and economic practicalities, are no different in character nor more demanding of computer science and engineering than are many of the other current

applications involving the integration of intellectual and machine functions. Thus, while its challenges may have been excessive for the capabilities of the past, it is not unique in the computer system research and development of the present. 4F

(Examples of other relevant problems illustrating similar characteristics) 4F1

The application of computer technology to libraries is in its infancy. Experience to date indicates that automation involves a series of technical, intellectual and economic problems which must be solved by highly skilled people with combined competence in computer technology as well as bibliographical technique. Cost elements involve systems analysis and design, hardware, file conversion, and the expense of operating the system. Because of current economic factors it is not practical at the present time to consider computer technology for large scale text storage or information retrieval. However, libraries contain numerous files for the control of book ordering, serial receipts, binding, cataloging, and circulation are suitable for automation under appropriate circumstances. 4G

Until recently it was anticipated that programs developed to automate operations in one library couldbe transferred to other institutions, thus saving the cost of duplicative design. Unfortunately, because of differences in hardware configuration and operations requirements among libraries, the transfer of automation programs has serious limitations. 4H

An investment of $70,000 to $100,000 may be required to design the mechanization of a single library operation, i.e., acquisitions or circulation. New money must be provided for these development costs as they are in addition to the budgets required for operating regular library services. In addition, funds must be found to pay for file conversion and to operate the system (hardware and staff costs). This suggests that only a few of the larger libraries can justify the expense of location automation. Alternative approaches are being developed by creating service agencies to provide automated programs for consortia, or regional groups of libraries such as those in Ohio and New England. The University of California is designing a system based on the concurrent automation of the nine libraries within its system. 4I

One of the few constants which comes from experience is that there should be no anticipation of an immediate cost saving from automation. However, the application of computer technology can produce more effective service for patrons and

more efficient records for library management. Automation can accommodate future increases in work load at less increase in labor costs than a manual system. 4J

There are some special features that must be kept in mind during design. The overall system is complex, but it must be used, for the most part, by persons with special training. The input to the system is not in the control of the system, although the content of the system provides part of the source material for intellectual activity which provides that input. The storehouse of knowledge that the library function represents in general is a cumulative, continually growing, non-self purging file for which no acceptable purging criteria exists. Obviously, this leads to a heirarchy of memory functions, so that "the library problem" is really a heirarchy of problems, not a single one. 4K

While not truly unique in character but certainly extreme in magnitude is the problem of "getting started". To be useful, an information file must span sufficient scope that the user can depend on it to contain a sufficient percentage of the existing information thata user needs to lead him to useful conclusions. Otherwise, the situation, whether intended to be fully operational or pilot experience, is unreal. For example, the MARC pilot bibliographic information service is "useful" to the extent that it supplies an appreciable percentage of bibliographic input of the receiving organization. In situtions where it does not, it can be regarded only as an experiment, the worth of which cannot really be evaluated. Building the files, whether content or control and access, to a level of true usefulness represents a large investment without immediate payoff. 4L

Human knowledge has been recorded for centuries in the form of readable hard copy. Libraries have developed on that basis. The impact of the use of computers in business, government, and intellectual activity have presented the library function with a new challenge, information recorded in machine readable form which is not organized and formatted as a publications would be. 4M

A well known example is the 1970 U.S. Census which serves as a prototype of a new information form. A relatively small amount of data from this machine-readable data base will be published. Effective utilization of the bulk of the Census information is difficult. First, purchasing the xxx reels of magnetic tape requires an investment of $xxx. Secondly, the information is not arranged on the tapes for effective use. A considerable investment must be made to compact the data in a more efficient format. The third

problem is the cost of computer time to run the tapes in response to inquiries. These issues present no technical difficulty; however, libraries are not organized or funded to effectively resolve this type of problem. 4M1

If libraries are to serve as the national memory function, they must expand their interests and capabilities to handle and make available information from such media. Various audio/visual material are other frontiers to conquer. present activities in these areas can only be regarded as exploratory. Allowances for cataloguing and accessing such media should be made in developing computer-based library systems. 4M2

In contrast to other information transfer media, e.g., television entertainment, news, and serial publications, there is no associated source of revenue such as provided by advertising for the examples cites above. The tradition of public and institutional support of libraries has removed them in the past from the mechanism of a marketplace. Because they must solve this problem of revenue acquisition in innovative ways in addition to technological development, a double burden is imposed. 4N

Because the responsibility of the memory institution includes the organization of existing knowledge in anticipation of unknown or unstated future requirements, it cannot exist solely on the basis of justification by current needs. Thus, the motivation, guidance, and support for building this resource needs specific attention, apart from the economics of using the resource once it exists. 4O

When a specific application or mission-oriented interest develops, such as lunar and planetary exploration in the 1960's or "pollution" in the late 1960's and for the 1970's, information is needed from many fields, from the "hard" sciences, economics, social sciences, politics and government, etc. Information which has been organized along traditional lines in the past must be reorganized or somehow accessible via new intellectual routes. Things which before were unrelated must be brought together under new categorizations. This means that the memory institution cannot assume an organization for is information that is static and should provide as much basis as is practical for automatic reorganization when new needs are recognized. This represents the statement of an open ended design problem, the solution of which in fact can only be approximated and not totally achieved in our present state of ignorance about the fundamentals of information. Also, the postulation of a value measure for such a requirements for comparison with costs to satisfy it is

futile, especially when the statement of that value measure for satisfying known needs has not been worked out. 4P

There are enormous variances in the usage of any given item in a library or handled by an information system. There is an age-old argument over the storage of little or never used material. Certainly, anything done to segregate the active material from the inactive can help to lessen the amount of information that must be managed in ways for maximum accessibility and responsiveness. Major libraries now have off-site, high density storage for this purpose, but it is never very popular with the library patrons to have part of a collection of interest to them removed from immediate availability. However, it is unlikely that the national memory that the library function represents will be a single level storage with any technology. 4Q

Similarly, there are wide variations in the need and demands of information seekers. Whether they should or not, some persons never consult already stored and organized information. Others make almost constant use of it. If society is to be served, any solution to the library problem must take this into account and use methods of operation and financing that does not deter the frequent user and does entice the present non-user to learn and begin to get value from stored information. 4R

The development management for the design and installation of computer-based library and information systems faces special problems. In general, the computer systems industry has a spotty record for the reliable delivery, on time and on budget, for large, complex automation projects. Software engineering has not generally reached the level of organization and control that exists in hardware engineering practice. (Reference reports of NATO conferences on software engineering) Both library science and computer science are complex and specialized areas. The information for each subject specialization has characteristics and traditions determined by the subject content itself and its history. Effective decision making in the technical management of library and information system automation projects requires at least some technical insight into the various areas and a facility for communication with the specialists of all. The basic design problem requires technical specialists who are competent to some degree in multiple fields or smoothly functioning teams of technical specialists who can at least communicate with each other. All of these characteristics extend the already difficult task of managing the design and installation of any large automated system. 4S

INVESTIGATION OF SOME KEY ACTIVITES - AND THE PRESENT STATE OF THE ART 5

In order to provide an empirical base for the work of this Panel in addition to published information and the direct experience of its members, several visits were made to discuss the problem, the technology, and current achievements. These visits were chosen to sample a broad range of activities, not to attempt a comprehensive survey or evaluation of all present library automation projects. These visits also provided a way to augment the knowledge of the individual Panel members and to blend that knowledge into a nucleus of common experience for the Panel. 5A

The organization and projects visited were: 5B

Bell Telephone Laboratories, Murray Hill, New Jersey, Main Technical Library - Illustrating the use of a quick response computer system in the management of the collection of a small network of libraries serving high technology research laboratories. 5B1

Harvard University Widener Library - A very large major university library having a unique cataloging system. 5B2

Project INTREX, MIT - A research laboratory for the study of highly interactive information retrieval, including text image access. 5B3

TIP-MIT - An operational information retrieval experiment using a time shared computer system for bibliographic access. 5B4

National Library of Medicine - A large scale national library devoted to a single specialization coupled with abstracting and indexing activities and research on interactive bibliographic access. 5B5

Library of Congress - The de facto national library covering all fields and present source of machine readable cataloging for current English language monographis. 5B6

Institute for Library Research, University of California, Berkeley, California - Headquarters for a long term project in library automation for the nine major campuses of the University of California. 5B7

Stanford University - Spires/Ballots Project - Specification, design and development of an integrated data

facility in support of library and information retrieval activities. 5B8

Stanford Research Institute - Augmented Human Intellect Research Center - A long term project and research laboratory on man-machine systems for accomplishing intellectual tasks. 5B9

IBM, Los Gatos, California, Technical Library - Experimental computer-based system for managing a single technical library. 5B10

