A PUBLIC ACCESS INFORMATION SYSTEM FOR THE HUDSON RIVER FOUNDATION

by

Richard S. Marcus

Room 35-414
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA 02139

January 20, 1986 **M.I.T. Report LIDS-R-1531**

This report was prepared for the Hudson River Foundation under grants 3/84X/4 and 4/84X/4 and constitutes the combined final report for these grants.

Hudson River Foundation
for Science and Environmental Research
122 East 42nd Street
New York, NY 10168

Hudson River Foundation Report HRF/GR-86/01

50272-101

REPORT DOCUMENTATION PAGE	1. REPORT NO. HRF/GR - 86/01	2.	3. Recipient's Accession No.

4. Title and Subtitle
A Public Access Information System for the Hudson River Foundation

5. Report Date Jan. 20, 1986 (Date of preparation)

6.

7. Author(s)
Richard S. Marcus

8. Performing Organization Rept. No.
LIDS - R - 1531

9. Performing Organization Name and Address

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA 02139

10. Project/Task/Work Unit No.

11. Contract(C) or Grant(G) No.
(C) 3/84X/4 and
(G) 4/84X/4

12. Sponsoring Organization Name and Address

Hudson River Foundation for Science and Environmental Research
122 East 42nd Street, New York, N.Y. 10168

13. Type of Report & Period Covered
Final Report: 4/84-12/85

14.

15. Supplementary Notes

16. Abstract (Limit: 200 words)

The Hudson River Foundation seeks to develop a Public Access Information System (PAIS) which would enable access to environmentally-related information concerning the Hudson River by both professionals (e.g., researchers) and the public at large. Easy, user-friendly access to summary, index, and catalog information from scattered, remote sites is desired. To investigate means to achieve these goals experiments were undertaken using an intermediary retrieval assistance system, CONIT, which is being researched at the Massachusetts Institute of Technology. Selected people from the Hudson River community were given the opportunity to use CONIT to access primarily bibliographic data on both Hudson-River specific and other topics from over 300 databases on 3 retrieval systems. Over 40 search sessions and 20 questionnaire reponses were analyzed. Results indicated that the intermediary assistance technique was effective for making bibliographic access more convenient and showed prospects for aiding access to other kinds of information (e.g., numerical, messages) as well. However, cost-benefits considerations led to the conclusion that Foundation resources not be expended at this time in the development of computerized systems but rather that existing systems should be utilized in the development of a PAIS with mixed computer and more standard manual modes.

17. Document Analysis a. Descriptors

Information retrieval; Information systems; Computers; Experimental design; Experimental data; Systems analysis; Data retrieval; Documents

b. Identifiers/Open-Ended Terms

CONIT; Environmental Information Systems; Intermediary Search Assistance Systems; Hudson River

c. COSATI Field/Group 5B, 88B, 44G, 44T

18. Availability Statement	19. Security Class (This Report)	21. No. of Pages
Release unlimited	unclassified	37
	20. Security Class (This Page) unclassified	22. Price

TABLE OF CONTENTS

Tables

1. Goals of the Hudson River Foundation

The Hudson River Foundation [HRF85] is an endowed public foundation which sponsors scientific, economic, and public policy research and educational programs on matters of environmental, ecological, and public health concern relevant to the Hudson River. One of the functions of the Foundation is to develop ways to manage information about the Hudson River's environmental characteristics, resources, uses, and dynamics. The Foundation wishes to ensure ready access to all relevant scientific, economic and public policy information, whether or not such information is generated by projects under its own sponsorship. The Foundation is investigating two approaches to information management, a Public Access Information System and a Scientific Database Information Management System.

In the Public Access Information System, the Foundation seeks to develop a system which has summary, index, catalogue and inventory functions and which provides an overview of information in the most general sense (who, what, where, when). The Public Access Information System will have remote-access capabilities and will be highly "user-friendly". In general, the objective is to provide for easy access to information on the Hudson River. It is expected that users of the Public Access Information System will be individuals who need information about the Hudson in order to make decisions or as a basis for research, teaching, or other activities. These users may not necessarily be trained in the use of computers or have a knowledge of computer languages.

With the Scientific Database Information Management System (SDIMS), the Foundation seeks to develop a system for organizing and storing scientific databases relating to the Hudson River and environs. We expect SDIMS to be used by people with training and experience in quantitative methods and the use of computers. However, since different disciplines may be accustomed to different computer languages, this system will also be "user-friendly" to the extent possible.

2. Research Goals and Objectives of This Investigation

This investigation sought seven major objectives:

(1) Determine the extent to which newly devised intermediary computer network techniques can serve as the basis for information systems for the Hudson River Foundation.

(2) Explore the possibility that such Hudson River Foundation information systems can serve as a new distributed-database, national-network model for accessing and disseminating multidisciplinary information both intra-group and inter-group in the United States for all sectors of our society.

(3) Develop demonstration systems by which to investigate these concepts experimentally.

(4) Devise experiments to test the efficacy of these concepts.

(5) Run the experiments, analyze the results, and evaluate the concepts.

(6) To the extent that the results of the experiments are positive, leave in place for the Hudson River Foundation a working prototype Public Access Information System.

(7) In any case, devise plans for the further development of Hudson River Foundation information systems, either as a direct extension of the prototype system or with a modified approach dictated by the experimental results.

3. Relevance of Project to Hudson River Foundation

The Hudson River Foundation has properly emphasized in its annual reports (as noted in the above section) the importance of providing access to information about the Hudson River and related matters, including both information developed directly under Foundation sponsorship and information (a considerably larger amount) developed outside such sponsorship. In these same documents the Foundation has wisely indicated its concern that any systems devised to meet these needs must be effective yet within the means of the Foundation's resources and, therefore, take proper cognizance of the possibilities for using or building upon existing systems, rather than developing entirely new systems. The experimental CONIT system developed at M.I.T. under several research grants is illustrative of one such means to utilize existing systems. We have investigated the possibility of capitalizing on these previous major research efforts in order to devise a plan for a Public Access system that meets the Foundations needs.

4. Approach

There are several classes of alternate approaches for providing the information systems desired by the Hudson River Foundation. The first kind of approach is simple and direct: study what is wanted and build a system with the desired features. The problem with such an approach is fairly obvious: it is simply too costly for the Hudson River Foundation to develop major new systems of the kind needed.

Some would counter that the newly emerging microcomputer era offers a means to provide the needed functionality within existing resource limitations. While microcomputers will play an increasingly vital role in the development of information systems in general and, in particular, for the Hudson River Foundation (as we shall indicate below), they are not currently constituted to provide the desired functionality in terms of multiple, simultaneous access to extremely large databases. Even for modest-size databases there have as yet been no adequate solutions to the problems of efficient development and distribution of new and modified software for many microcomputers scattered in the field.

