Surplus Notes from (7010,) Draft 1, NAS INFOSYS Panel Report

Note: These miscellaneous notes were left over from the 31 Mar, 1 Apr work session (at ARC) of the Informtion Systems Panel of the National Academy of Sciences: Ron Wiggington, Jim Skipper, Joe Eachus, Jack Kettler, and DCE. 1

See (Journal, 7010,) for Draft 1, that was taken back to the CSEB by Wiggington. 1A

<ROW>MISNOTES.RLW;4, 1-APR-71 12:14 BER ; 2

Miscellaneous notes and opinions, RLW 2A

See page 90 (Summary) of Seybold book for some powerful statements relating to: 2B

visions in employment of new technology 2B1

management of it 2B2

e.g., Management readiness is one of the most fascinating imponderables and it is compounded of ignorance, prejudice and inability to formulate clear-cut policy decisions. Much that happens is therefore achieved in spite of management by low ranking technologists who are hardly aware of the implications that paralyze their supervisor -- Lowell Hattery, IEEE Meeting, Washington, 1969. 2B3

If there is only one message that can be transmitted from our assigned task of viewing the technological status and prospects of the computer field as it applies to libraries and the information problem, it is that much adequate technology is already here, and more is coming rapidly -- all we need to do is to learn how to apply it and get on with the design job. 2C

We need only to show that there is enough capability to start and that there is a reliable trend of improvement. We do not need to produce an optimized design. 2D

In the Library field and the abstracting and indexing community, measures of quality or goodness or success are those associated with bigness -- number of volumes or pieces handled. Good measures to judge service are not available. Those that are -- are so context sensitive that comparison is difficult. "Customer satisfaction" is as much a topic of sociology -- and as unpredictable -- as it is of technology or technical adequacy of service capability. Willingness to pay is one indication, but information is a strange commodity and may not be judgeable by normal market indications. 2E

A figure of merit should be a combination of size (# of items), complexity (# of interlinkages), accessibility (i.e., ability to locate items of specific interest without too much scanning "relevance"), reliability (won't miss something, i.e., "recall"), response time (from request to service, single loop time), cost (background and direct charges, user effort). 2E1

Input problem is solvable (technologically) but not solved, i.e., various keyboard equipment, computer aided editing (proofing) seems feasible, some scanners appearing (but too often require retyping using a controlled font e.g., CDC 915). Source data capture is a high hope. However, conflicting (non convertible) interfaces occur, again technologically bridgeable, but multiple efforts in conflict can prevent success. 2E2

The gap between the dreams of the information transfer revolutionary -- i.e., what is conceivable in instant, interactive service and the real world of today's libraries, publication, retrieval methods and services, etc. is an enormous gap. 2F

Most of the activity deals with how better to perform the traditional functions of management and access -- a worthy objective within the state of technology and economically suitable if the opportunities for sharing work (avoiding duplicative processing) are really exploited. However, the work and technology today are indeed inadequate with respect to the storage and transfer of the information itself in digital form. 2F1

Decision support fact retrieval from highly refined simply organized data banks is practical (e.g., stock market). However, the more scholarly the subject and complex the organization of the information and the necessity for the questioner to explore the conceptual space, the farther we are from real computer-aided information systems. 2F2

The phenomenon of technological obsolescence precludes a premature commitment to a particular level of technology. Thus the design must not take such advantage of special features that it is locked in. The problem of transferability is a problem over time as well as progation of use. 2G

Technological obsolescence is bad because a capital expense must be liquidated over time and if a new equipment or

method undersells or overperforms relative to a large committed but unamortized capability, the old capability is not used and there is a financial failure. 2G1

Technological obsolescence has been a very serious factor in the computer field. Is it slowing down now? 2G2

It is a truism that conversion of the "national information system", and library systems in particular; to the dreamed of, instant response, all-encompassing, automated storehouse of man's knowledge will not (cannot) occur in one grand and glorious great leap forward -- even if there was unanimity on such an objective to be reached. 2H

How would it be paid for? Who would have access to it? compare to allocation of resource problem met in a large computing center serving many individuals and projects. 2H1

The C. C. Holt paper contains many interesting comments on the role of information in decision making, categorization of decision making, and merging roles of library and computer center as centers of scholarly activity. Much discussion of systems of standards. Also on page 39 an incisive statement of what can be done. 2I

Those who have the good fortune of having (effectively) unrestricted use of Xerography in management of an activity have experienced the multiplication of ability to handle (printed) information in high volumes quickly. Information can be disseminated rapidly in parallel in an organization, successive copying with marginal notation can speed up the analysis and use of the material. Individuals can review and incrementally change and add to bodies of text (and drawings) without excessive manual retyping, recomposition (in the literary sense, not printing), etc. The desire to make this transmission and manipultion even more responsive with the aid of electronics (communication and computers) is simply an improvement of approach (technique) not a change in basic method insofar as incremental modfication and replication of human interpreted text is concerened. 2J

Extensive use of computer technology in information transfer implies a (near) simultaneous advance in 2K

information available in machine readable form (e.g., as a by product of printing technology) 2K1

organizations to organize and maintain such collection and to provide intellectual access routines to it. 2K2

facilities for moving and delivering that information 2K3

education of a user community in how to use the new capability. (This part of the development equation must not be overlooked -- people, including users are part of the system.) 2K4

The Dial paper on Urban Information Systems contains -- in the introduction -- an excellent analysis of why the fact (reality) of urban information systems does not match the theoretical possibilities and the technological potential. 2L

There is an interaction between information transfer and power or control. That which is now impeded by the "viscosity" and delay of transfer will cause enormous side effects if computer technology succeeds in speeding up the flow and making access easier and more complete. We do not speak to these issues of desirability of improvements (we take them as given) but we note that it may be desirable (necessary?) to exercise discipline and control consciously rather than leaving that to nature. 2M

Therefore, into an optimistic view of what can come about, not only technologically feasible but economically viable if it comes about on the basis of an overall environment, a note of caution must be inserted. Can the U.S. society manage and control the processes made available? The impact of strikes which interrupt the fast response of information and action once a decision has been made, etc. For example, has televison -- the instant communication of information about action at a distance -- made the spread of public unrest and violence more rapid? Can unscrupulous use of the information gathering and transfer mechanism be a danger as great as the benefits envisioned? 2M1

There are three kinds of reactions to the possible threat: 2M2

1) Retreat in horror of the possibility of danger and give up the benefits. 2M2A

2) Ignore the problem and accept whatever happens. 2M2B

3) Educate the population to be able to function in such an environment and include such legal safeguards as

necessary to insure that the individual can defend himself. 2M2C

The first is defeatism, the second is blindness. The only rational alternative is the third. 2M2D

In terms of computer technology (hardware) the single most important item to consider is memory -- capacity, cost, access times, transfer rates,organization (physical and software), type (write only, read only or read/write; digital or image), back up, etc. The elctrical communication technology is coming (or here) with the necessary bandwidths (some cost problems, considerable controversy among common carriers, data utilities, and cable TV), CPU capabilities and speeds are more than adequate and improving constantly (except perhaps for very sophisticated text processing which can be merged into operational activity when practical. The second most important area needing attention is adequate consoles (OK now for minimal typewriter oriented uses, but full character sets and line graphics capabilities are as yet insufficient or too expensive!) at an acceptably cheap price. The electronics capability and economy are here to do the job -- if an agreed and adequate set of technical specification existed. 2N

Caution against the expectation that "The Computer" a deified all powerful mechanism can replace. Thinking, i.e., do more than facilitate the traversing of information paths through an organized universe. 2O

The judgment and selectivity of human intelligence coupled to suitably organized information manipulated by fast automatic machinery is a more complete description of the practical subject of importance. 2O1

Don't forget assimilation and arriving at new insights, conclusions, etc. (i.e., digestion). Getting access is necessary but not sufficient. Getting access to material in a form that can be manipulated -- with automated aids -- is essential for real benefit to users. (from multiple sources) 2P

Kinds of information and corresponding roles: 2Q

Flowing information as control for decision making 2Q1

Scientific and engineering 2Q1A

business 2Q1B

government 2Q1C

and "transactions" as means for 2Q2

executing decisions 2Q2A

ordering materials and services 2Q2B

reporting 2Q2C

and stored information as 2Q3

reservoir of knowledge to be selected from for specific use: 2Q3A

science and engineering 2Q3A1

business 2Q3A2

government 2Q3A3

education 2Q3A4

scholarly activity 2Q3A5

comparison base for judging the accuracy, newness, relevance etc of new or otherwise "flowing information". 2Q3B

In order to make progress in a "future extendable way" it is necessary to have a "future network model" as a long-term goal at a sufficient level of detail that the important design features are available to guide gradual development. I.e., some current design decisions must be made against the future context not rigidly cost optimized today. (Seed money to finance these temporary extra costs is one way to inspire action along these lines). 2R

However, the future cannot come into existence at once or in one push. Today's operational development must be done on accomplishable things but structured in such a way that the future can be built on top of it. 2R1