University of Chicago Library - A major research library, with a long term automation project and a user of MARC tapes. 5B11

University Computing Company, Dallas, Texas - A commercial computer service bureau which operates a nationwide computer utility and which has plans for a data transmission network. 5B12

New York Times Information Bank - An information service with remote bibliographic and text access intended both for support of the day-to-day news media operations (i.e., automated news "morgue") and for an information service operated on a commercial basis. 5B13

In addition, discussions were held with the Defense Documentation Center and participants in the Federal Library Automation Project sponsored by the U.S. Office of Education. Many more library and information-oriented automation projects were brought to the attention of the Panel by individual contact and through the many publications and reports of such projects. 5C

From these various projects, it is quite clear that useful automation has been brought to bear on the management and aids to bibliographic access of industrial and other libraries specializing in current materials, particularly where there is a well developed computing base used for general computational and business support of the environment in which the library fits, such as exemplified in the BTL and IBM activities. These results and the progress of several groups have shown that these methods can be applied locally to large libraries and closely knit groups of libraries, provided that the problem is undertaken as a serious large scale design and engineering project and that an appropriate multi-disciplinary design team is employed. There is nothing technologically deficient for automation of purchasing, inventory control, and handling of bibliographic information. Economic payoff can

come in saving labor normally used for repetitive tasks by being able to manage the collection more efficiently and in being able to use cataloging information supplied by others. There have been economic barriers in the financing of the high costs of local development of suitable software and there has been no satisfactory record of multiple use of such systems to spread the cost. 5D

For the most part, the computer science and engineering industry knows how to do the individual functions that library and information system users would like to have done. What hasn't been achieved is to be able to do them cheaply, rapidly, and reliably in the environment of a large evolving system. 5E

TIP and the SRI system both show what can be done and the flexibility which can be achieved in a research environment. INTREX demonstrates some of the hardware possibilities on a laboratory basis. Yet, at the present time, we can easily get more users simultaneously accessing the card catalog at a single million-volume library than we can get simultaneously operating consoles on the largest time-sharing computer system in existence. 5F

Also, it does not take very much to bring even the most sophisticated of present-day systems to a crashing halt. While recovery can usually be achieved with a minimum of information lost (the system being implemented for The New York Stock Exchange guarantees no loss), nevertheless the casual user is unlikely to tolerate such interruptions unless the benefits far outweight the inconvenience. 5G

An equally important problem is the need for discipline and precision in interacting with today's automated systems. Such systems are very intolerant in terms of their ability to adapt to imprecise inputs. Again, while we know how to deal with at least a measure of imprecision (format errors, spelling errors, unexpected responses, etc.), the techniques are expensive and cannot begin to deal with all such "errors". Hence, the orientation and training of users must be carefully and slowly carried to the point where they know exactly what the system expects of them, how it reacts when they do the unexpected, and finally how to recover when the system leaves them in the lurch. In the scientific or computer-oriented environments in which today's large scale computer systems typically operate, such training is both relatively easy to give and readily accepted. However, it's not what can be expected of a casual library user and perhaps not even of most library operatingpersonnel who may not be machine-oriented to begin with. 5H

A lesser problem, but still important today, is the difficulty in organizing very large files the rapid access for retrieval and reasonable access for maintenance, while still retaining reliability in operation and flexibility in adding new types of data as they become available. Typically, today's large systems are designed around a particular file organization, and any changes other than the most trivial are apt to mean expensive reprogramming and file conversion. More flexible systems such as some of the military's intelligence and command and control systems gain much of their file flexility through interpretive programming approaches which sacrifice speed and increase cost. The designers involved in most of the activities we visited recognize and cited this problem. 5I

A factor which made a very significant impact on those members of the Panel who had not had extensive experience in large library operations was the massiveness of the general library problem. In such institutions as the Library of Congress and the Harvard Widener Library, the physical size of the collection, the extent of and activity around the catalog card files, the amount of material awaiting processing, the glacial pace of manual methods, and the constant requirement for exercising human intellectual judgment in processing are all sobering things to observe. While these huge institutions are only one end of the library and information system spectrum, they serve to emphasize the complexity and magnitude of the problem of managing and accessing the "national memory". 5J

There are several large projects for applying computer technology to major research and academic libraries or regional cooperatives. The Panel visited only a few of them. In such projects as are carried out at the University of Chicago, Institute of Library Research, University of California (Berkeley), and the Spires/Ballots Project at Stanford University, the problems of large computer system design are fully recognized although not yet completely solved. While these cited examples represent a wide variation of environments many common themes exist. The efforts are aimed at economically viable operation systems, not experiments. The computing power for the systems must be obtained from general purpose computing centers and the scheduling algorithms and changing algorithms often discriminate against the large file, continuous use processes needed in the information processing systems. The development of multi-disciplinary design staffs has been difficult. The system developments have turned out to be more difficult and taken longer than originally expected. All of these projects have a significant amount of work yet to do to reach a fully operational status, and all are facing serious budget crises in routine operations and development activities. 5K

As commercial computing service bureaus develop into nationwide network operation, they provide a third alternative for libraries and information processors as compared to dedicated equipment or shared use with other parts of the host institution (e.g., the use of the university computing center by an academic library). For this reason, the Panel visited one such company, University Computing Company, Dallas, Texas, to sample non-traditional views on the problems and prospects. Typical of several such endeavors, the equipment capability exists today or is rapidly coming to provide some classes of data base service with remote access from all regions of the country, but there are no customers for it. Various reasons were cited such as high cost of preparing the data base, too much individual variation in customer requirements, and problems of design in the basic processes well above the level of mechanization. 5L

In contrast to other application areas, where computing and data processing application programs can be developed by the computer utility and offered to many customers, there are only a very few offerings of library support processes by software developers. While those that do exist may operate quite well, there are few actual customers for them. 5M

The problems of computer technology transfer were observed on two levels. As is true with much of the application software development of the past and present, the computer programs are simply not designed from the start to be transported, even between two installations having the same model of computer. They are incompletely specified and documented, sensitive to specific configurations of secondary storage, I/O equipment, and terminals, and dependent on specific versions of operating systems. Transfer between two types of computers is even more unlikely. With diligence and skill on the part of the receiving institution, ideas can be transfered, portions of programs might be usable if common file forms are chosen, and one institution can learn from the struggles, achievements, and mistakes of another. The library management and information access systems that have been built so far have grown with the evolving understanding of the automation target. 5N

This aspect of technology transfer can be improved by careful engineering for that purpose from the beginning once the system problems are understood and concepts proved out in prototype or pilot systems. However, the more difficult level of system transfer involves the transfer of the entire concept of doing business. A computer program system is only the vehicle for accomplishing a set of functions and at the present state of the art, such program systems mirror the

operating environment, rules, and individuality of the specific institutions supported. Until much greater compatibility of conventions and procedures is achieved in information processing, organization, storage, accessing, and managing, the complete transfer of systems will not be possible. 50

INFORMATION TECHNOLOGY 6

In very large part, the computer hardware appropriate to an electronic information system is the same as that appropriate to a commercial data processing system. Where this is true, there has been little effectiveness by those iterested in information systems in influencing cost or performance of the hardware. Fortunately, keen competition for the very large market represented by data processing now provides the influence, beneficially. Similarly, the capacities and costs for communications media are influenced by the powerful forces of business and personal communications, the broadcasting industry, and other general requirements. The appropriate strategy of libraries and information systems is to be in position to use new capabilities in these fields when they occur, not to carry on basic developments in the hardware for operational use. Also if the information activities were not so fragmented as a markt their economic leverage potential could be considerably greater. 6A

For two decades, there has been continuous improvement in the cost/performance ratio of electronic computational and data processing systems, an improvement which as been sufficiently uniform that statistical extrapolation for two, three, or four years has yielded quite good forecasts. Two major influences have contributed to this improvement - competition among suppliers of such systems and assiduously pursued research and development programs by government agencies and by private industry. 6B

Both elements are necessary for continued improvement. The extent to which competition alone can sustain improvement is limited by manufacturing costs, and without competition, a vendor would have no incentive to pass along to the user the cost benefits which result from research. 6C

During 1970, there was drastic curtailment of research and development programs both by the government and by private industry, particularly of longer team programs whose benefits could be expected in the marketplace in three to five years. There is scant reason to believe that there will be any substantial resumption of research and development in 1971. 6D

It seems likely that statistical extrapolation will continue to yield good forecasts for the early years of the present decade, but that in the middle years there will be a flattening of the improvement curve. This trend must be watched closely so that operational library and infoomation system plans can be adjusted to be consistent with the rate of development of the computing industry. 6E