Microcomputers can now, of course, be effectively used to handle *certain functions* needed by the Foundation. The Northstar and IBM PC systems currently being installed at Foundation headquarters should prove very useful for particular word processing and logging activities, among other functions

A second approach to providing the desired large-scale access to databases would be to use existing systems like, for example, DIALOG or SDC ORBIT. There are a number of

problems with this approach. Such systems are fairly difficult to use; they require considerable training and regular practice to master and maintain proficiency. Furthermore, information needed is regularly scattered across databases and systems and the heterogeneity of such databases and systems further exacerbate the learning and access problems.

Another kind of problem with this approach is that the specialized information and functions desired by the Foundation are not easily incorporated. Private-file databases on these publicly accessible systems provide only a partial, and rather costly and awkward, solution. Modifying these very large (and rather antiquated) systems to tailor them to one's particular needs is not much easier than building anew.

The approach we are propounding is to use the newly developed technique of the *intermediary computer system* by which easy and effective access to a *network* of existing multiple, heterogeneous information systems is achieved. Furthermore, the modular, flexible and moderate-sized intermediary system can be run efficiently and modified relatively easily to incorporate specialized Foundation needs. Such an intermediary system, in its ultimate form, may be briefly defined by the functions it should perform: (1) converse with the end user in a simple, easy-to-learn, easy-to-use language; (2) assist the user to formulate his problem and identify the appropriate information resources (e.g., databases) that can be utilized to solve the problem; and (3) automatically connect to these information resources, translate the user's request into the language required to interrogate those resources, and translate the responses to those interrogations into a form appropriate to the user's needs.

On several research projects we have developed experimental intermediary systems, under the generic name CONIT (COnnector for Networked Information Transfer), which we demonstrated are as effective in some contexts as expert, human information-specialist intermediaries for assisting end users in accessing and retrieving information from the several hundred databases of the three major online bibliographic retrieval systems: Lockheed's DIALOG, System Development Corporation's ORBIT, and ELHILL (MEDLINE) of the National Library of Medicine. Our conclusions are based on the analysis of controlled experiments in which we have compared the results of searching done by end users with the CONIT computer intermediary against those achieved by human expert intermediaries searching on the same problems without CONIT. (For details see the Marcus paper [MARC83b].)

Intermediary system technology can solve a good part of the problems encompassed in the concept of a 'Public Access System' and it can also serve to provide a framework for resolving many of the issues involved in the 'Scientific Database System' area. Intermediary systems have just recently reached the operational stage. Two well-known ones are CSIN (Chemical Substances Information Network) of the Environmental Protection Agency [MCGI83] and Sci-Mate of the Institute for Scientific Information [LEFK82]. For both of these our research work served as initial conceptual models. Our enhanced CONIT experimental systems are still in advance of the operational systems in some ways, especially in ease of use and generality of subject area and assistance modes. (For a more complete discussion of the various efforts in this area see paper by Lisanti [LISA84] and two Marcus papers [MARC83b] and [MARC85].)

One example scenario follows. The intermediary system would make possible easy and effective access to existing systems for individuals in the Hudson River Foundation community. Such systems could include those of Lockheed DIALOG, SDC ORBIT, and the National Library of Medicine MEDLINE which contain several hundred databases of general interest including at least a half dozen specifically covering environmental concerns for the Hudson River and elsewhere. The Hudson River Foundation would ensure that information

that it and its grantees developed would be incorporated into appropriate nodes in the information network. Such nodes include the existing databases on the above-mentioned systems, new databases on these systems, databases directly at the intermediary level, and databases at a separate local Hudson River Foundation microcomputer (to which the intermediary could also connect). For example, reports generated by the Hudson River Foundation and its grantees could be incorporated in and retrievable from existing databases: either specifically environmental or energy-related ones like ENVIROLINE (Environment Information Center), ENVIROBIB (Environmental Periodical Bibliography), POLLUTION ABSTRACTS, WATER RESOURCES, ELECTRIC POWER INDUSTRY ABSTRACTS, Aquatic Sciences and Fisheries Abstracts, Oceanic Abstracts, Water Literature, WATERNET, WATER QUALITY, WATER RESEARCH, AIR POLLUTION, DOE ENERGY, Electronic Power Research Institute (EPRI), ENERGYNET, ENERGYLIT, or ENERGYLINE and/or more general ones like the National Technical Information Service (NTIS) reports file. Directories for such information as environmentally concerned community groups in the Hudson River area could be maintained on the intermediary and/or the local computer. There would be the possibility for developing additional databases for the Hudson River (or, say, river basins in general) at various points in the information network.

Once such a scenario was designed and put in place the Hudson River Foundation community would have uniform and effective access to the information (both internally and externally developed) which it needs and other communities would have similar access to information developed by the Hudson River Foundation. As other organizations followed this path, a more effective, efficient, and coordinated information network would be developed with shared costs that would be far superior to the uncoordinated conglomeration of systems toward which we are now trending.

In the earlier (Phase I) work for the Hudson River Foundation analysis was accomplished that reviewed the Foundation's needs for information systems and the corresponding resources available to the Foundation for development and maintenance purposes. This analysis has fortified the opinion that the intermediary network approach is most appropriate to the Foundation's needs.

5. Specific Tasks in Phase II Plan

The basic plan for Phase II was to provide to the Hudson River Foundation an experimental Public Access system deriving from the CONIT intermediary retrieval assistance system. The existing CONIT system already provides effective user-friendly access to a majority of the information in the world's published literature and, therefore, to most of the information desired by the Hudson River Foundation community. Besides the databases specifically mentioned in the scenario above, there are a total of about 300 databases accessible with CONIT including those in the physical sciences including SCIENCE ABSTRACTS (INPSEC), SCIENCE CITATION INDEX, CHEMICAL ABSTRACTS, and SPIN (Physics); in engineering including ENGINEERING INDEX (COMPENDEX) and six metallurgical files; in medical and biomedical areas including MEDLINE (Index Medicus), EXCERPTA MEDICA, BIOSIS (Biology), and five files with both bibliographic *and* numeric data on toxicology; in geosciences including GEOARCHIVE and GEOREF (American Geological Institute); plus numerous files in agriculture, food, business, management, social sciences, politics, humanities, patents, and education.

In addition to access to these many databases, we considered augmentation of the CONIT system in several ways to provide further assistance specifically designed to support the Hudson River Foundation community. Additional files to be created of specific interest and utility to the Foundation were considered. Typical candidate files include: lists of Hudson River Foundation grantees with their addresses and abstracts of their research projects; summary statements of Foundation purposes, research support principles, current needs, and application procedures; lists, contact points, and relations to Hudson River concerns of political and other organizational units; lists of databases within, or external to the Scientific Database with summary descriptions of contents and means of access; and specific summary information, possibly in tabular form, of particularly useful numerical information derived from the Scientific Database.