<ROW>RLWNOTE.NLS;3, 1-APR-71 11:32 BER ; 3

On the problem of framework of study, expected results, strategy of report. 3A

Surplus Notes from (7010,) Draft 1, NAS INFOSYS Panel Report
Wigginton, 31 Mar 71

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. 3B

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. 3C

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. 3D

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. 3E

Point 1: to assess the adequacy of technology to meet the needs of library and information systems at the national level. 3F

To satisfy this point we must: 3F1

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. 3F1A

outline the technology that applies (hardware, software: e.g., input, processors, storage (image, digital), output, distribution and communication, programming, operating systems,data management systems). 3F1B

provide a statement of adequacy or deficiency at present or predictable (perhaps with reflections on past deficiencies ane 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/. 3F1C

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. 3F1D

Point 2: to provide guidance for the application of the results from existing exploratory and pilot projects. 3G

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. 3G1

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. 3G2

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). 3G3

We should describe the possible mechanisms for exploiting results obtained and the strategy of establishing operational capability based on them. 3G4

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

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. 3H1

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) 3H2

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. 3I

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. 3I1

Given that points 1 and 2 have been properly handled, these points should be handled with an attitude of: 3I2

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. 3I2A

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). 3I2B

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. 3I3

<JERNIGAN>ARON.NLS;1, 1-APR-71 18:17 MEJ ; 4

On the "Instabilities of hardware and software". 4A

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. 4A1

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. 4A2

The continuity of business calculations is both in the algorithm 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. 4A3

Surplus Notes from (7010,) Draft 1, NAS INFOSYS Panel Report
Wigginton, 1 Apr 71

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 disimilar hardware but it is also necessary to be able to bridge between successive versions of the operating system for the same hardware. 4A4

The computer industry is far from having compatible operating systems (except where one manufacturer specificaally 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 availablee 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. 4A5

<ROW>ASKIPPER.NLS;2, 1-APR-71 21:49 DCE ; 5

/Note: suggest the following branch replace statement 4I of RLW 8 Mar 71/ 5A

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. 5B

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. 5B1

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. 5B2

<JERNIGAN>CSKIPPER.NLS;1, 1-APR-71 17:28 MEJ ; 6

Types of decision that this report might help. 6A

What R&D work needs to be done. 6A1

Should local library consider automation. 6A2

Why, i.e., possible benefits. 6A2A

Costs. 6A2B

Cooperative possibilities. 6A2C

Service bureau or public utility concept. 6A2D

Myths and realities. 6A2E

Changes required in 6A3

Federal information policy. 6A3A

Consumer habits, i.e., concept of free library. 6A3B

Funding structure of libraries and for national information needs. 6A3C

Stimulation for change and to inform. 6A4

Communications industry. 6A4A

Computer hardware community. 6A4B

Attitudes of: 6A4C

(a) Policy makers 6A4C1

(b) Operating heads. 6A4C2

Deployment of resources: 6A5

Government granting agencies 6A5A

Foundations 6A5B

Institutions. 6A5C

Common purpose defined for interdependent activities. 6A6

Publishing 6A6A

Indexing abstracting 6A6B

Libraries 6A6C

Major decision points in working toward a system. 6A7

Continuing organization for national planning of establishing priorities (National Commission on Libraries and Information Systems). 6A8

Technical problems 6A9

Storage, location, transmission of information. 6A9A

<ENGELBART>MEJDSKIPPER.NLS;1, 12-MAY-71 10:06 DCE ; 7

Digestion Problem. 7A

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. 7A1

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. 7A2

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. 7A3

Multiple Use 7B

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. 7B1

Cataloging copy prepared by the Library of Congress is distributed to approximately 20,000* subscribers through the annual sale of over 80 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. 7B2

* NDB Editor - see latest LC Annual Report. 7B2A

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. 7B3

Page 20 - reference to Harvard. 7B4

Harvard University Widener Library - A very large major university library which has used computer technology to produce a book-form shelf-list. 7B4A

(The Problem of Library Budgets and Their Structure) 7C

Many libraries may find it difficult to fund improvements 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. 7C1

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. 7C2

<ENGELBART>MEJCEACHUS.NLS;1, 12-MAY-71 10:05 DCE ; 8

The Case for Microform. 8A

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. 8A1

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

the page was ppinted. 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. 8A2

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. 8A3

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. 8A4

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. 8A5

Of the various required and desired properties of microform images and their embodiment, no pair is contradictory, and the following set encompasses them all. 8A6

(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. 8A6A

(2) The medium which bears the image must be of very long life. 8A6B

(3) The image must be as small as possible, consistent with the resolution requirements. 8A6C

(4) Images must be contained in such a way as to facilitate their being located and moved by machine. 8A6D

(5) A machine must exist which can locate and move images. 8A6E

(6) A machine must exist which will make copies retaining all the above properties except for some small degradation of image incurred by copying. 8A6F

The Case Against Microform 8B

production of microform images reaquires the allocation of funds, equipment, and manpower which are in short supply. 8B1

There are not formal specifications for film and equipment to produce images with the required properties. 8B2

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. 8B3

<ENGELBART>NASREC.NLS;6, 13-APR-71 11:50 DCE ; 9

**** 9A

Aborted, earlier start: 9A1

d. C. Engelbart 25 February 1971 9A1A

Recommendations 9A1B

In something as complex as The Library System therer are problems and possibilities in abundance--and when explored through its interior labyrinth by concerned, knowledgeable and practical people from different disciplines, the interlinkage of problems of convention, prior investment, sheer bulk, user preferences, gadget ***promoters, financial squeeze, un-businesslike planning frameworks, budget squeezes, over-burdened operational staff, etc. repeatedly entrap and defeat expeditions sent into map a campaign. 9A1C

Our expedition traversing the terrain rapidly and looking for the technical problems of supporting a campaign, concludes that at the "tactical level" of supporting actual Library operhtions environment produced by the current "strategic" framework is not appropriate for taking advantage of a "mechanized assault." 9A1D

Miscellaneous nwe notes: 9B

Considerations of the site visits the Panel made: 9B1

How do we integrate our observations and assessments into the report? 9B1A

What sort of framework do we take for this purose? 9B1B

If, indeed, we stick with the "Market-Development" framework, then it puts this assessment into a special light: 9B1B1

Most of the things we saw would actually not be practical for the solution of an individual library's operation. And, in the "Network-Marketplace" sort of solution, their approaches would be evaluated much differently. 9B1B1A

For example, the cost of the computer programming, the cataloging, the storage space, etc. would be prorated over many users in the marketplace. 9B1B1B

But, must also realize that the validity of some of the approaches would be affected by the way in which this marketplace will (likely) be constituted: 9B1B1C

For instance, most likely will have a standard cataloguing base, and special forms of cataloguing (e.g. as used in INTREX) would call for a large-scale modification in library standards, while other forms could be selectively (optionally) added where user organizations chose. 9B1B1C1

Could list off the different projects we studied, assessing their relvance, spceial significance, etc., with respect to the hypothesized "market environment." 9B1B2

BTL circulation-control system: 9B1B2A

Would think that a central service organization could offer this sort of service to any library -- assuming that standard document (biblio) description was used ... 9B1B2A1

What other assumptions and conditions?? 9B1B2A1A

Etc... (INTREX, TIP, Stanford, Chicago, ARC, NO Times -- the visited ones -- plus perhaps commenting on other well-publicized systes such as ORBIT, DIALOG, etc.) 9B1B2B

Surplus Notes from (7010,) Draft 1, NAS INFOSYS Panel Report

<JOURNAL>7011.NLS;1, 22-MAY-71 0:11 DCE ; Title: Author(s): Douglas C. Engelbart/DCE; Keywords: nas; Clerk: DCE;
Origin: <ENGELBART>NASMISCAPRIL2.NLS;8, 22-MAY-71 0:08 DCE ;
Surplus Notes from (7010,) Draft 1, NAS INFOSYS Panel Report

DCE 22-MAY-71 11:58 7012

Draft 1 of Final Report, NAS INSOFYS Panel

Exported version, taken off by Wigington, had Header of, "NAS/CLR Panel Report: Draft, RLW 2 April 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 yndustries, 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 $509,800,000. 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 188,000,000 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 1968 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 toally 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 burea 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-orinted 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 biliographic access of industrial and other libraries speicalizing 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. 5O

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 amalgamate 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 1968 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 algorithm 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 disimilar 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 specificaally 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 availablee 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

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 faill 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, s9nce 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 pnotographic image of the page rather than on the page itself. In particular, if the pages of a volume have been photograpnned 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 reaquires 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

(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

DCE 22-MAY-71 11:58 7012

Draft 1 of Final Report, NAS INSOFYS Panel

<JOURNAL>7012.NLS;1, 22-MAY-71 12:00 DCE ; Title: Author(s): Douglas C. Engelbart/DCE; Keywords: nas; Clerk: DCE;
Origin: <ENGELBART>ADRAFT.NLS;12, 22-MAY-71 11:22 DCE ; NAS INFOSYS-Panel Draft 1 of Final Report ; Exported version, taken off by Wigington, had Header of, "NAS/CLR Panel Report: Draft, RLW 2 April 71". Draft developed during Panel session at SRI, 31 Mar and 1 Apr 71