At the present time, the large data processing system out-performs the smaller one on a dollar for dollar basis. It was at one time a rule of thumb that one could buy a factor of ten in performance for a factor of two in cost. It was therefore advantageous to amalganate many functions into a single centralized system, even though substantial communications costs may be incurred by doing so. However, it has been the recent trend that much more improvement in performance per dollar has been achieved in small systems than in large ones, and thus the relative advantage of the large system has diminished. It will be prudent in the future to give attention to this trend, rather than to assume without question that the centralized system is less costly than are several smaller ones to perform the same set of functions. 6F

The general trend of costs of achieving individual electronic functions is downward, but this does not usually result in devices of lower cost. Rather, it results in devices of greater performance without greater cost. The individual function cost reduction may be expected to continue, but probably with no sudden spectacular improvement of the magnitude represented by the change from discrete to integrated circuits. 6G

In commercial data processing, there is a requirement for large volumes of machine-accessible storage, as presently typified by the rotating magnetic disc. Based on this requirement, the competition among vendors, and known but unpublished research now in progress, it may be forecast with a high degree of certainty that the cost to the user per bit of such storage will dimish by a factor two or three within the next very few years. It is a reasonable forecast, but of less certainty, that within the decade the factor could reach ten. 6H

In addition to Panel views on this subject, an independent source of organized opinion is The Bernstein Report. It is the culmination of a year-long effort by Bernstein and others of the Research and Development Division of the U.S. Naval Supply Command to gather and distill expert opinion as to the future of information processing technology. While the resultant forecasts are opinions and, as such, their accuracy

is open to question, the thoroughness with which the technological community was canvassed leaves no question that the opinions were the best available at the time. The greatest shortcoming is that these are the opinions held by experts in 1966 rather than those held today. There is no comparable record of more up-to-date opinion in the public domain. 6I

Direct access memory is of the utmost importance in an electronic information system. It represents the major portion of the hardware cost for the central installation, and its size determines how much information the system can make immediately available. It is necessary because access to information stored on less expensive magnetic tape is cripplingly slow. 6J

A typical medium-sized data processing system may have a hundred million bits of direct-access storage, costing perhaps a tenth of a cent per bit. The very largest systems may have one hundred times as many bits, with the cost per bit down by a factor approaching ten. Since this very largest capacity is only equal to the full text of about a thousand books, the future size and cost of direct-access storage is of much interest. 6K

On this score, the opinion of the technological community, as expressed by the Bernstein Report is indeed optimistic. Specifically, this opinion holds that there will be: 6L

1) At a cost much less than .001 cent per bit, mass memory (=direct-access storage) of 10↑11 to 10↑12 bits, average access time under one second. 6L1

Earlier time - 1972 6L1A

Highest expectancy time - 1978 6L1B

Confidence level - 8, on a one to nine scale 6L1C

2) At a cost of .00001 center per bit, mass memory of 10↑10 bits, ten millisecond access time, 6L2

Earliest time - 1973 6L2A

Highest expectancy time - 1976 6L2B

Latest expected time 1978 6L2C

Confidence leve 5. 6L2D

3) At a cost of less than .0001 cent per bit, mass memory of 10^13 bits, access time under one second, 6L3

Earliest time - 1974 6L3A

Highest expectancy time - 1977 6L3B

Latest expected time - 1980 6L3C

Confidence level - 5. 6L3D

on the "Instabilities of hardware and software". 6M

In the past, computer-based information handling or management systems have been designed on the basis of specific hardware, specific languages appropriate for that hardware, and specific operating systems. In many cases, particularly in the case of libraries, these "application systems" (as viewed by the computer field) have been operated on central facilities of the institutions of which the library is a part. As changes in hardware and operating systems have occurred rapidly and continuously, the foundations for the information activities have been unstable, causing continual redesign and operating problems. The basic reasons for this involve the continuity factors relating to the specific environment of library and information activities, the system design requirements and techniques to permit survival in that environment and the computer science and engineering tools available to carry out those designs. 6M1

The continuity of scientific calculations rests in the algorithms in the programs to do those calculations. As programming languages have developed, especially FORTRAN, the exportability or moveability of such algorithms has become a fairly routine process. 6M2

The continuity of business calculations is both in the algorithms and the files they operate on. COBOL provides for the moveability of business calculation algorithms. The files involved, in the large majority of business situations, are relatively small (compared to published information files) and highly formatted and simply functional (as compared to the variability of general information files). Both characteristics make the conversion problem, from one system to another -- when conversion is necessary -- relatively early. 6M3

The major continuity of library and information systems is in the files of the system and the portions of the software

that interface with them. The development of programming languages is aiding the ability to design "transferable processing algorithms", but the development of general data base organization and management software is still deficient in terms of what can be purchased. The situation is complicated by the complex interaction between the computer operating system and data base software compounds the problem. Thus, it is not enough to have ways to bridge between dissimilar hardware but it is also necessary to be able to bridge between successive versions of the operating system for the same hardware. 6M4

The computer industry is far from having compatible operating systems (except where one manufacturer specifically emulates the software of another) and the data base management problem has hardly been touched for commercially available and supported systems. Those things that are available are "file" management systems not "data base" management systems and even at that level, are a part of the corresponding operating system, and not sufficiently stable yet from one version of an operating system to another. 6M5

Notes: 6N

Additional material will be collected in this section relating to current status, trends, and expectations for terminals communications, microforms, OCR, etc. 6N1

Software comments on file organization, programming languages, etc., are also needed. 6N2

Comments about long term desires for digital storage of content, but present impracticality, should be made. This is what leads to recommendations that large scale systems handle original documents by images on microforms. 6N3

The problems of computer operating system dependence should be pointed out. 6N4

Problems and Impacts (outline) (Many of these points are showing up in the verbiage already prepared - what belongs here and what belongs elsewhere will be determined by final editing) 60

Diversity of traditions and constituencies 601

Operational cooperation and work sharing 602

Keeping in step with general level of development of technology - both plan ahead and be realistic 603

Transferability of design and operations 604

Design for transferability 604A

Local vs. networks - differences in design approach 604B

Compatibility and standards - appropriate level 604C

Unambiguous identification and representations (e.g., bibliographic elements) 604D

Past and present technology 605

Storage - size and cost, organization 605A

Communications - bandwidth and cost 605B

Character sets - input, storage, output 605C

Physical collection (present) - lack of suitability for automated handling 605D

Specific device independent system design and problems of validity over extended periods of time. 606

The input problem 607

Current material as part of publication activity 607A

The backlog problem 607B

Levels: Bibliographic, the collection itself 607C

Planning and development management 608

Other influences 609

Publishing industry - libraries must develop consistent with the forms of material available 609A

Data transfer utilities and common carriers - a maze of regulations, franchises, tariffs, corporate interests 609B

Governments activities and involvement 609C

Copyright - the ownership dilemma - payment methodology, billing, accounting and bookkeeping 609D

Budget squeeze 6010

DISCUSSIONS AND CONCLUSIONS 7

Is there a national problem in making recorded knowledge available to humans? 7A

There are certainly problems on a national scale. Throughout the country there are entities, many of them called "libraries", which attempt to make available to each member of the group they serve that part of recorded knowledge that would be useful to him in his endeavors, and fail1 to varying extent to do so. 7B

Were it technically and economically feasible to have at some one place all knowledge recorded in machine accessible form, with machinery and communicatons to make the records human-accessible simultaneously and independently by as many people as so desire, then it would be proper to say there is a single problem with a single solution. Since this is not technically feasible, nor likely to become so in the short or medium range future, there is no need now to raise the question of its economic feasibility. If it ever becomes technologically feasible, at some distance point in the future, to consider such a monolith, it can be evaluated on its own merits in the environment at that time. It iss important at this time not to let such a distant vision lead us away from what can be done now, nor the present impossibility of extreme desires deter attempting what is practical. 7C

For example, to a user, the maximum in convenience would be provided if he could express his needs for information verbally to a machine using the vocabular grammatical form and conversational exchange that he would to a colleague or to a reference librarian, and have the machine respond with printed material conforming to his needs. 7D

If providing the user with maximum convenience in the above sense is considered as a goal, then the achievement of that goal is indeed in the remote future. Even so, it may be informative to look at technological and economic problems involved, to see if it is feasible to move in the direction of the goal in the foreseeable future. 7E

First, providing the facility for the initial voice input, with correct machine interpretation of a large vocabulary and complex grammatical structure, spoken by a person to whom the machine has not been previously conditioned, is a technological problem of great magnitude. Because there is

little visible economic reward for solving the problem, the problem is likely to remain unsolved indefinitely. 7F

Less ambitious versions of voice input to a computer do exist. Characteristically, they have a dozen or two words of vocabulary, require prior conditioning of the machine to the particular speaker, and lack software to interpret input other than as a sequence of single words to be acted on individually. 7G