The initial mode of access could be the one we have successfully employed in providing information to users about the several hundred databases accessible through the existing CONIT: namely, a hierarchically-organized, menu-selected structure. Such an addition would require certain revisions to our current CONIT system. We anticipate special arguments would need to be added to the SHOW command to enable users to display the lists of files and the files themselves. The explanations of how to use these additional features must be incorporated into the intermediary system. Once that is accomplished we have a basically new information system -- an augmented intermediary system -- as an experimental Hudson River Foundation Public Access Information System: one that provides access to the world's literature *plus* access to specifically Hudson River related information *not* (yet) available in the public databases. It may be noted that CONIT has been structured as a rule-based system with emphasis on modularity for ease and flexibility of modification. One would seek to retain these features for the Public Access system so that, in particular, the Foundation itself (or others in its employ besides the M.I.T. developers) may easily modify and add information as desired.

Besides the search retrieval, and display features of the augmented intermediary system itself, we have considered further augmentation of the Public Access system by allowing users to take advantage of other functional capabilities of the host Multics computer system on which CONIT resides. These features include several electronic mail systems and a computer conferencing system. The mail systems will permit the Hudson River Foundation community to communicate with each other via this highly effective medium as well as with any person in the world who has access to any computer with network mail connections to M.I.T. Multics (including hundreds of computers around the world). We plan also to set up at least two computer conferences: one on general issues of water resources and one on concerns specific to the Hudson River Foundation. A computer conference can be thought of as a collective computer mail arrangement in which every message added becomes immediately available to each participant in the conference. The conferencing system allows participants to "attend" or drop out as they see fit, as well as to add items, reply to others, and review any past item when they wish.

The mail and conference systems are state-of-the-art and, as such, are moderately easy to use. We have also considered the possibility of incorporating these functions *within* the intermediary framework and, thus, make them *very* easy to use,

6. Experimental Use of CONIT

6.1 Background and Preparations for Experiments

An experiment was run in which members of the Hudson River Foundation community were given the opportunity to perform searches using the CONIT intermediary system. Preparations for the experiment included setting up computer accounts and the detailed design of the basis for experimental usage. A special account on the CONIT system was set up to allow for access by Hudson River users. A particular user identification name and password were designated for access to MULTICS, the M.I.T. computer, on which CONIT resides. The Hudson River CONIT account was set so that the maximum cost of a single session could not exceed $100. In addition, a Hudson River Foundation account was obtained for the DIALOG retrieval system. The password for this account was entered into the tables for the Foundation CONIT account so that access to DIALOG was then possible through CONIT for Foundation users. Tables in the foundation account were also loaded with identification and password information for M.I.T. accounts for National Library of Medicine ELHILL (Medline) and the Systems Development Corporation (SDC) ORBIT retrieval systems. (Most databases of interest to the Foundation are available on DIALOG; therefore, it was felt not essential to initiate separate Foundation accounts on the other two systems since the extra administrative expense of using the M.I.T. accounts for those systems would be small.)

Other plans for the experiment were made. It was decided to allow use by two groups of people: (1) all those who had research grants from the Foundation (a total of 25 organizations receiving grants were identified with this group) and (2) all members of the Hudson River Panel -- an advisory group to the Foundation (this group totaled 19 with inclusion of the Science Officer of the Foundation).

A packet of instructional material was prepared for each principal investigator or panel member from the two groups. One component of the packet was a cover letter (see Appendix A) explaining the nature of the experiment. It was explained, for example, that each user was to be initially limited to $200 in usage costs but that additional usage could be arranged if the user or his organization could arrange for payment of such usage, or, perhaps the Hudson River Foundation might support, or partially support, additional usage.

Also enclosed in the packet was instruction on how to use CONIT. This included a reference manual, which is essentially a compendium of the online explanation messages from CONIT -- i.e., all the instruction in the manual, plus other instruction, is available to the user online and it is *not necessary* to read the manual -- although *some* users find it helpful. Separate instructional information was inserted in the packet detailing the procedures for accessing MULTICS and CONIT, including procedures for connecting through the TELENET and TYMNET networks -- only one of the potential users was resident at the time in the Boston area where a local telephone call permitted access. Finally, incorporated in the packet was a questionnaire (see Appendix) to be filled out *after* CONIT use was attempted or when it was determined that such usage would *not* be attempted in the framework of the experiment.

6.2 Experimental Data

The instructional packet was mailed to the 44 panel members and grantees in the third week of July, 1984. Most of the usage of CONIT on the Hudson River account was experienced in the period July 24, 1984 through September 25, 1984. We have accumulated statistics of use for this period and present them below. A few subsequent CONIT Hudson River usages were recorded through November, 1985; these will be described afterwards.

Data collected were derived from several sources. Firstly, each CONIT session was fully recorded on the MULTICS computer -- this included all user inputs and CONIT responses (including the translated responses from requests made to the remote retrieval systems). Secondly, there were a few telephone calls *from* users -- or would-be users -- concerning difficulties in connecting to, or operating, CONIT plus a few telephone calls by us *to* users for debriefing purposes. Finally, we have the responses to the questionnaires.

The panel members were not distinguished from the grantees by any significant differences in quantitative or qualitative measures in their searching or responses to the questionnaires. Therefore, we have chosen to characterize the results by considering the experimental base of 44 organizations as a whole without distinguishing the two component bases further.

We identified CONIT uses from 16 different organizations during the stated time period -- i.e., 36% participation by the 44 organizations. In several organizations more than one person participated in the searching, either in the same session or in different sessions. We identified 19 different persons as actually performing the search at the terminal. In addition, it appears that another 3 to 6 persons participated in the searching by being present during the session or directing the nature of the search beforehand.

The number of sessions recorded was 39. Most users had only a single session, a few had two or three sessions, and one user had 8 sessions.

How effective were the searches? Of the 39 total sessions 12 could be characterized as merely "test" sessions. In these sessions users made no apparent attempt to perform any searching as such. In most of these the main purpose seemed to be to test whether access to CONIT could be accomplished at all. In a few of these sessions some explanations of the use of CONIT were elicited, but no attempt to apply the explanations to searching was evident (it was fairly clear that understanding or following the explanations was *not* the problem inhibiting search attempts).

The 27 sessions in which an attempt *was* made to perform retrieval operations exhibited a wide range of search topics. Most of topics were clearly relatable to the central concerns of the foundation; examples: "PCB's in fish", "hatchery trout growth", "effects of aryl phosphates on fish and invertebrates", "phytoplankton in Hudson ecosystem", "Hudson River water quality", "ozone effects on true growth", "biofouling", "air quality models", "subterranean cables", "acid rain effects on streams", "xenon in atmosphere", "pollutant effects on lobsters", "Spottail Shiners", and "trichlorbenzene toxicity". In addition, there were several author searches and several searches on various topics that *could* have direct impact on ecological concerns (e.g., a search on information about a company that presumably had been involved in river pollution) plus a few searches on topics relating to current concerns that would directly affect searchers' work (e.g., searches on information about computer hardware and software).