NAS INFOSYS-PANEL Draft 1, APPENDIX A: CONCLUSIONS AND RECOMMENDATIONS

Note: This material was witten by me (Engelbart, updated from my January Positon Paper, and was taken by Wiggington on April 2, from our Panel working session of the prior two days, along with the body of Draft 1 (7012,), as a candidate adjunct to the Draft. 1

INTRODUCTION 2

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: 2A

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. 2A1

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. 2A2

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. 2A3

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. 2A3A

Thus, technology will likely open the floodgates to the

production of a vastly greater volume and diversity of recorded material. 2A4

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.) 2B

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. 2C

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. 2D

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? 2E

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. 2F

TIMELY SOLUTION REQUIRES SIMPLE PLAN, BUT BOLD AND MASSIVE 3

It really does seem to me that there is a basic consideration evident here, in the Library Problem: 3A

There is a very complex interlocking among the economic, political, traditional, attitudinal etc. factors of the problem. 3A1

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, 3A2

To make these many changes an adjustments within a complicated Solution-framework would multiply horribly the difficulty of implementing the Solution. 3A3

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. 3A4

The "Bold and Massive" aspects would be necessary, I feel, from both psychological and technical considerations. 3B

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. 3B1

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. 3B2

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. 3C

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. 3D

LIBRARY DIRECTORS AREN'T THE ONES TO BEAR THE SOLUTION-BURDEN 4

Where is the "industry" that can provide "architectural" and "builder" services? 4A

Whose responsibility is to develop library automation? 4B

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: 4C

Library managers are appointees whose operating budget is an appropriation, within an institution (university, county, corporation). 4C1

They are not trained to innovate new products, build new markets, etc. 4C2

They have no Board of Directors and stockholders to decide on a long-term plan and commit resources. 4C3

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? 4D

IS REORGANIZING THE LIBRARY INDUSTRY A VALID POSSIBILITY? 5

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). 5A

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? 5B

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. 5B1

Tthe term "information-utility industry" is brought forth regularly for analogy in discussions of this sort. To risk

old-natting 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. 5B2

Similarly, an informaion 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. 5B3

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. 5B3A

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.) 5B3B

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. 5B3B1

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. 5B4

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. 5B4A

COMPUTER NETWORKS, UTILITIES, AND RESOURCE SHARING ARE THE SOLUTION FOR SERVICING LIBRARY-SYSTEM AUTOMATION 6

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: 6A

Bringing the costs down, and 6A1

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.) 6A2

Bringing costs down will in turn come from two things: 6B

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.) 6B1

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

microfilming, computer-form transcription, computer typesetting etc. 6B2

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. 6C

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. 6C1

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. 6D

Some other needs: logistics, inventory, intelligence, personnel, purchasing, publishing, merchandising. 6D1

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. 6D2

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. 6E

In flexibility, range of services and customers, costs, competitive evolution, etc. 6E1

INTERACTIVE COMPUTER AIDS IMPORTANT, BUT ONLY IN CERTAIN AREAS OF LIBRARY AUTOMATION, AND MUCH PUSH ON THEM ANYWAY FOR OTHER USES 7

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. 7A

BUT: 7B

one doesn't use his private automobile for all of the transportation sevices, and 7B1

interactive computer services won't be used for all of the information manipulation services, either. 7B2

Ships, trains, trucks, airplanes, buses, etc. are an extremely important part of the transportation industry, and 7C

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. 7D

VALUE TO BREAK-THROUGH TECHNOLOGY, BUT VAST IMPROVEMENT POSSIBLE WITHOUT IT 8

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. 8A

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. 8B

COMPUTER-SYSTEMS INDUSTRY NEEDS SIGNIFICANT MATURATION 9

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." 9A

Consider the industry that exists to provide new buildings: 9B

There are skilled, experienced firms of architects who can work with a client to develop an overall plan to suit his

needs -- where specal attention and interactive dialogue is given to the the "user features," i.e. to the features that the building's users are aware of. 9B1

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. 9B2

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. 9C

Most architects (of computer sytems) want the buildng job, too, or are even actually the sellers of the concrete and steel as well. 9C1

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. 9C2

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. 9D

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. 9D1

For instance, there must evolve cleaner conceptualization for system functions and system levels; and both pogramming 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. 9E

MICROFORM FOR GENERAL WORKING HARD COPY 10

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: 10A

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, rearding computer technology's effects upon the quantity of information. 10A1

These strongly imply that full-sized paper copies must be superseded by smaller, cheaper physical copy (microform) -- or ultimately, by electronic storage. 10B

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. 10C

A TRIAL SHOT AT A "SIMPLE, BOLD, MASSIVE PLAN 11

General Considerations: 11A

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. 11A1

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. 11A2

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. 11A3

Specific Recommendations: 11B

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. 11B1

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. 11B1A

produce a standard catalog description of every useful library item, in computer form, and make the basic records available in some open, wholesale, manner. 11B1A1

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. 11B1A1A

produce a high-quality photographic master of every useful library item, and make reproductions available in some open, wholesale manner. 11B1A2

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. 11B1A2A

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. 11B1B

This is directly supportive of, but not equivalent to, nor limited to support of, what is currently called "Library Networking". 11B1B1

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.). 11B1B2

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. 11B1B3

It would be reasonable to give public support to DMN, with the same sort of justification as for highways, waterways, harbors and airports. 11B1B4

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. 11B1B5

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.) 11B1C

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. 11B1D

Comments on the above recommendations: 11C

Copyrights -- 11C1

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, 11C1A

it is obviously of very large economic value to have this access be high speed and economical. a 11C1B

The speed and cost of accessing full-sized, hard-copy information forms are hopelessly outclassed by such as micro-image or digital forms. 11C1C

It seems obvious that new solutions will have to be found for the control and economic return associated with usage of privately created information. 11C1D

No matter what significantly improved information system is developed. 11C1D1

System-Development effectiveness: 11C2

It is an old, nearly-true statement that "You can do anything you want with a computer, if you can only describe it properly." 11C2A

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. 11C2B

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. 11C2C

Marketplace 11C3

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. 11C3A

As it is now, in order to take a significant step toward "modernization", he is faced with a very unsatisfactory situation. :e has really 11C3B

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.) 11C3B1

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. 11C3C

The architect of his choice needs a lexible range of products and services to design into the system. 11C3D

The architect needs to have a flexible market of builders to select from. 11C3E

The range of building jobs (in nature and quantity) needs to support a healthy market of architects and builders. 11C3F

They need to have a healthy, active community of suppliers and support services (sub-contractors, etc.). 11C3F1

When the library has to install all of its own compyter equipment, operate a sophisticated system

(economically), as well as create a drastic change in its regular 1103G

DCE 22-MAY-71 12:15 7014

NAS INFOSYS-PANEL Draft 1, APPENDIX A: CONCLUSIONS AND
RECOMMENDATIONS

<JOURNAL>7014.NLS;1, 22-MAY-71 12:16 DCE ; Title: Author(s): Douglas C.
Engelbart/DCE; Keywords: nas; Clerk: DCE;
Origin: <ENGELBART>ADOUG.NLS;8, 22-MAY-71 12:13 DCE ;
NAS INFOSYS-Panel Draft 1, APPENDIX A: CONCLUSIONS and
RECOMMENDATIONS

NAS INFOSYS-Panel Draft 1, APPENDIX B: Miscellaneous Contributions

Note: 1

See (7012,) for body of Draft 1, which together with APPENDIX A (7014,) was taken back to the CSEB by Wiggington. 1A

These miscellaneous notes were left over from the 31 Mar, 1 Apr work session (at ARC) of the Informtion Systems Panel of the National Academy of Sciences: Ron Wiggington, Jim Skipper, Joe Eachus, Jack Kettler, and DCE. 1B

<JERNIGAN>CSKIPPER.NLS;1, 1-APR-71 17:28 MEJ ; 2

What R&D work needs to be done. 2A

Should local library consider automation. 2B

Why, i.e., possible benefits. 2B1

Costs. 2B2

Cooperative possibilities. 2B3

Service bureau or public utility concept. 2B4

Myths and realities. 2B5

Changes required in 2C

Federal information policy. 2C1

Consumer habits, i.e., concept of free library. 2C2

Funding structure of libraries and for national information needs. 2C3

Stimulation for change and to inform. 2D

Communications industry. 2D1

Computer hardware community. 2D2

Attitudes of: 2D3

(a) Policy makers 2D3A

(b) Operating heads. 2D3B

NAS INFOSYS-Panel Draft 1, APPENDIX B: Miscellaneous Contributions
SKIPPER: Types of decision that this report might help