Computer voice output is a much easier problem. Without much question, its implementation will remain ahead of the input. 7H

Next, having received its input, the machine must perform internal actions in accordance thereto. If the machine contained an intellect equivalent to that of a human being, this step would not be difficult. However, the machine does not and will not contain such an intellect. The machine is and will remain constrained to go through previously determined (that is, programmed) procedures in response to input. The range and complexity of responses of which the machine is capable are primarily dependent on its software, and secondarily the body of data to which it has access. 7I

To be fully useful in the spirit of this discussion, "data" must be indigital form. If "data" is to include the content of even a few thousand books, two economic problems stand out. 7J

The first is the cost of hardware to store in digits the ?trillions? of characters needed, in directly machine-accessible form. Present prices for storage and hardware eliminate from consideration most potential uses of full text machine accessible "libraries". That it is present costs that impose the barrier is to be emphasized, because it is this of all costs in the handling of information by computers that seems most amenable to attack by technology. For the purpose of continuing this discussion, we will assume that that attack has been carried to a successful conclusion, and that cost per bit for large stores has been reduced by a factor sufficient to cause the cost of hardware for storage to cease to be the dominating one. Then the second economic problem comes to the fore, which is the cost of getting the characters into machinable form in order that they be entered into the now low-cost storage. 7K

We note here that the cost of going from one machineable form to another is trivial compared to cost of generating a machinable form by a person pushing keys with his (her) fingers. The cost of producing billions of characters in this fashion is again prohibitive. 7L

This barrier is not one that waiting for technological progress will overcome. If overcome it is to be, it will be by conscious anticipation of the great reduction of cost of storage, and prior stock-piling of candidate text in machinable form. 7M

One possible source of machinable text is the publisher. Some publishers at the present time make paper or magnetic tapes at the time of type-setting (or the equivalent) for their own purposes. If such taps are on hand, and their formats fully known, it is a simple machine operation to transform them to whatver medium and format that is most suitable for entering into a large store. 7N

It is possible that some publishers who do not at present conform to this practice might be persuaded to do so with a very modest monetary inducement, since benefits to them would result, as well as the stock-pile being augmented. 7O

Optical character recognition offers a second possibe source, although it is less attractive than is generating machinable form at the source. Present OCR devices have uncomfortably high mis-recognition rates, and while it is reasonable to predict improvement in this, it is not reasonable to predict perfection. A page-by-page error-correcting scan by a human will probably remain necessary for a very long time. Additionally, there is the practicalproblem of presenting a page of bound book to an optical scanner in such a way tht the entire page is in sharp focus. However, that problem is the same for a camera as it is for a scanner, and we are led to the thought of the OCR device operaing on a photographic image of the page rather than on the page itself. In particular, if the pages of a volume have been photographed for the purpose of reproduction, some simple constraints would make the product suiable for optical character recognition. The constraints are that the framework in which the photographic images are embedded facilitate individual page images being located and moved by machine, adthat the resolution be appreciably higher than the minimum required for legibility by a person. 7P

The Case for Microform. 7Q

There is a variety of purposes that can be served by photographic images of pages of text, neither mutually exclusive nor necessarily all applying to a particular image. 7Q1

First, it may be desired to preserve the information content of a page beyond the lifetime of the paper on which

the page was printed. For this purpose there is little constraint on size of the image other than for storage considerations, and except for esthetics, the resolution need be only above the threshold of human legibility. However, the medium on which the image is made must be of very long term stability. 7Q2

Second, it may at some time be desired to transmit by electrical means a copy of the page to a remote point. It is much more convenient to present the scanner with a photographic image than with the page itself. Since there is degradation of quality in any transmission method, the resolution of the image must be well above the threshold of human legibility in order that the received image reach that threshold. Convenience is markedly improved if the image can be located and moved into the scanning position by a machine rather than a person being required for that task. 7Q3

Third, it may at some time be desired to put the textual content of the page into digitized form. The alternatives are that it be key-punched, which is both slow and expensive, that it be digitized by an Optical Character Recognition device scanning the original page, or digitized by an OCR device scanning a photographic image of the page. Both convenience and cost militate toward the last of the three, if the image is adequate for the purpose. "Adequacy" comprises the same properties needed for image transmission, but to a greater degree. Since the error rate of an OCR device is heavily dependent on the sharpness of the character image that can be formed within it, the premium on high resolution of the photographic image is increased. Since digitizing is much more likely to be a production task -- that is, once undertaken, it will be for many pages rather than for isolated ones -- than is image transmission, the requirement for machine location and movement of images becomes mandatory rather than merely desirable. 7Q4

Finally, it may be desired to have additional copies of a page or a collection of pages, at the site where the originals are or at one or more other places. If there is a copying device which can both utilize and preserve in the copy the features needed by the other uses of the image, then generating useful copies can be a very low cost operation. For material of which multiple copies are desired the cost of making the initial photographic image -- which may have been unpleasantly high -- is not at all horrifying when prorated among the copies. 7Q5

of the various required and desired properties of microform images and their embodiment, no pair is contradictory, and the following set encompasses them all. 7Q6

(1) The image resolution must be high enough so that the degradation encountered is making a photographic copy of the original image followed by using the copy as source for electrical transmission of the image does not result in an illegible final image. 7Q6A

(2) The medium which bears the image must be of very long life. 7Q6B

(3) The image must be as small as possible, consistent with the resolution requirements. 7Q6C

(4) Images must be contained in such a way as to facilitate their being located and moved by machine. 7Q6D

(5) A machine must exist which can locate and move images. 7Q6E

(6) A machine must exist which will make copies retaining all the above properties except for some small degradation of image incurred by copying. 7Q6F

The Case Against Microform 7R

Production of microform images requires the allocation of funds, equipment, and manpower which are in short supply. 7R1

There are not formal specifications for film and equipment to produce images with the required properties. 7R2

There is not an agreed-on physical embodiment for images that can be located and moved by machine. If there is a diversity of embodiments, there must be a corresponding diversity of machines, and as a community effort the whole thing becomes futile. 7R3

The "Library and Information Problem", as an automation development problem, is not a single homogeneous problem because there is a wide range of functions, sizes, financing and environments for libraries and the information production and demand activities with which they interface. What costs are affordable in one situation are not allowable in another. The appropriate operating mode for a small library does not provide the same unit costt and functional performance as for a huge library. The method of serving a graduate student doing original research is not appropriate for a business

executive. We saw many of these variations in our site visits. 7S

Those entities attempting to serve a large group of people are generally facing difficulties stemming in part from size and increase of size. As the body of recorded knowledge expands - and this will be the case so long as civilization as we know it continues to exist - so does the burden of housing records of that knowledge. As the community served increases in size, so does the diversity of interests from individual to individual and thus the fraction of the total of knowledge useful to some member of the community. On one hand, an entity serving a community of static size has increasing requirements, and one serving a community increasing in size at a linear rate has requirements that increase at a rate much faster than linear. If the increased requirements are met by expanding the function in traditional library fashion, the cost of doing so increases not less than linearly with the requirements. One the other hand, those who defray the costs, whether they are a legislature or the managers of an endowment fund, are unwilling and beyond some point, unable, to allocate an ever-increasing fraction of the total available resources to some one function, no matter what its importance. 7T

One concludes that the problems met in attempting to serve a large, growing, diverse community are not to be solved locally with more buildings and more employees, and that if solution is to be achieved, it must be done either by altering the problem or going beyond the local level, or by both. 7U

The establishment of the National Library of Medicine is the prime example of a serious attempt to do both. If successful in the long run - and there is no basic reason to believe it cannot be - it may well substitute for a hundred libraries or portions of libraries each a tenth the size of the National one, and perform service in excess of the combined ability of the hundred. 7V

The alteration of the problem here is the change from diverse" to "uniform" in the nature of the community being served. The going beyond the local level is that the community served is geographically dispersed, not merely that the Library is a National institution. 7W

That the problem can be altered at the local level to the benefit of all concerned has been shown by some universities, and some non-university libraries and other information activities. In each case, it has been by departing from traditional methods to take advantage of electronic means for handling and conveying information. Efforts in this direction

have benefited from commercial work done elsewhere in information handling in that hardware replication is immensely cheaper than hardware development. They have, however, suffered in the past from the lack of off-the-shelf input and output devices capable of handling an appropriate character set, been hindered by the costs of storing or handling large files, and injured by the growing pains of the adolescent computer industry. 7X

Some success has been achieved in attempts to go beyond the local level without significantly altering the problem. In particular, there is vast duplication across the country in cataloguing the same book independently in many places even if it may be done the same way. The Library of Congress attempt to remedy this with MARC II is commendable, but sometimes frustrated by the incompatibility of the cataloguing system used by a library that might otherwise benefit, the slow start-up time for extending a pilot operation to full scape, and the threshold of usefulness (i.e., in scope of coverage) which must be exceeded for it to be economically helpful to a recipient and thus be taken serious. 7Y