Of the 27 sessions featuring actual retrieval attempts 21 (78)% were successful in that some document references *were* retrieved. Analyses revealed that of those 21 sessions in which documents were retrieved in 19 cases (90% of the 21 cases and 70% of the 27 cases) at least some of those documents were *relevant* to the user's topic. (No *quantitative* measure of the degree of *usefulness* -- as opposed to mere relevance -- was obtained -- but see below for *number* of documents found.) Of the 8 sessions in which no relevant documents were found analysis showed that in one case (13%) there *were* no relevant documents in the databases, in 3 cases (38%) system bugs prevented successful searching, and in 4 cases (50%) inadequate search strategy was the reason for failure.

Most users searched on only a single topic in any one session; approximately three users searched on two or three topics in a single session. The number of search requests explicitly made by users in one session ranged from 1 to 10; the median being about 2 or 3. Of course, a single CONIT search request on a phrase with n words will result in n+1 actual searches -- one for each word and one for the combination; the median number of actual searches was thus about 5.

Analysis indicates that the number of relevant documents retrieved ranged from one to 100 or more; a median of about 10 relevant documents was observed.

We now turn to costs for running CONIT. Dollar costs are derived from the costs of using the M.I.T. MULTICS computer and the costs of the retrieval systems themselves. Costs for the MULTICS computer typically range from $5 to $20 per hour connected; the variation results from differing charges by MULTICS for different times of day (cheaper on nights and weekends) and for different amounts of computational resources used (e.g. charges per unit of actual processor time and storage or input/output units and for network connections). [NOTE: Charges are quoted based on *direct* costs; for its own facilities and accounts on remote systems which it administers M.I.T. assesses an overhead charge (currently 61.5%) while for remote systems usage *not* on M.I.T. accounts (e.g., the HRF DIALOG account), *no* such charge is made.]

For the commercial retrieval system the cost typically ranges from $50 to $100 per hour, primarily depending on the database in use and network connection. Thus, when intensive use is being made of the retrieval systems themselves, the proportion of the costs for using CONIT itself (i.e., the MULTICS costs) is typically in the range of 10% to 20%. However, to the extent that no other system is connected (as when users are learning CONIT in the first few minutes of a session) the proportion of total costs which relate to MULTICS will rise.

For the experimental usage session costs were observed to range from $1 to $114. Typical session costs were in the range $25 to $50; the minimum cost for a session producing relevant documents was $11.

Session times ranged from 3 minutes to 74 minutes; typical sessions fell in the range 20 to 40 minutes. The minimum session time for finding relevant documents was 15 minutes.

The number of different databases searched was 29. Although most of the searching was done on DIALOG, there was some searching done on the other two systems which have some important databases not available on DIALOG (e.g., TOXLINE); also on occasion a connection to DIALOG may not be possible but the desired database is also on ORBIT or NLM, which may be accessible at that time.

As previously mentioned, after the 39 sessions recorded and compiled under the statistics described above, there have been a few additional uses of CONIT. In particular, we have recorded 11 sessions by users, only one of whom had previously used CONIT during the intensive two-month experimental period. Two of these users did not get beyond testing connecting to CONIT and making a few requests for general information about using CONIT. The other two users did make effective utilization of the system. One performed a search for literature written by a given author and for documents citing the work of that author (and the searcher himself). The other had, over the course of two weeks, 8 sessions covering a variety of topics with many relevant documents found.

Before evaluating these experimental results we should summarize the responses to the questionnaires. Twenty questionnaires were filled out and returned to us from 18 different organizations. Fifteen of the respondents had run CONIT sessions; almost all of the organizations involved in the searching did also, then, return a filled-out questionnaire. Five respondents had no CONIT sessions; reasons for failure to use CONIT included: lack of available equipment with which to make access (2 respondents), lack of time to make the attempt (2 respondents), and lack of interest (1 respondent).

There was a variety of terminal/computer/modem equipment used in, and more so available for, access to CONIT. All but two respondents had access available via local (as opposed to long-distance) telephone calls.

Respondents reported the number of potential users of CONIT at their respective organizations in the range of 1 to 200 with a median of 12. The number of databases available through CONIT of potential interest to the respondent, was reported in the range of 1 to 60 with a median of 10. The expected number of usages was in the range of 15 per week to one per year; an individual (personal) expected usage of once per month was typical.

A wide variety of information was of interest to the respondents in terms of computerized retrieval. Emphasis was placed on literature about the Hudson River and by Hudson River area researchers, in particular access to the often hard-to-find so-called "gray" (or "fugitive" or "non-published" -- e.g., report) literature. Mention was also made of summary information, bibliographies, database directories, and data on (Hudson River) water quality, flow, temperature, larva, and estuarine species; and museum specimens. Only a few responses were made to the questions about the desirability of computer functions other than retrieval from databases: 4 respondents wanted the ability to perform messaging, conferencing, and/or polling, and one mentioned (improved) computation and analysis.

6.3 Experimental Evaluations

We would characterize the overall results of the experimental operation of CONIT in a generally positive fashion. Most users were able to obtain information they needed fairly quickly and easily. The comments made by users generally in a range that could be described as "moderately favorable" to "enthusiastic". At the most positive end of the spectrum were several comments stating that the new technology represented by CONIT would now permit them (the respondents) and other retrieval-system inexperienced (e.g., non-professional) users to perform searching more effectively and quickly and without having to resort to expert human intermediaries.

User-stated concerns were recorded in three areas: (1) there were some bugs in the CONIT system that caused unreliable operations on occasion; (2) some of the explanations

given by CONIT were either unclear or overly wordy; and (3) there are (or should be) easier or better or cheaper mechanisms for retrieving information.

Another negative aspect to the experiment was what we would consider the relatively low usage made of CONIT. As we pointed out before, fewer than 40% of the organizations given the opportunity to use CONIT actually did so in the two-month intensive part of the experimental period. Actually that amount of usage could be considered fairly high under normal circumstances; however, the strong urging given by the Foundation to its grantees and panel members might have been expected to lead to a greater amount of "free" usage. What is even more significant is the low usage in the 14 months following the initial period: only 4 users, only two of which did substantive searches and only one of which was a repeat user from the initial period. These actual usage statistics contrast with the much higher "expected usage" reported on the questionnaires.

We do not have detailed data from non users to confirm the reasons for the relatively low usage of CONIT. However, we can identify a number of reasons which, in some combination, most likely account for the bulk of the cause for low usage.

One reason is the unavailability of computer or terminal equipment by which to access the M.I.T computer. Even for those who theoretically have such equipment available, there is a major obstacle for many in *taking advantage* of those facilities if they do not regularly use that equipment; learning "strange" equipment and procedures is a barrier. Even though CONIT *itself* may be very easy to use, *access* to it can be (perceived as) difficult for the uninitiated without something like a smart autodialer modem.

Of course, a lack of interest or felt need for bibliographic information would inhibit use; although from the perspective of professional or personal need one could argue that most persons *should* search regularly, to a large body of those persons attempts to perform such searches are observed more in the breach. The Hudson River Foundation could serve itself and society by both *providing* means to do this information retrieval and *encouraging* its community to do so. In this regard it is not clear the extent to which additional CONIT usage might have been elicited by stronger statements from the foundation that such usage was being encouraged beyond the initial experimental period.