Deployment of resources: 2E

Government granting agencies 2E1

Foundations 2E2

Institutions. 2E3

Common purpose defined for interdependent activities. 2F

Publishing 2F1

Indexing abstracting 2F2

Libraries 2F3

Major decision points in working toward a system. 2G

Continuing organization for national planning of establishing priorities (National Commission on Libraries and Information Systems). 2H

Technical problems 2I

Storage, location, transmission of information. 2I1

<ROW>RLWNOTE.NLS;3, 1-APR-71 11:32 BER ; 3

On the problem of framework of study, expected results, strategy of report. 3A

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. 3B

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. 3C

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. 3D

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. 3E

Point 1: to assess the adequacy of technology to meet the needs of library and information systems at the national level. 3F

To satisfy this point we must: 3F1

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. 3F1A

outline the technology that applies /hardware, software: e.g., input, processors, storage /image, digital/, output, distribution and communication, programming, operating systems,data management systems/. 3F1B

provide a statement of adequacy or deficiency at present or predictable /perhaps with reflections on past deficiencies ane 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/. 3F1C

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. 3F1D

Point 2: to provide guidance for the application of the results from existing exploratory and pilot projects. 3G

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. 3G1

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. 3G2

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). 3G3

We should describe the possible mechanisms for exploiting results obtained and the strategy of establishing operational capability based on them. 3G4

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

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. 3H1

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) 3H2

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. 3I

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. 3I1

Given that Points 1 and 2 have been properly handled, these points should be handled with an attitude of: 3I2

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. 3I2A

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). 3I2B

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. 3I3

NAS INFOSYS-Panel Draft 1, APPENDIX B: Miscellaneous Contributions
SKIPPER: NAS/CLR Note D -- 31 Mar 71

<ENGELBART>MEJDSKIPPER.NLS;1, 12-MAY-71 10:06 DCE ; 4

Digestion Problem. 4A

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. 4A1

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. 4A2

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. 4A3

Multiple Use 4B

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

DCE 22-MAY-71 12:32 7015

these programs constitute cost savings and improved efficiency for every library in the country. 4B1

Cataloging copy prepared by the Library of Congress is distributed to approximately 20,000* subscribers through the annual sale of over 80 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. 4B2

* NDB Editor - see latest LC Annual Report. 4B2A

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. 4B3

Page 20 - reference to Harvard. 4B4

Harvard University Widener Library - A very large major university library which has used computer technology to produce a book-form shelf-list. 4B4A

(The Problem of Library Budgets and Their Structure) 4C

Many libraries may find it difficult to fund improvements 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. 4C1

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. 4C2

<ENGELBART>MEJCEACHUS.NLS;1, 12-MAY-71 10:05 DCE ; 5

The Case for Microform. 5A

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. 5A1

First, it may be desired to preserve the information content of a page beyond the lifetime of the paper on which the page was ppinted. 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. 5A2

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. 5A3

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. 5A4

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. 5A5

Of the various required and desired properties of microform images and their embodiment, no pair is contradictory, and the following set encompasses them all. 5A6

(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. 5A6A

(2) The medium which bears the image must be of very long life. 5A6B

(3) The image must be as small as possible, consistent with the resolution requirements. 5A6C

(4) Images must be contained in such a way as to facilitate their being located and moved by machine. 5A6D

(5) A machine must exist which can locate and move images. 5A6E

(6) A machine must exist which will make copies

retaining all the above properties except for some small degradation of image incurred by copying. 5A6F

The Case Against Microform 5B

Production of microform images reaquires the allocation of funds, equipment, and manpower which are in short supply. 5B1

There are not formal specifications for film and equipment to produce images with the required properties. 5B2

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. 5B3

<JERNIGAN>ARON.NLS;1, 1-APR-71 18:17 MEJ ; 6

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. 6A

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. 6B

The continuity of business calculations is both in the algorithm 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. 6C

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 disimilar hardware but it is also necessary to be able to bridge between successive versions of the operating system for the same hardware. 6D

The computer industry is far from having compatible operating systems (except where one manufacturer specificaally 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 availablee 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. 6E

<ROW>MISNOTES.RLW;4, 1-APR-71 12:14 BER ; 7

See page 90 (Summary) of Seybold book for some powerful statements relating to: 7A

visions in employment of new technology 7A1

management of it 7A2

e.g., Management readiness is one of the most fascinating imponderables and it is compounded of ignorance, prejudice and inability to formulate clear-cut policy decisions. Much that happens is therefore achieved in spite of management by low ranking technologists who are hardly

aWare of the implications that paralyze their supervisor -- Lowell Hattery, IEEE Meeting, Washington, 1969. 7A3

If there is only one message that can be transmitted from our assigned task of viewing the technological status and prospects of the computer field as it applies to libraries and the information problem, it is that much adequate technology is already here, and more is coming rapidly -- all we need to do is to learn how to apply it and get on with the design job. 7B

We need only to show that there is enough capability to start and that there is a reliable trend of improvement. We do not need to produce an optimized design. 7C

In the Library field and the abstracting and indexing community, measures of quality or goodness or success are those associated with bigness -- number of volumes or pieces handled. Good measures to judge service are not available. Those that are -- are so context sensitive that comparison is difficult. "Customer satisfaction" is as much a topic of sociology -- and as unpredictable -- as it is of technology or technical adequacy of service capability. Willingness to pay is one indication, but information is a strange commodity and may not be judgeable by normal market indications. 7D

A figure of merit should be a combination of size (# of items), complexity (# of interlinkages), accessibility (i.e., ability to locate items of specific interest without too much scanning "relevance"), reliability (won't miss something, i.e., "recall"), response time (from request to service, single loop time), cost (background and direct charges, user effort). 7D1

Input problem is solvable (technologically) but not solved, i.e., various keyboard equipment, computer aided editing (proofing) seems feasible, some scanners appearing (but too often require retyping using a controlled font e.g., CDC 915). Source data capture is a high hope. However, conflicting (non convertible) interfaces occur, again technologically bridgeable, but multiple efforts in conflict can prevent success. 7D2

The gap between the dreams of the information transfer revolutionary -- i.e., what is conceivable in instant, interactive service and the real world of today's libraries,

publication, retrieval methods and services, etc. is an enormous gap. 7E

Most of the activity deals with how better to perform the traditional functions of management and access -- a worthy objective within the state of technology and economically suitable if the opportunities for sharing work (avoiding duplicative processing) are really exploited. However, the work and technology today are indeed inadequate with respect to the storage and transfer of the information itself in digital form. 7E1

Decision support fact retrieval from highly refined simply organized data banks is practical (e.g., stock market). However, the more scholarly the subject and complex the organization of the information and the necessity for the questioner to explore the conceptual space, the farther we are from real computer-aided information systems. 7E2

The phenomenon of technological obsolescence precludes a premature commitment to a particular level of technology. Thus the design must not take such advantage of special features that it is locked in. The problem of transferability is a problem over time as well as progation of use. 7F

Technological obsolescence is bad because a capital expense must be liquidated over time and if a new equipment or method undersells or overperforms relative to a large committed but unamortized capability, the old capability is not used and there is a financial failure. 7F1

Technological obsolescence has been a very serious factor in the computer field. Is it slowing down now? 7F2

It is a truism that conversion of the "national information system", and library systems in particular; to the dreamed of, instant response, all-encompassing, automated storehouse of man's knowledge will not (cannot) occur in one grand and glorious great leap forward -- even if there was unanimity on such an objective to be reached. 7G

How would it be paid for? Who would have access to it? compare to allocation of resource problem met in a large computing center serving many individuals and projects. 7G1

The C. C. Holt paper contains many interesting comments on the

role of information in decision making, categorization of decision making, and merging roles of library and computer center as centers of scholarly activity. Much discussion of systems of standards. Also on page 39 an incisive statement of what can be done. 7H

Those who have the good fortune of having (effectively) unrestricted use of Xerography in management of an activity have experienced the multiplication of ability to handle (printed) information in high volumes quickly. Information can be disseminated rapidly in parallel in an organization, successive copying with marginal notation can speed up the analysis and use of the material. Individuals can review and incrementally change and add to bodies of text (and drawings) without excessive manual retyping, recomposition (in the literary sense, not printing), etc. The desire to make this transmission and manipultion even more responsive with the aid of electronics (communication and computers) is simply an improvement of approach (technique) not a change in basic method insofar as incremental modfication and replication of human interpreted text is concerened. 7I

Extensive use of computer technology in information transfer implies a (near) simultaneous advance in 7J

information available in machine readable form (e.g., as a by product of printing technology) 7J1

organizations to organize and maintain such collection and to provide intellectual access routines to it. 7J2

facilities for moving and delivering that information 7J3

education of a user community in how to use the new capability. (This part of the development equation must not be overlooked -- people, including users are part of the system.) 7J4

The Dial paper on Urban Information Systems contains -- in the introduction -- an excellent analysis of why the fact (reality) of urban information systems does not match the theoretical possibilities and the technological potential. 7K