(Plus excerpts from Skipper, Engelbart, and Salton statements. Also, the outline of points listed in Kettler's report of the discussion of January 27 (afternoon) provides ideas - please identify parts of your own material that fits and other fragments of text as candidates.) 7Y1

POINTS TO CONSIDER FOR CONCLUSIONS: 8

1) Adequacy/inadequacy of technology 8A

computers 8A1

terminals/I/O devices 8A2

Storage 8A3

Communications 8A4

2) Is solving the Problem essential to society/U.S.? 8B

3) Adequacy/inadequacy of software, file handling, etc. 8C

4) Is there a national problem - requiring a national solution? 8D

5) Importance of the interaction between the intellectual and the technological problem. 8E

6) Effects of the non-technological areas. 8F

7) Political and economic problems 8G

a. applications area 8G1

b. plateauing of required technology 8G2

8) Own facilities vs. network supplied services 8H

9) Role of minicomputer 8I

10) 8J

a. Software exportability; not currently 8J1

b. Role of network in transferring the use of a function, not the function itself. Established but essential. 8J2

11) The "correct" level of "standardization". 8K

12) The "sub-optimization trap" (i.e., cost comparison boundaries). 8L

13) The need for experimentation/exploration. What types? What time frame of expected payoff: 8M

OCR 8M1

facsimile transmission 8M2

file-filling problem, i.e., preparing for future use of image and digital collectsion. 8M3

14) positive action now is necessary in anticipation of information in the digital form. 8N

15) Role of "libraries" for data collections for use and for pilot experience (e.g., census) 8O

16) Interrelationships with access to periodical literature through A&I services. 45-50 per cent of current library publications purchases is for periodicals. 8P

RECOMMENDATIONS 9

1) Commit to, urge on, and accelerate unified cataloging and indexing in computer form (to build into the on-going operational system) with elimination of costly duplication. Provide for both hard copy and computer form output. 9A

2) Take steps to set up a high capacity communications based network marketplace. Someone should develop model processing and service nodes for the marketplace that are designed from the beginning to be exportable. 9B

a. Standard interface - electrical, code (characters), and protocol 9B1

b. Pieces of this are: standards, model nodes, physical net, franchising conditions and mechanisms, and incentives for joining. 9B2

3) 9C

a. Insure that current and new materials become available in high quality microform masters (color?) for later selection and conversion to distributable media. 9C1

b. Convert present holdings in some order of priority (recent materials first) to high quality microform master (color?) 9C2

The combination of computer-based intellectual access routes and image distribution is now a practical thing to consider for large scale development. 9D

(Items 3 and 4 must be done in a way consistent with appropriate copyright and ownership protection. Centralized and uncontrolled free availability is not necessarily implied here; assured availability in machine handleable form is intended.) 9E

4) Start banking digitally produced literature in anticipation of future developments which will make digital handling and storage of full text of active materials technologically and economically practical. 9F

a. newly produced material 9F1

b. active literature from within the existing collection 9F2

c. the rest, as practical. 9F3

If OCR develops, as predicted by some, the microform masters are a natural medium for input to the process for conversion to digital form, as well as being useful as a distribution medium. 9F4

5) Encourage standards at the correct level which ensure

compatibility for automatic transfer, yet permit technological innovation and improvement. 9G

6) Get the intellectual and organizational problems under control to use the technology that is coming anyway. Existing institutions need not be destroyed, but some need to be educated and urged to become consistent with the developing environment. 9H

7) Insure that memory institutions get into position to handle large banks of data produced originally in computer readable form. 9I

<ENGELBART>ADOUG.NLS;2, |-APR-7| |8:57 DCE ; .DPR=|; :JRNLP5706,
0|/24/7| |23|:|6 DCE ; .DIR=|; .HED=".MCH=72; .GCR;.HJH=2;DCE
2|-MAY-7| 23:46 70|0.MCH=65;
.HJH=|;ADOUG | Apr 7! (|900)
Material for: CONCLUSIONS and RECOMMENDATIONS";
| . | . 2 . 3 . 4 . 5 . 6 .
7
.SNF=72; .MCH=65; .SNB=0; .DLS=|; .SCR=2; .RTJ=0; .PGN=0;
.CODE['|]=|75B; 10

ADOUG | Apr 71 (1900)
Material for: CONCLUSIONS and RECOMMENDATIONS

INTRODUCTION |0A

Consider the technology that our Panel studies, for its possibilities and needs in remedying the "library problem" -- while we are examining possibilitis, trends, etc., we should keep the following in mind: |0A|

During the period in which computer technology is being extensively integrated into the Library System, there will be concurrent activity by other segments of our society, industrously integrating computer sytems into their workaday world -- toward helping in such "intellectual, knowledge-oriented" activities as their studying, formulating, communicating, teaching, deliberating, negotiating, planning, managing, etc. |0A|A

On a large scale this will have an overwhelming effect on the quantity of significant, recorded communication that will need to be stored and rendered retrievable for subsequent access -- i.e. the library problem is going to be made very much bigger by the very technology that provides hope for dealing with it. |0A|B

NOTE: My personal professional activity is very strongly oriented toward improving the information-management capability for working teams. I find that the increase in recorded dialogue that results is quite marked. Moreover, to increase the coupling between the intellectual efforts of different groups, it is important to make their workaday information available to each other -- i.e. to have library-like service over the memos, trial designs and plans, analyses, studies, etc. that are a constant product of an active group. |0A|C

This is the only way (in my opinion) that we can effectively harness the intellectual and knowledge resources of humans within large-scale activities -- i.e. within the social institutions wherein our most complex, urgent problems must be solved. |0A|C|

Thus, technology will likely open the floodgates to the production of a vastly greater volume and diversity of recorded material. |0A|D

The size of currently published information packets is influenced by factors within our past, non-automated

culture. Our emergent high technology could easily make practical the independent publication of paragraph-sized packets as relevant and valuable contributions to a recorded dialogue. (This, in fact, is seen to be very probable from my own experimentation.) |0A2

The "recorded dialogue system" for a social institution is certainly a critical component of what for an organism functions as a central nervous system. It is absolutely essential to the proper evolution of society that the production, management, retrieval and access of recorded dialogue be as effective as possible. |0A3

Viewed as but a part of society's "recorded-dialogue system, today's libraries may possibly show up as obsolete components -- the recorded-dialogue system of tomorrow may well distribute the various functions of a library in such a way that ordering, cataloguing, bibliographic searching, and physical accessing (and studying) are done in widely separated places and/or with distributed services. |0A4

Are we looking for a way to buttress the libraries with new technology, or are we looking for the best way to improve society's recorded-dialogue system? |0A5

Certainly, if we are looking for the best way to improve society's recorded-dialogue system, we must make the most of the resources of our current library systems, and should undoubtedly plan for evolution from the current situation in a harmonious sort of way that among other things doesn't needlessly abandon valuable assets. |0A6

TIMELY SOLUTION REQUIRES SIMPLE PLAN, BUT BOLD AND MASSIVE |0B

It really does seem to me that there is a basic consideration evident here, in the Library Problem: |0B|

There is a very complex interlocking among the economic, political, traditional, attitudinal etc. factors of the Problem. |0B|A

Any significantly effective solution (The Solution) is obviously going to have to make large and pervasive changes, which will cause many, quite traumatic adjustments among individuals and institutions, |0B|B

To make these many changes an adjustments within a

complicated Solution-framework would multiply horribly the difficulty of implementing the Solution. |0B|C

We therefore feel that The Solution has to have a framework that is simple and direct in its essential details. The complex changes and adjustments should follow from the many sectors and groups adapting relatively autonomously to to local situations with unambiguous, simple, straightforward directives and guiding actions from any central agency trying to "solve the problem" -- as contrasted with a plan that involves a lot of centralized (or inter-faction) planning, coordinating, monitoring and enforcing. |0B|D

The "Bold and Massive" aspects would be necessary, I feel, from both psychological and technical considerations. |0B2

Psychological, because there needs to be something in the air that is very clear and compelling about the changes to come, in order to get all parties that have to participate in the changes and adjustments to get about the long and detailed process of doing their part. |0B2A

Technical, in that the objective details of developing new organizational structure, roles, equipment, services, etc. really needs to be thought through and approached in a coordinated, total sense -- not a patching up of old ways. |0B2B

It is quite obvious to us that the changes have to be evolutionary -- so we are not intending the "Bold and Massive" to mean "discontinous, revolutionary". We mean for "Bold and Massive" to describe the goals and energies to be associated with the EVOLUTION of the Library System. |0B3