Finally, information systems users will naturally gravitate to those systems that they feel are easiest to use within their budgets. If some have convenient access to a good search service, especially one that does not charge directly on a per use basis, they may prefer to maintain that mode of information access. Also, admittedly, the current experimental, and still buggy nature of CONIT could be much improved.

7. Coverage of Online Databases and Bibliographies

One question we sought to investigate was where bibliographic references of interest were stored and available. Two classes of "interest" can be distinguished: (1) documents specifically mentioning the Hudson River with respect to environmental concerns and (2) other documents of interest to people involved in such concerns (especially where that interest relates specifically to those concerns -- as, for example, where the document discusses a pollutant that is of concern in the Hudson River without mentioning or considering Hudson effects specifically). Let us call class (1) "Hudson Specific" and class (2) "general".

While the Hudson specific documents obviously form a very important core of documents of interest, they are only a small fraction (we estimate less than 1%) of all documents of interest. The documents actually *sought* are also, although to a much lesser extent, predominantly in the general category. To support this latter statement we see that only a few of the searchers in our CONIT experiment actually requested Hudson-specific documents.

Nevertheless, as we have said, the Hudson-specific documents do form an important core and we have considered this core in analyzing the relative merits of various databases and bibliographies as sources of bibliographic references. Compared are all the 300+ online databases accessible through CONIT and three major Hudson River bibliographies: (1) "Annotated Bibliography of New York Bight, Hudson-Raritan Estuarine System and Contiguous Coastal Waterways: 1973-1981" [HORV84] prepared at the Marine Sciences Research Center of SUNY Stony Brook (SSBB); (2) a "Hudson River Bibliography" prepared by Peter S. Walczak et al as an appendix in a report prepared by Karin Limburg et al [LIMB85] at Cornell University (CB); and (3) Appendix (Section 7) of the Users Guide to the Hudson River Data Base, a report [KEIS85] by Richard Keiser and Ronald Klauda (KKB).

The three bibliographies are already Hudson-specific. We performed searches on documents mentioning "Hudson River" in 41 online databases (see Table 1 for list of databases). In 37 of these 41 databases Hudson River document references were found as shown in Table 2. For a few databases -- e.g., GEB, GEL-- the more general search on "Hudson" and "River" (words not necessarily occurring together as a phrase) was substituted. It was found the more general search occasionally produced false drops (e.g., on "Hudson Bay") and rarely found anything relevant which was not retrievable under the more specific search.

Besides the "Hudson River" searches we performed a search on "New York Harbor" or "New York Bight" in three selected databases [GWA, GWO, and GWW) to see how many *additional* references are found. As can be seen in Table 2, in those three databases as a group the number of references is more than tripled. The total number of references found by these searches is 4,691. Of course, many *duplicate* references are included; i.e., the same reference will be found in different databases. These results were printed offline and have been turned over to the Foundation.

A review of the results indicates that the relevance of the references to the hard-core environmental concerns of the Foundation is at, or close to, 100% for most of the databases. For the news and reports (UR) databases it appears to be in the range of 30 to 70%; for the books database (UBL), perhaps around 10%. Thus for UBL 6 of the 128 retrieved were indexed by a term beginning "environm:" and that term or either "fish" or "pollut:" were in 101 of the 186 in UNNN and in 27 of the 162 in the MAGAZINE database (UM).

Table 1. Selected Databases of Interest

AW AQUACULTURE [D] Growing of aquatic organisms (FISH,112)
GEB ENERGY BIBLIOGRAPHY [O] Energy Bibliography and Index (EBIB)
GED DOE ENERGY [D,O] Tech Info C, US Dept. of Energy 1983+ (DOEE,POWER,103)
GED1 DOE ENERGY 74-82 [D] File GED for 1974-1982 (104)
GEE ELECTRICITY [D] Electronic Power Research Institute (EPRI,241)
GEG ENERGYLINE [D,O] Enviroment Info. Ctr., Energy Info. Abs. (ENERGY,69)
GEL ENERGYLIT [O] Department of Energy (EDB)
GEN ENERGYNET [D] Environment Information Center's Energy Directory (ENET,169)
GEP ELECTRIC POWER [O] Electric Power Industry Abstracts (EPIA)
GF FISH AND WILDLIFE [D] U.S. F&W Reference Service (FWRS,957)
GGA GEOARCHIVE [D] Geosciences, Geosystems, England (GEOA,58)
GGR GEOREF [D,O] Geological Reference File, Am. Geological Inst. (GEOR,89)
GGM METEOROLOGY [D] Meteorology Soc. and Geoastrophysics (METEOR,29)
GPG POLLUTION ABSTRACTS [D] Data Courier (POLLUTION,41)
GPA AIR POLLUTION [D] Air Pol. Tech. Info. Ctr. of EPA (APTIC,45)
GVE ENVIROLINE [D,O] Environment Information Center (ENVIR,40)
GVB ENVIROBIB [D] Environmental Periodical Bibliography (EPB,68)
GWA AQUATIC [D] Aquatic Sciences & Fisheries Abs., life sciences (ASFA,44)
GWN WATERNET [D] American Water Works Assoc. (AWWA,245)
GWO OCEANIC ABSTRACTS [D] Oceanography, Marine Affairs (OCEAN,OA,28)
GWR WATER RESEARCH [D] Water Research Centre, England (AQUALINE,WR,116)
GWW WATER RESOURCES [D] Slected Water Resources Abstracts (SWRA,117)
MBG BIOLOGY [D] Biosciences Info. Serv. Bio. Abs. (BIOSIS,5)
MBE BIOETHICS [N] Ethics considerations in Medicine and Research (MEDETHICS)
MBL LIFE SCIENCES [D] Animal behavoir, biochemistry, ecology, etc. (IRL,LS,76)
MMM MEDLINE [N,D] Natl. Lib. of Medicine Index Medicus 1980+ (MEDICUS,154)
MME EXCERPTA MEDICA [D] Biomedical; Dutch Publisher 1982+ (EMBASE,72)
MTG TOXICOLOGY [N] 1981+ (TOXLINE)
MTG2 TOXICOLOGY 76-80 [N] File MTG for 1976-1980 (TOX76)
MTG3 TOXICOLOGY 65-75 [N] File MTG for pre1965-1975 (TOX65)
UBL BOOKS [D,O] Books cataloged by Lib. of Congress (LC/LINE,LCMARC,426)
NEDS [B] NEDRES (NOAA) directory of environmental databases
NED1 [B] Older entries for NEDS file
UBP BOOKS IN PRINT [D] R.R. Bowker (BIP,470)
UM MAGAZINE INDEX [D] Covers Articles in 370 popular magazines (MAG,47)
UNA1 ASSOCIATED PRESS NEWS CURRENT [D] Last three months (APNEWS,258)
UNCS CHRISTIAN SCIENCE MONITOR [O] Articles 1979+ (MONITOR)
UND DAILY NEWS INDEX [D] Current month for UNNN and UM (NEWSEARCH,DNI,211)
UNNI NEWSPAPER INDEX [O] US Papers: LA Times, Wash Post, etc. (NDEX)
UNNN NATIONAL NEWSPAPER INDEX [D] NY Times,Wall St Jrnl, Sci Mon (NNI,111)
UNU UPI NEWS [D] March 83 to 2 months ago (260)
UNU1 UPI NEWS CURRENT [D] Last three months (261)
UNWA WORLD AFFAIRS REPORT [D] Calif. Institute of Int'l Affairs (WAR,167)
UR REPORTS [D,O] Government-Sponsored R&D, Natl Tech Info Service (NTIS,6)