There is an interaction between information transfer and power or control. That which is now impeded by the "viscosity" and delay of transfer will cause enormous side effects if computer

technology succeeds in speeding up the flow and making access easier and more complete. We do not speak to these issues of desirability of improvements (we take them as given) but we note that it may be desirable (necessary?) to exercise discipline and control consciously rather than leaving that to nature. 7L

Therefore, into an optimistic view of what can come about, not only technologically feasible but economically viable if it comes about on the basis of an overall environment, a note of caution must be inserted. Can the U.S. society manage and control the processes made available? The impact of strikes which interrupt the fast response of information and action once a decision has been made, etc. For example, has televison -- the instant communication of information about action at a distance -- made the spread of public unrest and violence more rapid? Can unscrupulous use of the information gathering and transfer mechanism be a danger as great as the benefits envisioned? 7L1

There are three kinds of reactions to the possible threat: 7L2

1) Retreat in horror of the possibility of danger and give up the benefits. 7L2A

2) Ignore the problem and accept whatever happens. 7L2B

3) Educate the population to be able to function in such an environment and include such legal safeguards as necessary to insure that the individual can defend himself. 7L2C

The first is defeatism, the second is blindness. The only rational alternative is the third. 7L2D

In terms of computer technology (hardware) the single most important item to consider is memory -- capacity, cost, access times, transfer rates,organization (physical and software), type (write only, read only or read/write; digital or image), back up, etc. The elctrical communication technology is coming (or here) with the necessary bandwidths (some cost problems, considerable controversy among common carriers, data utilities, and cable TV), CPU capabilities and speeds are more than adequate and improving constantly (except perhaps for very sophisticated text processing which can be merged into operational activity when practical. The second most

important area needing attention is adequate consoles (OK now for minimal typewriter oriented uses, but full character sets and line graphics capabilities are as yet insufficient or too expensive!) at an acceptably cheap price. The electronics capability and economy are here to do the job -- if an agreed and adequate set of technical specification existed. 7M

Caution against the expectation that "The Computer" a deified all powerful mechanism can replace. Thinking, i.e., do more than facilitate the traversing of information paths through an organized universe. 7N

The judgment and selectivity of human intelligence coupled to suitably organized information manipulated by fast automatic machinery is a more complete description of the practical subject of importance. 7N1

Don't forget assimilation and arriving at new insights, conclusions, etc. (i.e., digestion). Getting access is necessary but not sufficient. Getting access to material in a form that can be manipulated -- with automated aids -- is essential for real benefit to users. (from multiple sources) 7O

Kinds of information and corresponding roles: 7P

Flowing information as control for decision making 7P1

Scientific and engineering 7P1A

business 7P1B

government 7P1C

and "transactions" as means for 7P2

executing decisions 7P2A

ordering materials and services 7P2B

reporting 7P2C

and stored information as 7P3

reservoir of knowledge to be selected from for specific use: 7P3A

science and engineering 7P3A1

business 7P3A2

government 7P3A3

education 7P3A4

scholarly activity 7P3A5

comparison base for judging the accuracy, newness, relevance etc of new or otherwise "flowing information". 7P3B

In order to make progress in a "future extendable way" it is necessary to have a "future network model" as a long-term goal at a sufficient level of detail that the important design features are available to guide gradual development. I.e., some current design decisions must be made against the future context not rigidly cost optimized today. (Seed money to finance these temporary extra costs is one way to inspire action along these lines). 7Q

However, the future cannot come into existence at once or in one push. Today's operational development must be done on accomplishable things but structured in such a way that the future can be built on top of it. 7Q1

<ROW>ASKIPPER.NLS;2, 1-APR-71 21:49 DCE ; 8

/Note: suggest the following branch replace Statement 4I of RLW 8 Mar 71/ 8A

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. 8B

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. 8B1

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. 8B2

<ENGELBART>NASREC.NLS;6, 13-APR-71 11:50 DCE ; 9

**** 9A

Aborted, earlier start: 9A1

d. C. Engelbart 25 February 1971 9A1A

Recommendations 9A1B

In something as complex as The Library System therer are problems and possibilities in abundance--and when explored through its interior labyrinth by concerned, knowledgeable and practical people from different disciplines, the interlinkage of problems of convention, prior investment, sheer bulk, user preferences, gadget ***promoters, financial squeeze, un-businesslike planning frameworks, budget squeezes, over-burdened operational staff, etc. repeatedly entrap and defeat expeditions sent into map a campaign. 9A1C

Our expedition traversing the terrain rapidly and looking for the technical problems of supporting a campaign, concludes that at the "tactical level" of supporting actual Library operhtions environment produced by the current "strategic" framework is not appropriate for taking advantage of a "mechanized assault." 9A1D

NAS INFOSYS-Panel Draft 1, APPENDIX B: Miscellaneous Contributions
Engelbart: Garbagy residue

Miscellaneous new notes: 9B

Considerations of the site visits the Panel made: 9B1

How do we integrate our observations and assessments into the report? 9B1A

What sort of framework do we take for this purose? 9B1B

If, indeed, we stick with the "Market-Development" framework, then it puts this assessment into a special light: 9B1B1

Most of the things we saw would actually not be practical for the solution of an individual library's operation. And, in the "Network-Marketplace" sort of solution, their approaches would be evaluated much differently. 9B1B1A

For example, the cost of the computer programming, the cataloging, the storage space, etc. would be prorated over many users in the marketplace. 9B1B1B

But, must also realize that the validity of some of the approaches would be affected by the way in which this marketplace will (likely) be constituted: 9B1B1C

For instance, most likely will have a standard cataloguing base, and special forms of cataloguing (e.g. as used in INTREX) would call for a large-scale modification in library standards, while other forms could be selectively (optionally) added where user organizations chose. 9B1B1C1

Could list off the different projects we studied, assessing their relvance, spceial significance, etc., with respect to the hypothesized "market environment." 9B1B2

BTL circulation-control system: 9B1B2A

Would think that a central service organization could offer this sort of service

to any library -- assuming that standard document (biblio) description was used ... 9B1B2A1

What other assumptions and conditions?? 9B1B2A1A

Etc... (INTREX, TIP, Stanford, Chicago, ARC, NO Times -- the visited ones -- plus perhaps commenting on other well-publicized systes such as ORBIT, DIALOG, etc.) 9B1B2B

NAS INFOSYS-Panel Draft 1, APPENDIX B: Miscellaneous Contributions

<JOURNAL>7015.NLS;1, 22-MAY-71 12:33 DCE ; Title: Author(s): Douglas C. Engelbart/DCE; Keywords: nas; Clerk: DCE;
Origin: <ENGELBART>NASMISCAPRIL2.NLS;15, 22-MAY-71 12:29 DCE ;
NAS INFOSYS-Panel Draft 1, APPENDIX B: Miscellaneous Contributions

NP Note about Journal-entry process

The process of getting a message or file sent off to the Journal now takes too much of the sender's time to bring out the kind of use for our Dialogue System that we envision in effective collaborative dialogue. Here are some thoughts I'd like to register in this regard, even assuming that it's quite likely that they are already in WSD's NP stack. 1

Would like an easy way for submission-control information to be supplied from user's file data. 2

e.g, from special-syntax leading-text data of the statement heading the submission packet 2A

How about a submission process like: "E J S1 CA CA" whereupon The Journal Processor: 3

interprets the statement, 3A

inserts JOURNUM DATE:TIME, 3B

Journal Processor might keep a small stack of ready numbers available, so that it can be immediately responsive if number-assignment feedback is desired. 3B1

copies into its domain the specified file material, all or parts of which may be specified by LINKs, and be from other files, and 3C

returns control to the user. 3D

The user's involvement in the Journal-entry process is now over. The Journal system executes the rest of its process as a separate fork. 3D1

The user needs comfortable assurance that this material really is safely held within the Journal files. 4

The Journal system could perhaps be implemented in such a way that before returning control to the user it can take such precautions as thorough Verify-File process, make severl copies (perhaps go so far as to reserve a tape transport just to add these raw-input items into for added safety). 4A

Perhaps, when the journal-entry process was all finished, the Journal Processor could insert a special "journal-entry verification" statement in his private "file-data control file." Consider the following (optional) conventions and service: 4B

If, for User ABC, there is a file under his name called "FMABC", for File Management file of user ABC., and 4B1

If there is a statement in this file with a (some standardized) name, "JOURentryRECORD", then 4B2

After completing all of its entry work and verification for Journal Entry nnnn, the Journal Processor will insert its verification entry as a statement named "Vnnnn" (for "Verification on entry NNNN") in the subplex under (JOURentryRECORD). 4B3

Included in this verification entry is a link back to the Journal-Insertion Control Statement that launched the entry-nnnn process. 4B4

After the Journal-entry request was trasacted with the user, thee Journal-entry Processor could insert the link (ABC,FMABC,Vnnnn:g) along with the entry number and date/time. When the user might run cross this material later, and wonder if he can now safely delete it, he can simply jump on this link to see if the entry process hs been completed and verified. 4B5