Achieving a high evolutionary rate, without discontinuities or traumatic shock, can be significantly abetted by building these attributes into The Solution Plan: Simple; Bold; Massive. |0B4

LIBRARY DIRECTORS AREN'T THE ONES TO BEAR THE SOLUTION-BURDEN |0C

Where is the "industry" that can provide "architectural" and "builder" services? |0C1

Whose responsibility is to develop library automation? |0C2

Libraries on the whole seem to be in a helpless position as compared to another kind of industrial plant that might be faced with similar conflicting pressures and demands: |0C3

Library managers are appointees whose operating budget is an appropriation, within an institution (university, county, corporation). |0C3A

They are not trained to innovate new products, build new markets, etc. |0C3B

They have no Board of Directors and stockholders to decide on a long-term plan and commit resources. |0C3C

Are we expecting these "plant managers" to renovate their industry, where it wouldn't at all be the plant managers who would carry this burden in other industries? |0C4

IS REORGANIZING THE LIBRARY INDUSTRY A VALID POSSIBILITY? |0D

The commodity of information has a strange way of being produced, marketed, maintained, etc. Any other industry operating this way would be unable also to cope with large-scale problems. For example, the steel industry wouldn't have efficient furnaces, the transportation industry wouldn't have jets (or even airplanes, what with the fantastic revolution they involved). |0D1

Is it at all possible that this information industry could be re-organized in such a way as to generate its entrepreneurs, capital, etc. within the usual "market" system? |0D2

For instance, how about charging full cost for library service, as cost-accounted in a business-like way. Then, for instance, the budgeting in a university would have to put funds into the hands of its academic departments that otherwise would be apportioned directly to the library, and the library would charge for service costs and perhaps seek profits. |0D2A

Tthe term "information-utility industry" is brought forth regularly for analogy in discussions of this sort. To risk old-hatting a concept in finishing this line of thinking here, we point out that an electric-power utility, for its larger customers, must impose a monthly "demand charge", which represents the costs of

guaranteeing that the agreed-upon quantity of power will be available upon demand. The utility company must establish an expensive, general system of energy converters, generators, transmission lines, and distribution facilities, and for each big customer he must install a specific system of lines, controls, transformers, and such coming into his plant that is adequate to meet his demands. The peak demand that a customer may make (i.e. how many megawatts maximum he can draw during a given time of the week/day) is very carefully negotiated and monitored because of its economic significance in the business of supplying power. |OD2B

Similarly, an information utility would have to charge big customers for "negotiated demand availability," as well as for specific services. For example, the history department would receive a hefty charge for maintaining a large, seldom-used collection in a condition of ready demand. |OD2C

We appreciate such things as the fact that this bill may exceed the current budget of a history department, and that it might be harder for budget allocators to give the money to that department than to the library; but a move such as this would put the decisions about financing library costs where they are more relevantly connected into cost/value tradeoffs, and would have the appropriate effect of distributing the budget-defense burden more to where it ought be. |OD2C|

The Library would get a small budget, and the History Department would have to fight for its "library" money (perhaps make deals with other departments for sharing carring costs of some collectons, etc.) |OD2C2

If, for instance, there actually must be (for all I know) $50,000 per year funneled into the library for each serious scholar, it would seem to me important to be open about it and not hide this fact from scholars and their backers -- show on the books what it does cost, pass the resources through the scholar's department, bill the department for library services and for on-demand availability, etc. |OD2C2A

I know that this line of thought drifts from the appropriate INFOSYS area. I guess that I am showing the results of recent struggles of my own to formulate the working relationships between elements of a large team so as to distribute responsibilites, freedom of choice, opportunity to adapt, etc. over the group, toward maximizing its net flexibility of developing and pursuing human goals. |0D2D

I should point out that one of the significant things that concurrently will be emerging from the harnessing of computer technology, will be new ways in which an organization can be structured into functional groups, and new speed and flexiblity in its inter-group negotiation and transactional administraton. This type of development really could make a significant difference in the ways in which an automatied information utility could operate. |0D2D|

COMPUTER NETWORKS, UTILITIES, AND RESOURCE SHARING ARE THE SOLUTION FOR SERVICING LIBRARY-SYSTEM AUTOMATION |0E

Like any of the other businesses, libraries will find it profitable to automate purchasing, inventory control, staff-task monitoring, etc. -- I really don't think that INFOSYS should worry overmuch about automation of these areas; it will come about very naturally, and be affected mainly by two things: |0E|

Bringing the costs down, and |0E|A

Getting the systems industry to be more mature in its ability to design, install, and operate effective service systems. (See discussion of this factor below.) |0E|B

Bringing costs down will in turn come from two things: |0E2

A large, active consumer market (for computer systems of this kind), so that there is competitive pressure to improve product designs, manufacturing efficiencies, marketing and maintenance services, etc. (This also depends upon a considerably matured computer-systems industry, which necessarily has to grow together with the its market.) |0E2A

Efficient use of capital investments -- as put into software, hardware, procedures development, cataloguing,

microfilming, computer-form transcription, computer typesetting etc. |0E2B

In this latter category, I can't see any other factor as important as "computer networking," with its wider-market participation (sharing) of hardware, software and data-base resources. |0E3

For instance, University Computing Inc. were very emphatic about the improved economy of maintaining and operating, large, multi-computer installations, as opposed to the same total computing/storing power distributed over a number of sites. |0E3A

other segments of our society (other markets) will want very similar services, and libraries will be able to share the larger-market benefits in buying their services from computer utilities. |0E4

Some other needs: logistics, inventory, intelligence, personnel, purchasing, publishing, merchandising. |0E4A

And really, libraries shouldn't be in the technology-management and operating business -- any more than they should be in the publishing business, or in the business of doing original-scholarship research and authorship to produce the journals they handle, or in generating their own electricity, or etc. |0E4B

A larger, more flexible, and more active marketplace must be established, in terms of range of equipment and services available for libraries to select from (and change between), and in terms of the number and diversity of library customers and needs "reachable" by service vendors dealing with transmission, storage, and processing. This makes the difference between impractical possibility and practical feasibility in contemplating a significant degree of automation for libraries. |0E5

In flexibility, range of services and customers, costs, competitive evolution, etc. |0E5A

INTERACTIVE COMPUTER AIDS IMPORTANT, BUT ONLY IN CERTAIN AREAS OF LIBRARY AUTOMATION, AND MUCH PUSH ON THEM ANYWAY FOR OTHER USES |0F

We are quite convinced that interactive computer service

will produce extremely high payoff in supporting the minute-by-minute intellectual endeavors of trained people. The modes of use will be not too dissimilar from the ways a private automobile supports the minute-by-minute transportation activitis of (trained) people. |0F1

BUT: |0F2

one doesn't use his private automobile for all of the transportation sevices. and |0F2A

interactive computer services won't be used for all of the information manipulation services, either. |0F2B

Ships, trains, trucks, airplanes, buses, etc. are an extremely important part of the transportation industry, and |0F3

Remote-batch processes, running at different levels of priority according to sensible scheduling algorithms (which are delicately adapted to many factors in the physical service system and in the users' working and information context), will be absolutely essential in the computer systems. |0F4

VALUE TO BREAK-THROUGH TECHNOLOGY, BUT VAST IMPROVEMENT POSSIBLE WITHOUT IT |0G

We feel that such as massive storage devices, digital communication systems, microform technology, as examples, will improve mightily in the next decade, and that The Solution can (and must) depend upon this impovement. |0G1

The expectable improvements will be adequate -- we don't have to depend upon hoped-for breakthroughs, i.e., establishing and launching the critical framework of The Solution needn't wait upon these technical developments. |0G2

COMPUTER-SYSTEMS INDUSTRY NEEDS SIGNIFICANT MATURATION |0H

More important than the "expectable technological improvements" mentioned above, as things that technology must do for The Solution, is to evolve a more mature "computer-systems industry." |0H1

Consider the industry that exists to provide new buildings: |0H2

ADOUG | Apr 7| (|900)
Material for: CONCLUSIONS and RECOMMENDATIONS

There are skilled, experienced firms of architects who can work with a client to develop an overall plan to suit his needs -- where special attention and interactive dialogue is given to the the "user features," i.e. to the features that the building's users are aware of. |OH2A

And there are many competitive firms of building contractors and sub-contractors to build what the architect planned. And there are many conventions for who monitors what, inspects what, etc., and for how the processes of bidding, negotiating and accepting are done. |OH2B

Relative to the possible improvements, the number and complexity of knowledge-support systems to be developed, their complexity and rate of evolving, etc., the computer-systems industry isn't in very good shape as far as providing really good architects and builders. |OH3