KEY -- listed for each file is: (1) CONIT code name; (2) CONIT full name;
(3) code letters of systems having the file [D=DIALOG, N=NLM, O=ORBIT B=BRS];
(4) short explanation of file;
(5) abbreviated and other alternate names for the file (in parentheses)

Table 2. Search Results and Overlap Figures

FILE	SEARCH	# DOCS	SSBB/OD	CB/OD
AW	A	0		
GEB	B	0		
GED	A	45		
GED1	A	247	10/20	
GEE	A	76		
GEL	B	3		
GEN	A	0		
GEP	A	395		
GF	A	110		1/5
GF	C	0		
GGA	A	28	12/20	
GGR	A	293	9/20	1/5
GGM	A	18	8/11	
GPG	A	90		
GPA	A	8	0/4	
GVE	A	158	9/20	
GVB	A	54	10/20	
GWA	A	121	16/20	3/5
GWA	C	316		
GWN	A	5		
GWO	A	83	17/20	3/5
GWO	C	291		
GWR	A	68	17/20	1/5
GWW	A	278	19/20	2/5
GWW	C	419	18/20	
GWW	E	215		
MBG	A	87		
MBE	B	0		
MBL	A	58	17/20	3/5
MMM	A	7		
MME	A	16		
MTG	A	115		
MTG2	A	61	10/20	
MTG3	A	32	5/9	
NEDS	B	11		
NED1	B	29		
UBL	A	128		
UBL	D	6	2/4	
UBP	A	57		
UM	A	162		
UM	D	27		
UNA1	A	111		
UNCS	B	3		
UNNI	B	25		
UNNN	A	186		
UNNN	D	101		
UNU	A	82		
UNU1	A	61		
UR	B	356	17/20	

Search Key: A=hudson & river; B="hudson river"; C=(new york) & (bight or harbor) not (A or B); D=A & (environ: or fish: or pollut:); E= basin: or canyon: or gorg: or estu:

The SSBB bibliography is, of the three printed bibliographies the largest with 2566 annotated references. The publication data coverage of SSBB is from 1973 to 1981. The CB bibliography has about 1250 references, of which approximately 350 are utility-company sponsored reports; CB references range from the 1950's to 1983. The KKB bibliography has approximately 500 references, mainly for utility company sponsored reports containing or describing data collections.

We analyzed overlap statistics in order to compare the coverage of online databases and the printed bibliographies. The column in Table 2 headed "SSBB/OD" shows how many of a sample of document references from the various online databases were in SSBB. The second (denominator) number is the size of the sample: generally 20 or as many of those references from the online database in the 1973-81 time range. The first (numerator) number is the number of the sample also in SSBB. In *no* case did SSBB have *all* the references found in the online database. In general, the percentages ranged from 45% to 90% although in one case (GPA), *none* of a small sample of 4 were in SSBB.

From the reverse perspective we sought to discover what fraction of the SSBB can be found in the online databases. For a random sample of 10 from SSBB 6 (60%) were found in the online databases. Note that 4 of these 6 did NOT specifically mention the Hudson River; 2 of these 4 mentioned the Hudson Bight or New York Harbor and 2 mentioned (or implied) only other regional coastal locations.

An analysis of the coverage achieved by the CB bibliography of the online databases was performed similar to, but smaller in scope than, the one analyzing the coverage of SSBB. The results are shown in the column headed "CB/OD". For 7 major online databases the coverage ranged from 20% to 60%.

Two online databases are worthy of particular note: NEDS and NED1. These two are the online databases for the National Environmental Data Referral Service (NEDRES) of NOAA. NEDRES provides, partly through these databases, a directory service of existing environmental data including lengthy descriptions of the data itself and contact points for further information. It thus goes considerably beyond most bibliographic data in its scope and depth on projects. The NED1 database is a closed file containing 12,000 data file descriptions generated between 1974 and 1980 as part of the Environmental Data Index (ENDEX) developed by the National Oceanographic Data Center. NEDS embraces a new format and is designed to eventually incorporate NED1 data and replace it. As of the beginning of 1985, NEDS contained only about 2,000 entries.

The two NEDRES files are mounted only on the BRS/SEARCH system which was used directly (it is *not* currently one of the systems accessible through CONIT). The number of "Hudson River" entries found was 40; 29 on NED1 and 11 on NEDS. In view of the extensive nature of the information on data files contained in the NEDRES databases, they could serve as a basis for a Hudson River online data directory if the current meagre number of listings was augmented. The SDIMS study by CCA [GOLD85] has been investigating that prospect.

In comparing the printed bibliographies we note that KKB appears to contain the most comprehensive coverage of the utility-company sponsored reports; in that regard it appears to be an augmented version of Section 1 of CB. Of course, KKB does not pretend to cover all types of documents as its smaller size attests.

Part of the larger size of SSBB compared to the others is due to its larger geographical scope: coastal areas outside the immediate Hudson River areas. Even within the Hudson River

itself, however, it is clearly quite comprehensive as is seen by the high percentage of its coverage of the references found in the core water-related online databases. The two main deficiencies in coverage of SSBB are: (1) small coverage of the utility-company sponsored reports and (2) a date of publication limited primarily to the 1973-81 range. On the other hand, SSBB includes extensive abstract/annotations, a sort by author and a keyword indexing scheme. KKB also go beyond a simple listing of references to classify and index the studies in various categories.

To emphasize the complementarity among the bibliographies we may state the following fact: out of a sample of 20 entries in SSBB (entries numbered 500-519) none were found in CB. Clearly, also, there is a complementarity among the online databases and between the databases and the bibliographies. We estimate each of the Hudson-specific references appears on the average in from two to three of the databases. This would make the number of *unique* references we found from the databases in the range of 1700 to 2500; i.e., still somewhat less than the number in SSBB. In fact, if we extended the search strategy beyond the few simple geographic location names, we estimate perhaps twice as many unique Hudson-specific references would be found. Nevertheless, in view of the wider time coverage and topical scope (e.g., news articles and non-strictly-environmental topics) we would expect an even larger number if the online databases were fully comprehensive. In fact, our analysis showed that an estimated 40% of the references in SSBB were not in the online databases. (Note that we found a similar degree of omission of coverage in the SSBB compared to the online databases.)