He'd generally wait to delete his entry-source material until he received this verifying input. 4C

He might even (someday) feel safe in having a process which deletes his source material autaomaticaly, after the full imput process is completed. 4C1

DCE 22-MAY-71 12:49 7016

NP Note about Journal-entry process

<JOURNAL>7016.NLS;1, 22-MAY-71 12:49 DCE ; (Expedite) Title: Author(s): Douglas C. Engelbart/DCE; Distribution: William S. Duvall, James C. Norton/WSD JCN; Clerk: DCE;

DCE 22-MAY-71 13:18 7017

To Local Board No. 80, re. C.H. Irby, from D.C. Engelbart, 5 April 71

Letter mailed to CHI's draft board to support his reclassification appeal

Augmentation Research Center
Stanford Research Institute
Menlo Park, California
94025
April 5, 1971 .LMS= : 1

Local Board No. 80
Selective Service System
124 West Figueroa Street
Santa Barbara, California
93104

(Regarding Mr. Charles H. Irby, 04-80-45-348) 2

3
Gentlemen: 4

This letter concerns two features regarding the work that Mr. Irby is engaged in at Stanford Research Institute. 5

1) Relevancy to National Defense in particular, and to national goals and problems in general: 6

About the work, in brief: 6A

We are involved in an experiment that has no equivalent anywhere in the world. It has been sponsored by the Department of Defense on a continuing basis for eight years, currently at the rate of $1.2 million/year. It is aimed at developing new tools, new methods, new organizational forms, etc., that will enable teams to be significantly more effective

at coping with problems of great complexity and/or urgency. The work heavily involves the use of computers, for manipulating text and diagrams in working notes, team dialogue, designs, plans, goals, etc. -- using computers in ways much different from their past applications. 6A1

Our approach involves using an experimental team that applies our experimental tools and methods in doing its every-day work. (Mr. Irby plays a central role in this team.) The team is engaged in a very complex task in which there is usually a good deal of pressure. The team members not only have to be very skillful, high-quality professionals in the realm of the team's work (chiefly, computer-system development), but they also have to be trained in the new roles and methods of working, and in the use of the new computer tools with which they constantly work throughout their days. They also have to be unusually flexible in adapting to the continual change of their working mode and environment, and it requires a high degree of "uncertainty tolerance" and team spirit to be able to operate well under these conditions: i.e. working very hard under constantly shifting circumstances, where

snafus in experimental tools or methods are a constant deterrent (it is not unlike trying to do intense professional work while taking care of the four kids). 6A2

About its national-defense relevance: 6B

After eight years of slow, often frustrating progress in basic techniques, our program is now launching into a new phase where the techniques we have developed for our own use will be shared with (tested by) other teams around the country, working on complex computer-research projects. One of these teams, at the Air Force's Rome Air Developement Center, will be evaluating our techniques for use in Air Force management situations -- such as for a staff team in support of the Air Force Chief of Staff at the Pentagon, or for large-projct managemnt, for logistics control (e.g., the Air Force has over $20 billion in equipment and parts, whose inventory must stay balanced, and that has to be gotten to the right places at the right time), etc. 6B1

The logistics branches of both the Army and the Navy have shown high interest, as well as have the new-systems people concerned with other sorts of military operations,

such as intelligence analysis, or command and control. 6B2

As an instance of potential applicability that is unusual in its specific place of application, but quite typical in the type of value offered to the users, we were recently visited by a man responsible for operating the "Situation Room" in the White House, to which the President and his top staff turn for immediate information on any world situation. It was quite clear from his reaction to our tools and techniques that there would be very valuable application possibilities in his environment, within a few years when we have learned more about training different kinds of teams, and have developed our tools and techniques for a broader range of applications. 6B3

For instance, even in the relatively fundamental team task of jointly developing a "team document" (plan, design, report, etc.), our tools would obviously be of considerable value -- they would have allowed, for instance, the recently issued President's State of the World Message to have been developed in considerably shorter time than the "approximatly four months" that it did take (providing for all of the staff

contributions, reviewing, re-writing, etc.). The value of simply cutting down the elapsed time, in getting the complex and urgent business of an executive staff done, is obviously very high in situations like this. 6B3A

About its more-general relevance to problems of national interest: 6C

Our work is also directly applicable to the complex problems faced in such as cancer research, urban planning management, and regional development, stability analysis for national economics, etc. -- the world of today is literally overwhelmed with problems where complexity and urgency demand team efforts of maximum effectiveness. 6C1

2) Difficulty in staffing such work: 7

As mentioned above, a person in the core of our experimental team has to be a very skillful professional, at the very forefront of "deep computer science," and also must be highly adaptable to a shiftng work environment and capable of working under sustained conditions of pressure and "frontier stresses." It is only in the younger group of computer-science professionals that we seem to be able to find suitable

candidates for such work. To add a new person to our core staff, we screen many applicants very carefully, and then must train our choice for almost a year even though he has the highest of technical qualifications. 7A

Mr. Irby is a key person in our core staff not only because of his role and his professional competence, but because he has an unusual mixture of perception regarding the human-team aspects of our problem, and because he demonstrates a very high level of dedication to the goals and progress of our program. Losing him would be a severe blow -- it would be a least a year and a half before we could balance the team again, if we were lucky enough to be able to do so. 7B

I am the manager of this research program. I view with respect and appreciation the problems and responsibilities faced by a draft board today. I wouldn't trouble you with a letter such as this if I didn't feel strongly and seriously that it indeed would be to the best interest of our national defense to keep Mr. Irby on this project. 8

Sincerely yours, 9
10

DCE 22-MAY-71 13:18 7017

To Local Board No. 80, re. C.H. Irby, from D.C. Engelbart, 5 April 71

Douglas C. Engelbart, Manager
Augmentation Research Center 11

DCE 22-MAY-71 13:18 7017

To Local Board No. 80, re. C.H. Irby, from D.C. Engelbart, 5 April 71

<JOURNAL>7017.NLS;1, 22-MAY-71 13:19 DCE ; Title: Author(s): Douglas C. Engelbart/DCE; Clerk: DCE;
Origin: <ENGELBART>CHIDRAFT.NLS;9, 22-MAY-71 13:14 DCE ;

JCN 22-MAY-71 14:46 7018

Proposal for Changes to the ARC TENEX File Group Write Access Configuration

Proposal for Changes to the ARC TENEX File Group Write Access Configuration

Proposal for changes to TENEX file group write access 1

When PDP-10 NLS was first put into operation the browse mode feature was not yet implemented. Users experienced some difficulty when viewing others' files, particularly when using L10 analyzer-formatter programs, even though they did not wish to change the files permanently. 1A

Therefore, TENEX user and directory grouping options were set to give read-write access to many users, including access to directory-only user names, such as NLS, Journal, etc. 1B

Now that browse mode is available in both TNLS and DNLS, we are finding that many of the present read-write accesses are causing files to be locked unnecessarily. 1C

This is further complicated by the current problem with the handling of locked files where NLS gives the user the message "file locked by XXX" when he loads a file locked by another user, but does not permit the user to continue viewing the file. 1C1

This will be corrected soon, leaving only the problem of files being locked that should not be locked, due to write accesses that are not really necessary. 1C2

I propose that the following file group write accesses be implemented by Ken Victor in place of all current ones: 1D

NLS and REL-NLS 1D1

Melvin, Victor, Hopper, Paxton, Irby, Church, Andrews, Duvall, Bass, Lehtman, Parsley 1D1A

META 1D2

Andrews 1D2A

PORGEN 1D3

Bass, Parsley, Irby 1D3A

JOURNAL and AJOURNAL 1D4

Duvall, Lehtman 1D4A

NIC 1D5

Watson, North, Norton 1D5A

MSR 1D6

Norton, Parsley 1D6A

NET 1D7

Melvin 1D7A

CATALOG 1D8

None (clerical process will be performed after connecting to CATALOG's directory to prevent accidental changing of Master Catalog Files) 1D8A

Others? 1D9

Note that no "people" usernames are permitted direct write access to other "people" usernames' files. 1E

Should anyone who does not have write access for another user's files have a (legitimate) need to write on any of them, he can connect to that user's directory, if he has the user's password. 1F

WSD note: Since all users need to create Journal files when entering Journal documents, Journal subsystem programs should be changed to temporarily connect the user to Journal's directory, to offset the removal of his access due to the above proposed changes. 1G

I would appreciate comments on this proposal by Wednesday, May 26th 5:00 PM, at which time I intend to schedule these changes with Ken Victor and Ed VandeRiet. 1H