Most architects (of computer sytems) want the buildng job, too, or are even actually the sellers of the concrete and steel as well. |OH3A

And their really isn't within the computer-systems industry a discipline that compares with user-oriented building architecture -- computer-systems "architects" are more like the structural engineers who know well how to make it strong or reliable, but aren't really trained to shape things senisitivvely to the users' subjective and objctive functional needs (much less to their aesthetics), or even very well to accommodate the cost/payoff need/value framework of the users. |OH3B

Also, the "builders" in the computer-systems industry are notoriously bad at meeting schedules and budgets, and their products aren't easily enough maintained, modified, or transferred onto other sites. |OH4

Their systems admittedly are often the most complex that man has designed, and apparently there isn't yet a (known) method for materially improving this situation. But if these complex systems are to serve in support of complex human operations, then there is an absolute need for the quality of both the production process and of the product to improve significantly. |OH4A

For instance, there must evolve cleaner conceptualization

for system functions and system levels; and both programming and problem-descriptive languages must improve. Also, the processes of design and of management must be improved significantly -- as must methodology, tools (including computer aids, of course), conventions, concepts, team-collaboration modes, documentation, methods of bootstrapping systems onto new hardware, new operating systems, or new compiler systems, etc. |0H5

MICROFORM FOR GENERAL WORKING HARD COPY |0I

Much more information to be available to a much larger clientele, is a direct trend; if the automated dialogue obtains, there will be an added increase in quantity and in activity; |0I1

And we are shifting to a knowledge-worker society -- where the product per person and the consumption per person will both go up. Also refer to the considerations in the Introduction, reading computer technology's effects upon the quantity of information. |0I1A

These strongly imply that full-sized paper copies must be superseded by smaller, cheaper physical copy (microform) -- or ultimately, by electronic storage. |0I2

Seems hard to imagine that a microform media won't eventually prevail over paper for the storage and dissemination of most of our published material. |0I3

A TRIAL SHOT AT A "SIMPLE, BOLD, MASSIVE PLAN |0J

General Considerations: |0J1

Global problems become hopeless when faced at a local level, and local problems are clumsily handled at a global level. Effective automation in the Library System involves problems of both kinds, with a very critical dependence upon global framework. |0J1A

Any other industry, to take advantage of technology, needs transportation and communication facilities that bring many types of suppliers, manufacturers, distributors and consumers within ready mutual access -- to provide a sort of physical marketplace in which local transactions can effectively deal with local problems. And, in order that participants in this marketplace can

effectively and freely serve, or be served from, a wide selection of other participants, there needs to be a minimal set of standards. |0J|B

It is our view that appropriate global steps could establish a Marketplace for The Library Industry that would permit its internal workings to evolve through local solutions and adaptations toward improvements of many kinds--wherein successive stages in the quality and quantity of automation sell themselves, not without risk and personal agencies, but within a scale of time, promotion, analysis, dollars, number of parties, etc. that can be handled periodically by local staff. Unless such an environment is created on a global scale, we see little hope for the magic of automation to be soon effective; and if it is created, we don't see that the technological challenges in providing suitable magic involve any remarkable technical developments beyond what seems naturally forthcoming. |0J|C

Specific Recommendations: |0J2

We recommend four specific global actions, that, if integrated within a global strategy which we are not qualified to complete, are aimed to promote the type of marketplace described above. |0J2A

1) Produce a commitment, probably at the Federal level, to push ahead on an orderly but massive scale the following two projects--which may involve several billion dollars. |0J2A|

Produce a standard catalog description of every useful library item, in computer form, and make the basic records available in some open, wholesale, manner. |0J2A|A

The MARC Project of the Library of Congress is dead center on this. From our point of view, it could be ***serve this and by being accelerated and expanded. |0J2A|A|

produce a high-quality photographic master of every useful library item, and make reproductions available in some open, wholesale manner. |0J2A|B

Two uses are seen: General availability of

micro-form copies, in whatever reduction and format desired; and provision of a uniform inputting media for automatic-reading processors which are likely to evolve, so that the material may be read into computers. |0J2A|B|

2) Produce a similar global commitment to develop and maintain a physical, information-transportation network to serve the Library Industry. There might be significant value here from improving the means for transporting physical objects (such as books and films), and this should be considered in a total-system plan. But with respect to potential gains from automation, our attention turns here to the value of an electronic Digital-Message Network (DMN) operated by some sort of central "Department of Library Systems" (DLS) to connect many types of producers, suppliers, brokers, customers, services, etc. |0J2A2

This is directly supportive of, but not equivalent to, nor limited to support of, what is currently called "Library Networking". |0J2A2A

DLS would see that every library could easily be connected to the DMN. The library would be provided with a sort of service box, having a prescribed, standard, physical interface to which may be connected any configuration of devices (typewriters, displays, printers, mini-computers, big computers, optical character readers, etc.). |0J2A2B

Because of the interface standards and the universal distribution of these DMN "ports", manufactures would develop many products from which to select; and there would grow to be many configurations through which a library's local terminal facility could evolve with minimal incremental perturbation. |0J2A2C

It would be reasonable to give public support to DMN, with the same sort of justification as for highways, waterways, harbors and airports. |0J2A2D

Any person or organization willing to offer service to the Libraries, or wishing to use their services, will find uniform regulations (and

charges) for having a similar DMN port established on his (its) premises. This doesn't necessarily mean that anyone should be able to connect--a fuller analysis of this Library DMN System may determine, for instance, that only "licensed retailers of information services" should have usage access. |0J2A2E

3) Launch a concerted, longer-term action directed toward capturing all relevant publications in digital form at time of their publication. (These assumedly to become available to the public subject to copyright laws.) |0J2A3

4) If any specific push is made in computer technology, in the name of a national information system, it should be for improving the industry's effectiveness in large-system development, operation, and modification. |0J2A4

Comments on the above recommendations: |0J3

Copyrights -- |0J3A

The speed and flexibility with which information can be accessed (located, transmited to user, displayed to him) is a very basic factor in any national information information system. xx.and, |0J3A1

it is obviously of very large economic value to have this access be high speed and economical. a |0J3A2

The speed and cost of accessing full-sized, hard-copy information forms are hopelessly outclassed by such as micro-image or digital forms. |0J3A3

It seems obvious that new solutions will have to be found for the control and economic return associated with usage of privately created information. |0J3A4

No matter what significantly improved information system is developed. |0J3A4A

System-Development effectiveness: |0J3B

It is an old, nearly-true statement that "You can do

anything you want with a computer, if you can only describe it properly." |0J3B|

And the Panel seems to feel that the bottlenecks toward making significant improvements in the national information system aren't in the computer-technique domain -- because indeed computer technology CAN or will be able to do the necessary things for improving the national xxxx system. |0J3B2

But, when the aggregate of functions for an effective, large information system is integrated into one coherent system, the basic problems design, coordination, and management of very complexx tojects hits thee of thresholdthreshold our resent capabilitiies. |0J3B3

Marketplace |0J3C

A library manager needs to be provided with a number of choices for improving his cost/effectiveness when the pressures reach a certain point for improving performance and/or decreasing costs. |0J3C|

As it is now, in order to take a significant step toward "modernization", he is faced with a very unsatisfactory situation. :e has really |0J3C2

A plethora of devices that look provacative in the advertisements, no lack of suggestions and pressures to make use of them (from the people who read and believe the future-possibilities extrapolations.) |0J3C2A

What he needs would be the equivalent of a selection of architects who can help him design what he needs within the limis of his budget and in the frame of his particular need/value structure. |0J3C3

The architect of his choice needs a lexible range of products and services to design into the system. |0J3C4

The architect needs to have a flexible market of builders to select from. |0J3C5

The range of building jobs (in nature and quantity)

needs to support a healthy market of architects and builders. |0J3C6

They need to have a healthy, active community of suppliers and support services (sub-contractors, etc.). |0J3C6A

When the library has to install all of its own computer equipment, operate a sophisticated system (economically), as well as create a drastic change in its regular |0J3C7

MISCELLANEOUS: SKIPPER NOTES ||

Digestion Problem. ||A

Briefly stated, the overall problem is that: The increase in knowledge and the development of new intellectual disciplines has greatly stimulated the rate of publication. For example, it is estimated that scientific literature has been increasing at an average rate of 6% per annum for the past quarter century. In addition, certain qualitative constraints do not exist in limiting the quantity or assuring the substantive quality of various new types of literature. The editorial review process in evaluating articles to be included in scholarly journals is lacking in the publication of technical reports. ||A1

Another factor involved in the information problem is the interdisciplinary character of contemporary knowledge. Formerly, literature could be acquired and organized for a limited discipline, such as chemistry. Today, information related to chemistry from fields such as mathematics, engineering, physics and biology must be made available to and used by the chemist. ||A2