Two reasons for the lack of complete coverage in the online databases were noted. In the first place, some of the online databases go back in coverage only 5 to 10 years. Also, the early years of the online databases generally have less complete coverage than the recent years. We estimate that possibly one half of the 40% omission figure is due to this earlier incompleteness in the online databases.

In the second place, there still are a number of documents that do not get included in the online databases. We noted an occasional journal article or special type (e.g., Master's Thesis -- the THESES database contains only PhD theses); however, the primary document type still not completely covered in the online databases is the report. Thus, for example, none of a sample of 20 utility-company sponsored reports from CB was found in the REPORTS (UR) database. We discuss in Section 9 below some means to redress this lacking.

8. CONIT for Interfacing Other Functions

As was pointed out in Section 5, there are a number of functions besides retrieval of bibliographic references from already existing databases which are desirable in the information systems for the Foundation. One such function is the ability to *create* one's own bibliographic database. A second desired ability is to create and search easily databases of all kinds, including *numerical* databases.

We sought to investigate one means to achieve these goals, through interfacing CONIT with an existing general-purpose database management system (DBMS). We chose to experiment with interfacing CONIT to RDMS (Relational Database Management System), a relational DBMS that resides on the M.I.T. Honeywell Multics computer.

Software techniques were devised for calling RDMS routines from CONIT and for receiving and passing RDMS responses in CONIT. Thus RDMS could be treated similarly to the other remote retrieval systems to which CONIT interfaces. Of course, RDMS is not "remote" (i.e., on a computer remote from MULTICS) and so there were no time delays or costs associated with networked intercommunications between CONIT and RDMS.

A small sample database on RDMS was constructed for containing some of the information in one table of the data description guide in the Cornell Environmental study. This table was Table 7.16 entitled "Hudson River Open Literature Highlights: Algae and Macrophytes". This table lists and describes 13 studies in which data were collected on the given topic. (See copy of this table in Appendix B).

Four columns from the table were chosen to be represented in the RDMS file: (1) *Reference,* (2) *Description, (3) Years of Study.* and (4) *Sample Location. Reference* contains the year of publication and author(s) of the document (listed in the CB printed bibliography) that describes the study. *Description* is a short (up to 9 words) abstract of the nature of the study. *Years of Study* contains the first and last years of the study. *Sample Location* gives a diagrammatic picture of where the samples for the data were taken as measured by by river mile. These 4 columns contain the bulk of the table information and represent a cross section of the textual, numerical, and author-name data important to data directory information.

Twenty three RDMS fields were created to represent the selected data. Three fields capture the "Reference" information: a year of publication (numeric) field and two textual author fields (the second author field is null if there is only one author -- the Reference column does not list more than two authors). Two numeric fields capture the beginning and ending year of the study. The Sample Location data was also encoded in two numeric fields in the following way: one field contained the *lowest* river mile sampled and the second field the *highest* mile sample (this encoding does not capture all Sample Location data; several *discontinuous* locations would be merged and a location outside the river-mile spectrum -- e.g., New York Bight -- would be missed).

The Description was stored as a single textual field. In addition, 15 index-term fields were used to capture words and phrases from the description to be used for search purposes. (Relational database management systems generally are limited in their ability to handle the repetitive textual fields typically needed for bibliographic and textual retrieval and must resort to subterfuges like this.)

CONIT commands were modified to handle basic retrieval operations from this "Hudson" database much as they do for databases from the remote retrieval systems. One command selected this Hudson database and created the necessary software environment. There were two search commands, one did searches by author and the other by index term. One difference from other CONIT search commands was that instead of CONIT search automatically naming the result of a search, the user had the ability (and requirement) to name search results. One simplified print command was also provided to display the fields and relations resulting from a retrieval. In summary, while the retrieval functions in CONIT for interfacing to the relational database are still minimal, we have demonstrated that a basic set of bibliographic retrieval functions can be incorporated in CONIT for accessing a local database.

Beyond these bibliographic-type retrieval functions, the full-range of retrieval operations possible with the RDMS system is also achievable by calling on the commands of RDMS directly via a "pass-through" technique. Thus, for example, one could search on studies mentioning a particular topic done in a certain range of years that took samples beyond a certain river mile. Also, results derived from the main database are stored as "relations", just as are all relational databases, and therefore serve as "mini" databases of their own and can be searched directly (without having to re-search the *main database* and *intersect* the results with other search results, as with standard bibliographic systems).

To make the CONIT interface to RDMS a truly useful and user-friendly operation -- comparable to what CONIT now accomplishes for its three interconnected bibliographic retrieval systems -- would require extending command facilities and associated translation and explanation procedures. This would be fairly straightforward but requires considerably more effort and resources than were available to us on this grant.

9. Conclusions and Recommendations

9.1 Overview

The experimentation and analyses of these investigations supports our original hypothesis that the newly emerging techniques of computer intermediary systems can provide one useful basis for developing the Public Access Information System (PAIS) for the Hudson River Foundation, as well as similar systems for many other organizations. The principal advantage of these techniques derives from the economics achievable through *sharing* -- computer systems and databases can be shared. In fact, the *simultaneous,* online sharing of programs and databases actually *enhances* the economy without degrading performance (up to some, usually high, saturation point in utilization of the hardware and software).

While these intermediary system techniques should *ultimately* be useful in the kind of information systems of which PAIS is representative, there are three reasons that prevent these techniques from providing an *immediate* basis for *fully* handling the desired PAIS functions. First, current intermediary systems do not fully satisfy PAIS functions. Secondly, the Hudson River Foundation by itself does not have the resources to support the development of modified and improved systems of this type that *would* be fully satisfactory. Thirdly, the Hudson River Foundation community does not yet have the computational facilities or the motivational factors that would make access to, or felt need for, such systems at a high level for the general population.

To illustrate the magnitude of the problem let us consider the intermediate solution of maintaining a *private file* of bibliographic data on one of the existing retrieval systems. This solution has the potential merit of making Hudson-River-related bibliographic references available in a single, centralized database accessible to all who have, or can use, accounts on a major retrieval service. The Foundation would not have to make and maintain any retrieval system of its own but only provide input data in a suitable digital form to the collection of publicly-available databases. (Actually, a *private* file is only *publicly* available to the extent mutually agreed upon; use of such private files may be limited to only those users of the retrieval system who have been given special access codes.) The Fish and Wildlife Reference Service (FWRS) maintains two such files on Dialog and NEDRES maintains two files (previously described), although more of a data directory nature than a strictly bibliographic nature, on the BRS system. An analysis of these operations led us to conclude that it would cost the Foundation more than $100,000 per year to maintain such a private file. In light of expected only low usage the revenues achievable to the Foundation if it chose to try to recompense itself for such service would not defray costs significantly.