JCN 22-MAY-71 14:46 7018

Proposal for Changes to the ARC TENEX File Group Write Access Configuration

<JOURNAL>7018.NLS;1, 22-MAY-71 14:46 JCN ; (Expedite) Title: Author(s): James C. Norton/JCN; Distribution: Walter L. Bass, Roger D. Bates, Mimi S. Church, William S. Duvall, Douglas C. Engelbart, Martin E. Hardy, Fred P. Hocker, J. D. Hopper, Charles H. Irby, Mil Jernigan, Harvey G. Lehtman, John T. Melvin, Jeanne B. North, Cindy Page, Bruce L. Parsley, William H. Paxton, Jeffrey C. Peters, Jake Ratliff, Barbara E. Row, Ed K. Van De Riet, Dirk H. Van Nouhuys, Kenneth E. Victor, Don I. Andrews, Richard W. Watson, Marilyn F. Auerbach, Beauregard A. Hardeman/WLB RDB MSC WSD (note the changes since you saw this) DCE MEH FPH JDH CHI MEJ HGL JTM JBN CXP BLP WHP JCP JXR BER EKV DVN KEV DIA RWW MFA BAH; Keywords: ; Clerk: JCN;

DCE 23-MAY-71 9:26 7019

Visitor Log: Donald Streeter, IBM Yorktown, Director of Computing Systems Department

Donald N. Streeter, Director, Computing Systems Dept.
Research Division
Thomas J. Watson Research Center
Post Office Box 218
Yorktown Heights, New York 10598
(914) 945-1758. 1

He is organizing a esssion on "Systems, Man and Cybernetics" or something, in which to consider the computer as part of the system. 2

I told him that I couldn't spare the time to consider participating (although I wasn't sure whether or not he was asking me toconsider it), but that I'd like to help locate possible participants or comment on topics and plans. 2A

Also, he is spending about 20% of his time working on a monograph that has tentativelybeen ad for inclusion in IM's planned series on progrmming (Bob Glaser was the editor in chief, but he has left IBM). Don's monograph is on computer usage in re as it has been experienced at yorktown. 3

Talked of managing computer resources -- accounting, scheduling, priorities and billing. 4

He said they are currently implementing a system at Yorktown, for much the same reasons as motivate us at ARC to consider billing our own people -- that the resource utilization can only grow to have a balanced cost/benefit profile among a system of tool if the developers have an awarenss of the resources cnsumption under use. 4A

Mentioned Prof. Norman Nielsen, of the Business School at Stanford (also assoc. director of the computer center) who published a paper on "allocation and pricing policies" in the communications of the ACM, Aug 70. Said that this paper was extrememly useful to them, and that they had imported Nielsen as a consultant when they were formulating their own plan. 4B

IBM's Research Lab at Yorktown may join the ARPA Net. 5

I described some of the BC notions, and told him that he would be a very likely guy for me to start pursuing with participation possibilities someday. 6

Gave him copies of the 1970 RADC and NASA reports, and biblkography sheets. 7

DCE 23-MAY-71 9:26 7019

Visitor Log: Donald Streeter, IBM Yorktown, Director of Computing Systems Department

<JOURNAL>7019.NLS;1, 23-MAY-71 9:26 DCE ; Title: Author(s): Douglas C. Engelbart/DCE; Clerk: DCE;

JTM 24-MAY-71 8:51 7020

IMP Interface Purchase (Preliminary)

We should get together and talk about this

JTM 24-MAY-71 8:51 7020

IMP Interface Purchase (Preliminary)

Discussion with BBN IMP people at Spring Joint concerning buying an IMP interface 1

I spoke to both Bob Kahn and Frank Heart on 18 May about buying a BBN style IMP interface which sits on the PDP 10 I/O buss 1A

TIME ESTIMATE: 4-5 weeks from firm order 1A1

COST ESTIMATE: 11.5K +/- 1K 1A2

According to Frank, the interface is considered a standard production item (i.e. no problem to build except for finding technician time) and comes with complete documentation 1B

JTM 24-MAY-71 8:51 7020

IMP Interface Purchase (Preliminary)

<JOURNAL>7020.NLS;1, 24-MAY-71 8:51 JTM ; (Expedite) Title: Author(s): John T. Melvin/JTM; Distribution: Ed K. Van De Riet, Roger D. Bates/EKV RDB; Keywords: ; Clerk: JTM;

DVN 24-MAY-71 11:30 7021

Access to write on impersonal files

Replie to (Journal,7018,)

DVN 24-MAY-71 11:30 7021

Access to write on impersonal files

I need to write on NLS files to continue to add carifying comments to the NLS Status file. I have written on MSR files regularlly in creating and updating baseline records. My guess is others have and that MSR should be completely open. 1

2

DVN 24-MAY-71 11:30 7021

Access to write on impersonal files

<JOURNAL>7021.NLS;1, 24-MAY-71 11:30 DVN ; (Expedite) Title: Author(s): Dirk H. Van Nouhuys/DVN; Distribution: James C. Norton, Bruce L. Parsley/JCN BLP; Keywords: groups access; Clerk: DVN;
Origin: <VANNOUHUYS>JOURDRAFT.NLS;1, 24-MAY-71 11:22 DVN ;

KEV 24-MAY-71 13:34 7022

notes on changes to group configuration (journal, 7018,)

KEV 24-MAY-71 13:34 7022

notes on changes to group configuration (journal, 7018,)

Note that the subsstem, SNDMSG, which allows sending messages to people that will be seen at login time, only allows you to send messages to members of your own group. Thus, for all practical purposes, if we adopt this new grouping, we loose SNDMSG. 1

2

in addition to the groups proposed in (journal,7018,) i would like to propse the following groups for "SYSTEM" files: 3

3A

user SYSTEM, SRI-DLM, KEV-SRI, BBN-ORG 3B

3B1

Victor, Melvin, Hopper, Irby 3B2

3C

SUBSYS 3D

3D1

Victor, Melvin, Paxton, Irby, Bass, Duvall 3D2

3E

SYSUTILITY, EXEC, SOURCES 3F

3F1

Victor, Melvin 3F2

3G

KEV 24-MAY-71 13:34 7022
notes on changes to group configuration (journal, 7018,)

<JOURNAL>7022.NLS;1, 24-MAY-71 13:35 KEV ; (Expedite) Title: Author(s): Kenneth E. Victor/KEV; Distribution: Walter L. Bass, Roger D. Bates, Mimi S. Church, William S. Duvall, Douglas C. Engelbart, Martin E. Hardy, Fred P. Hocker, J. D. Hopper, Charles H. Irby, Mil Jernigan, Harvey G. Lehtman, John T. Melvin, Jeanne B. North, James C. Norton, Cindy Page, Bruce L. Parsley, William H. Paxton, Jeffrey C. Peters, Jake Ratliff, Barbara E. Row, Ed K. Van De Riet, Dirk H. Van Nouhuys, Don I. Andrews, Richard W. Watson, Marilyn F. Auerbach, Beauregard A. Hardeman/WLB RDB MSC WSD DCE MEH FPH JDH CHI MEJ HGL JTM JBN JCN CXP BLP WHP JCP JXR BER EKV DVN DIA RWW MFA BAH; Keywords: groups; Clerk: KEV; Origin: <VICTOR>KEV.NLS;2, 24-MAY-71 13:29 KEV ;

DCE 25-MAY-71 8:36 7023

Contact Report: DCE visit to Loren Bright, NASA Ames

On 18 May 71 I went to Ames Lab to visit: 1

Mr. Loren Bright,
Drector, Research Support
NASA Ames Research Center
Moffett Field, Calif. 94036
961-1111, ext 2685 1A

I went to begin exploring the possiblities for our augmentation techniniques being used to support the documentation activities of ILLIAC IV's operations. 2

I presented to him a general descriptive outline that is equivalent to (NABC, NNNN). 3

He seemed to feel that this might be interesting to the ILLIAC people at Ames. He isn't directly enough associated with the details to provide much evaluative information, or to give much evaluative judgement about the potential value to Ames. 4

Mel Pirtle (From Berkeley Computer Corporation) has just recently been hired by Ames to head up their ILLIAC-IV operation -- which will probably be a eparrate "Division" under Bright. Pirtle was out of town at the time. Bright and I deferred any further discussion for a meeting between Pirtle and me. 5

DCE 25-MAY-71 8:36 7023

Contact Report: DCE visit to Loren Bright, NASA Ames

<JOURNAL>7023.NLS;1, 25-MAY-71 8:36 DCE ; Title: Author(s): Douglas C. Engelbart/DCE; Clerk: DCE;

Meeting Log: Committee A, 24 May 71

Attendees: RDB DCE CHI JTM JCN BLP WHP EKV RWW 1

Discusion Notes: 2

After discussing the news brought back from the meetings of the Network Working Group, Dick outlined the task topics which seemed important to him with respect to meeting our NIC developments. 2A

We knocked thse items around a bit, then DCE circulated a printout of the "Stages" branch below. The approach that was established by meeting end was: 2B

DCE will concentrate heavily upon formulating and promoting the Stage 1 Plan (including proposal to ARPA). 2B1

About Stage 1: 2B2

JCN will be chief pusher. 2B2A

He is to be the chief catalyst, coordinator, balancing negotiator, facilitator, or etc., responsible for keeping the planning action moving. Like: 2B2A1