Because of the large mass of currently published literature, cataloging, indexing, and abstracting techniques are becoming less effective in leading the reader to relevant information. Society requires better intellectual and physical access to information than can be provided by existing manual methods. ||A3

Multiple Use ||B

It is pointless to talk about the design of systems to improve access to information unless adequate attention is

given to the foundations of the system. Any national program, whether it be manual or based on computer technology, is dependent on comprehensive collecting of relevant literature and its prompt identification and description in bibliographic terms. These responsibilities can be met most effectively on a centralized basis. It is for this reason that the three national libraries in Washington, D.C., constitute the foundation for a national information system. They should be encouraged to continue meeting this responsibility and the Congress should recognize that investments of Federal funds in supporting these programs constitute cost savings and improved efficiency for every library in the country. ||B1

Cataloging copy prepared by the Library of Congress is distributed to approximately 20,000* subscribers through the annual sale of over 60 million* catalog cards. As it costs from three to five times as much to catalog without LC copy, the magnitude of the savings is apparent. Of equal importance is the fact that centralized bibliographic description provides a standard which is essential for national bibliographic control. ||B2

* NDB Editor - see latest LC Annual Report. ||B2A

Of immediate significance for this inquiry is the availability of cataloging data from Library of Congress for currently published books in machine readable form. As the data base becomes expanded by time and the inclusion of additional languages, it is apparent that additional libraries will use this source for preparing local cataloging records. It is also significant that acceptance and expanding use of the MARC II format for these records is providing a defacto if not a de jure standard. This gives libraries the capability of interchanging bibliographic records for the mutual benefit of their constituents. ||B3

Page 20 - reference to Harvard. ||B4

Harvard University Widener Library - A very large major university library which has used computer technology to produce a book-form shelf-list. ||B4A

(The Problem of Library Budgets and Their Structure) ||C

Many libraries may find it difficult to fund improvements

ADOUG | Apr 71 (1900)
Material for: CONCLUSIONS and RECOMMENDATIONS

in their services. It was mentioned earlier that substantial amounts of new money must be provided for systems design work. It is not reasonable to expect that this expenditure can be carried by reducing existing budgets. Cooperative acquisition programs usually give the library access to a greater depth in resources, rather than saving an equivalent amount of money from the acquisition budget. ||C1

Cost recovery capability could be included in new information services. However, the public has come to expect "free" library services as the right of taxpaying citizens. The attitude may frustrate cost recovery. In addition, many libraries do not receive supplemental income -- such as fines -- as a matter of institutional policy. This may be difficult to change. ||C2

MISCELLANEOUS: RLW NOTES AND OPINIONS |2

On the poblem of framework of study, expected results, strategy o report. |3

<ROW>RLWNOTE.NLS;3, |-APR-71| ||:32 BER ; .RTJ=0; |3A

on the problem of framework of study, expected results, strategy of report. 13A1

We will take the points 2D1,...,2D5 of the 8 Mar draft as the gross statement of the desired output of the panel work and thus by implication the prescription of what is to be treated in the report. Points 2D1 to 2D3 are not sequential and must be handled in parallel. The report does not necessarily have to be structured according to these five points. 13A2

We accept as a working hypothesis that recorded and organized information is a vital component in the nation's business, government, educational, and national defense systems. 13A3

We accept the hypothesis, well supported by other work, that present traditional methods for performing the functions now embodied in libraries and related information production, processing, and using activities are not able to cope with the rising volume of information and increased timeliness demanded by modern society and are breaking down. 13A4

Other than what we can cite from site visits, literature, and personal experience, I expect us to develop very little original analytical fundamentals. In cases unsupportable by citation of straight forward reasoning, we may state opinions and identify the associated hypothesis as being in need of proof. Our recommended approach to that proof can also be given as something that should be undertaken. A similar appproach to important parameters to be determined can be taken. 13A5

point 1: to assess the adequacy of technology to meet the needs of library and information systems at the national level. 13A6

To satisfy this point we must: 13A6A

outline , but not establish de novo, the "needs" of library and information systems that are apt to be helped by computer systems and associated technologies. 13A6A1

outline the technology that applies (hardware, software: e.g., input, processors, storage (image, digital), output, distribution and communication,

programming, operating systems,data management systems). |3A6A2

provide a statement of adequacy or deficiency at present or predictable (perhaps with reflections on past deficiencies and the specific developments that have or will overcome them) as individual facilities, techniques, capabilities for specific levels of mechanization and automation (e.g., management of information collections, access to collection, storage and delivery (distribution) of information once accessed, manipulation of information once delivered). |3A6A3

The impacts and trends of costs must come into the assessment of adequacy. The most difficult part of the judgment to be expressed and its support is the need for some kind of standard to judge acceptable level of cost or "affordability" so that the "adequacy" is not only a logical sufficiency but also implies the range of practical choices for real life institutions. We do not expect to find or develop a measure of value of information, content, or timelieness, that can be used in classical value/cost trade off analysis. |3A6A4

point 2: to provide guidance for the application of the results from existing exploratory and pilot projects. |3A7

To handle this point we must categorize the types of exploratory and pilot projects (and, I would add, preliminary operational experiences) which have been and are being carried on. |3A7A

We should cite specific examples of these categories encountered in our site visit or of which we have learned through publication and other personal experience. |3A7B

We should express our judgment as the adequacy or inadequacy of knowledge or accomplishment in these categories (although I think we must avoid attribution of failure of any specific project and use great care in assigning better success to any specific project over any other one -- conditions for success or failure, however, are necessary items to be treated). |3A7C

We should describe the possible mechanisms for

exploiting results obtained and the strategy of establishing operational capability based on them. |3A7D

Point 3: to identify weak areas in the understanding, design and use of computer systems as needed for effective and efficient information systems. |3A8

In handling this point we have the opportunity to speak to the question of why hasn't progress been made more rapidly and how can it be accelerated. Here is the place to bring in: the non-technological factor of national policies, ownership problem, management problems, total systems design vs. computer process design, how information processing is different than business and scientificculation and records, continuity of files and processes over time and start up transients as complications of the design of the "steady-state" system, the transferability problem and other myths. |3A8A

This section of the report is largely the expression of interpretation and opinion of the panel which form the basis for the strategy of action to get "progress or "results" (i.e., to solve the needs we have accepted as a hypothesis) |3A8B

Point 4: to provide technical guidance for use by planners, system designers and operating officials concerned with information handling systems. Point 5: to recommend national targets for improvements in the weak areas. |3A9

These points are largely the challenge to provide recommendation for national actions. In stating them, the charge to identify specific accomplishments to be sought have been omitted. Perhaps we should recast these points to include that emphasis. |3A9A

Given that Points 1 and 2 have been properly handled, these points should be handled with an attitude of: |3A9B

If we had the administrative decisions and funding decisions to make, what projects would we start or what achievements would we regard as desirable (and reachable), i.e., what would we want OE, NSF, National Commission on Libraries, PSAC, etc. to want to cause to happen and to make appropriate decisions to implement. In making these recommendations we would want other competent technical colleagues to

agree for the most part on the technical validity of the basis for the objectives. |3A9B|

In treating these points, it is acceptable, even desirable to state specific questions to be answered (i.e., that we couldn't answer), hypotheses to be tested (i.e., we have only opinion to offer, and specific numbers and relationships to determine (i.e., specific research, surveys, or analysis to perform). |3A9B2

In summary, we want the report to represent a reasonable management basis for decision and action with a sufficient technical validity of rationale that technical specialists cannot easily invalidate the bases for recommendation. This means that we do not need to spell out all details, but should concentrate on specific and "doable" achievements to be used as guides for actions. |3A9C

Types of decision that this report might help. |4

What R&D work needs to be done. |4A

Should local library consider automation. |4B

Why, i.e., possible benefits. |4B|

Costs. |4B2

Cooperative possibilities. |4B3

Service bureau or public utility concept. |4B4

Myths and realities. |4B5

Changes required in |4C

Federal information policy. |4C|

Consumer habits, i.e., concept of free library. |4C2

Funding structure of libraries and for national information needs. |4C3

Stimulation for change and to inform. |4D

Communications industry. |4D|

Computer hardware community. |4D2

Attitudes of: |4D3

Deployment of resources: |4E

Government granting agencies |4E1

Foundations |4E2

Institutions. |4E3

Common purpose defined for interdependent activities. |4F

Publishing |4F1

Indexing abstracting |4F2

Libraries |4F3

Major decision points in working toward a system. |4G

Continuing organization for national planning of establishing priorities (National Commission on Libraries and Information Systems). |4H

Technical problems |4I

Storage, location, transmission of information. |4I1