Does every guy have his contract capacity filled; does every task have a pusher, and a buyer; 2B2A1A

And, no planning item should go unquestioned regarding such attributes as: who the buyer is; what requirements it is to meet; who will see that its planned implementation will meet them; who made the estimates on resources; what alternate-task possibility exists that might have higher relative payoff to us; 2B2A1B

He doesn't particularly do the balancing, except in areas where he is the natural coordinator/pusher (like for financial coordination, MSR pushing, clerical-support, etc.) 2B2A2

But he sees to it that the right person does get connected with an unbalanced situation for the purpose of resolving it. 2B2A2A

BLP will devote full time to helpin Jim in Stage 1; result of Stage 1 may well be a plan for Stage 2 that puts Bruce back to work as a programmer. 2B2A3

Early action should include developing mutually agreed proposal for roles and assignemnts thereto for covering the various coordinating (advisory, surveillance, ...) roles. Check finally with DCE, and then slip on our new roles and try them out for a few months. 2B2B

During this stage, RWW should make himself available to help wherever feasible. 2B2C

DCE rung in periodically for reviewW, and on call whenever special need arises. 2B2D

Baseline portrayal to be by pencil or chalk or whatever best serves to expedite the process. 2B2E

About Stage 2: 2B3

Most of the details should come out of Stage 1. 2B3A

Stages 3

Set up ARC planning structure by outlining four time-specified stages: 3A

Stage 1: Immediate near future, lasting a week or two 3B

Stage 2: Into mid or late summer 3C

Giving NIC its big kickoff push, while also providing the planning (trial designs, etc.) and preparatory work for later stages. 3C1

Important items to include, besides direct and early NIC developments: 3C2

ARC organization, management, accounting 3C2A

Operations 3C2B

Hard-copy local-printout system 3C2C

File-library system 3C2D

Stage 3: Up to Feb 1972 Duration of current contract) 3D

Wrapping up some specifics with respect to the proposed work of team augmentation, Network particiption, and NIC development and operation. 3D1

Orient as much as possible to support an optimum transition into Stage 4. 3D2

Stage 4: Next contract period, and beyond 3E

Continue with ARC-internal, Team-Augmentation bootstrapping. 3E1

Continue with NIC development and service 3E2

Develop operational capability (computer services, development support for extended NLS tools, accounting, billing, scheduling, training, documenting, etc.) to support other groups using NLS for 3E3

development of: 3E3A

computer-systems documentation, and 3E3A1

RINS data-base 3E3A2

and use of the facilities and data for selective activities in the system-development domain. 3E3B

DGE 25-MAY-71 12:45 7024

Meeting Log: Committee A, 24 May 71

<JOURNAL>7024.NLS;1, 25-MAY-71 12:46 DCE ; (Expedite) Title: Author(s): Douglas C. Engelbart/DCE; Distribution: Marilyn F. Auerbach, Walter L. Bass, Roger D. Bates, Mimi S. Church, William S. Duvall, J. D. Hopper, Charles H. Irby, Harvey G. Lehtman, John T. Melvin, Jeanne B. North, James C. Norton, Bruce L. Parsley, William H. Paxton, Ed K. Van De Riet, Richard W. Watson, Don I. Andrews, James A. Fadiman/MFA WLB RDB MSC WSD JDH CHI HGL JTM JBN JCN BLP WHP EKV RWW DIA JAF; Clerk: DCE;

HGL 25-MAY-71 13:36 7026

New Note on Journal

Supercedes earlier note

HGL 25-MAY-71 13:36 7026

New Note on Journal

It is no longer necessary to run RECOVERNLS if there is a crash while the Hard Copy Journal Distribution operation is being run. It will be handled henceforth automatically by the new monitor. This note replaces the information contained in (journal, 6915,0:). 1

HGL 25-MAY-71 13:36 7026

New Note on Journal

<JOURNAL>7026.NLS;1, 25-MAY-71 13:37 HGL ;Title: Author(s): Harvey G. Lehtman/HGL; Distribution: William S. Duvall, James C. Norton, Barbara E. Row, Cindy Page/WSD JCN BER CXP; Keywords: journal, distribution,documentation; Clerk: HGL;

WSD 26-MAY-71 12:00 7027

DSS tasks

these will be entered and updated in baserec following our meeting today

DSS tasks

Note on time estimates 1

The estimates given are guesses for 'good working days', which means that slop must be added for meetings, bad days, etc. 1a

Over the long run, I would probably suggest a factor of 2 1b

Tasks for DSS 2

Groups in the identification system (2 days) 2a

Put an identification system submode into NLS (1 day) 2b

/Done/ Run JCN 110 program on saveit. 2c

/Done/ Fix ↑C so it is caught by Journal, and it aborts gracefully. 2d

/Done/Change the Journal Hard-copy distribution program so it: 2e

Sends copies to all of the authors (this may be done when building the distribution file) /Done/ 2e1

Address Journal copies, vis. Master Collection, Access Collection, Englebart Collection, Duvall Collection. /Done/ 2e2

/Done/ Change Link parsing to allow "--" as left delimiter. 2f

Allow user to specify sub-collection membership when entering document into Journal. 2g

Modify Journal to accomodate NIC. 2h

Include sub-collection information into Identification Record of user. 2i

Consider preliminary design of citation building system. 2j

Write program which converts Journal header statements into proper format for master catalog, and make corresponding entries. /Done/ 2k

/Done/Propose stage 0 file system (hopefully to coincide with stage 0 NIC) 2l

Impelement stage 0 file system (20 days) 2m

Propose Master Catalog Organisation 2n

Design long-range file system (10 days) 2o

/Done/Change update file to copy file for Journal 2p

/Done/Study reliability problem with respect to Journal 2q

Special attention to losing of number file /Done/ 2q1

Design set system with BLP (see--Journal, 6207:gw) (see--Journal, 6983:gw) (5 days) 2r

Fix up submodes to be usable from display NLS, specifically (5 days) 2s

Journal system 2s1

Collector/Sorter 2s2

Catalog Number system 2s3

Identification system 2s4

Implement Groups in identification system (2 days) 2t

Improve Journal response 2u

Background Processes (15 days) 2u1

segmenting of JCAT and CNUMBERS /Done/ 2u2

Re-organising of Journal File manageement (5 days) 2u3

Improving efficiency of string constructio stuff in L10 ??? 2u4

Impproving effeciency of file opening in TENEX ??? 2u5

Consideration o alternatate file mechanisms in TENEX for handling Journal files ??? 2u6

Implement on-line distribution (10 days) 2v

On-line dist for NIC (5 days) 2v1

/Possibly/ implement automaic RFC numbers (1 day) 2w

Change catalog numbering system to accept DATE TIME IDENT /SITE/ (10 days) 2x

Design and Implement Mail system (10 days)	2y
Back links (15 days)	2z
Stage 0 NIC work (5 days)	2a*
Journal System Maintenance (Continuing task)	2aa
Fix H.C. Distribution to use new output Processor (1 day)	2aa1
Figure out how to expunge deleted Journal Files	2aa2
Devise (if possible) a method for using other directories or taking corrctive action when Journal's directory becomes full (2 days)	2aa3
Change Journal to Connect to Journal directory when operating (involves TENEX diddling)	2aa4
Bring up system with automatic running of Recovernls (Ken's almost got this done, I think)	2aa5
Develop automatic way of moving Journal files to another directory, and updating jcat correspondingly. (1 day)	2aa6
Developing easy way to lock users from Journal for maintenance. (1 day)	2aa7
Journal Quality Assurance	2ab
A continuing task whenever a new system or output processor or ...is brought up.	2ab1
Secondary Distriution of Journal Documents. (5 days)	2ac
["[Done]"]:	3

WSD 26-MAY-71 12:00 7027

DSS tasks

<JOURNAL>7027.NLS;1, 26-MAY-71 12:02 WSD ; (Expedite) Title: Author(s): William S. Duvall/WSD; Distribution: James C. Norton, Richard W. Watson, Douglas C. Engelbart, Bruce L. Parsley, Charles H. Irby, William H. Paxton, Mimi S. Church/JCN RWW DCE BLP CHI WHP MSC; Keywords: dss tasks; Clerk: WSD;

DCE 29-MAY-71 8:47 7018

Note on Journal's Printout Directives, a problem in 7017

Bill: Journal printout of (Journal, 7017,) never dididid recover from an early directive in its body that set the left margin way over to the right in order to inset the return address of a letter. It was originally (Engelbart, Chidraft,), which printed all right. A slight new adjustment to Journal-directive conventions, huh? Sorry to be the bearer, etc. Doug. 1

DCE 29-MAY-71 8:47 7048

Note on Journal's Printout Directives, a problem in 7017

(J7048) 29-MAY-71 8:47; (Expedite) Title: Author(s): Douglas C. Engelbart/DCE; Distribution: William S. Duvall, Walter L. Bass, James C. Norton/WSD WLB JCN; Clerk: DCE;