In view of these observations it is our recommendation that the Foundation expend no further efforts at this time to develop a computerized PAIS of its own along the lines originally envisioned (i.e., major, online multi-access computer databases). Rather, the Foundation should seek to utilize the existing computer systems (e.g., the various retrieval systems, CONIT and other intermediary search assistants) and to develop the expertise within the Hudson River community to know how to use the computer systems and to utilize more standard information management techniques (e.g., printed or local computer [not networked] informational listings and files).

9.2 General PAIS Plans

There would be several modes of operation and functional areas related to PAIS:

1- Development of (additional) information files and procedures
2- Periodic Dissemination of Information
3- Handling of Spontaneous Queries
4- Recording and Analysis of PAIS operations

9.2.1 Development of Information and Procedures

This includes a collection of comprehensive data on Hudson River Foundation and Hudson River Foundation-sponsored activities as well as selected information on non-Hudson River Foundation data, including information about secondary sources (e.g., people, libraries, computer databases) from which more information is available.

Hudson River Foundation data would include an updated listing of all current and past supported activities (the kind of information currently reported in the Hudson River Foundation Annual Program Plan and Solicitation of Proposals) and reports related thereto, including Hudson River Foundation reports (e.g., the Annual Plan) themselves.

Non-Hudson River Foundation data could include a list of individuals and organizations, as well as computer systems and databases, with information on Hudson River concerns. This listing should contain summary descriptions of those concerns, including type of information for which each can serve as secondary sources, contact points (and/or access procedures for secondary sources), and type of funding available (if organizations that sponsor work). Also included would be current events and other Hudson River related information as well as lists of projects and bibliographies summarized by the type of information available in each.

Another type of information would be incorporated in the so-called Data Directory, envisioned in the Goldman/Bergman [GOLD85] SDIMS study and proposal, in which information about data files would be maintained in a Foundation microcomputer and/or in an outside database (e.g., NEDRES). Because the data files are associated with particular people, projects, organizations (and, often, reports), the data file records share much in common with the bibliographic records. Therefore, it is potentially very worthwhile to consider integrating the Data Directory type of information gathering and recording with attempts to capture at least some of the references to the "gray literature " on the Hudson that may be escaping inclusion in the online bibliographic databases.

Information and file development includes the collection/curation of printed lists, brochures, reports and monographs for transmission to, or review by, interested parties. The procedures to be developed include those to direct the other PAIS modes as described below.

9.2.2 Periodic Dissemination of Information

Collected data of the type mentioned above need to be regularly and periodically disseminated to individuals and organizations. One type of dissemination would be to persons and organizations for their direct and immediate use. The preparation and dissemination of

the Annual Plan to a selected distribution list would be an example of this type of dissemination. The development of a Hudson River newsletter for more frequent and comprehensive dissemination has been proposed and should be given favorable consideration in some form.

A second type of information dissemination would be to maintainers of databases like NTIS, NEDRES, ASFA, ENVIROLINE, etc. This would provide more convenient access to Hudson River related information to the world at large in addition to the Hudson River community itself.

9.2.3 Handling of Spontaneous Queries

Information requests come into the Foundation from the "outside world" (e.g., press, public, researchers and political/administrative officials) and from/through its immediate "family" (Board, Panel, and grantees). Request come in person, by telephone, and by mail (standard and, eventually, electronic). Using the data and resources developed and categorized as mentioned above, the recipient of the request should be able either to answer the request directly or refer it to an appropriate source. Included in this scheme is enhanced access to external systems and databases through CONIT.

9.2.4 Recording and Analysis of PAIS Operations

All requests for information and responses thereto, as well as complimentary and critical remarks, should be recorded for analysis of the utility and reputation of the Hudson River Foundation in general and its PAIS in particular.

9.3 Specific PAIS Plans

We have undertaken specific steps to help organize the effort to develop a Hudson River Foundation PAIS along the mixed-mode lines described above. Of course, our investigations have produced results which, in turn, have led to the general plans. Accounts have been set up on the DIALOG and the BRS Retrieval Systems specifically in the Foundation name. Not only are all of the (over 200) publicly available files thus accessible, but also we have arranged Foundation access to the NEDRES files and one private file of the Fish and Wildlife Reference Service (FWRS): file 947 on Dialog. It is not possible to gain special-permission access to the other FWRS private file which lists projects funded under the Dingell-Johnson and Pittman-Robertson acts. However, we have obtained a printed compilation [SOUS85a and SOUS85b] of the information in this database for the Foundation (it lists, for example, about 1234 projects of which 25 are from New York State).

As we have previously mentioned, access to CONIT was provided on an account set up within CONIT, under the MIT Multics computer and through that to M.I.T. accounts on the SDC ORBIT and National Library of Medicine Systems, as well as the Foundation account on Dialog. Should Foundation use of these various M.I.T. accounts grow to a level beyond which the M.I.T. project on which CONIT resides can support, it will be necessary to set up Foundation-specific accounts on those systems.

A quantity of information about these various systems and databases, and results of searching them, has been gathered for future Foundation use.

Specific steps have been taken to prepare procedures for submitting reports to NTIS for inclusion in its online database and printed listings. A manual [NTIS83] was obtained describing the general procedures for submitting to NITS. One step in this process has been completed: we have registered with NTIS a code to serve as a basis for a Hudson River Foundation scheme to assign report document numbers. The Foundation has been assigned the code consisting of the 3-letter prefix "HRF" as an internationally unique identifier for its report numbers.

Here is the scheme we propose for using this code. Let us consider three series of documents:

1. Reports by grantees describing their research sponsored by the Foundation.

2. Other reports/documents prepared by the Foundation itself (e.g., annual reports, state-of-river reports, etc.)

3. Other documents the Foundation would like to make available through NTIS (e.g., utility-sponsored reports not already so available).

Let us give specific "infix" designations to these 3 series as follows:

GR for grantee report
FR for Foundation-prepared report
OD for other document

The report number for the second grantee report of 1986 (for example) would then be: HRF/GR-86/02

With this coding scheme a searcher can easily isolate from the overall NTIS database any number of "virtual" Hudson River Foundation "mini" databases. Thus, searches on various positions in the Report Number field can retrieve all Hudson River Foundation documents, or subsets such as all grantee reports, or all Foundation-prepared documents in a given year, or all Hudson River Foundation documents of *any* type for a given year, as well as, of course, any given Hudson River Foundation document. Once a subset has been retrieved in this way the documents in the subset can be listed or further delimited by other searches -- e.g., finding all Hudson River Foundation documents discussing PCBs or authored by Jon Cooper.

Another step in setting up the submittal procedure is to devise plans for filling out a "Report Documentation Page" for each submission (see page 2 of this report). We propose this be done primarily by the author (e.g., grantee) with review and actual submittal done by the information specialist (to be hired) at the Foundation. The intellectual aspect of this step is the assignment of subject descriptors and identifiers. We have obtained reports that should help the information specialist oversee this aspect of the submittal.

One avenue toward the appropriate development of Hudson River Foundation information systems would be for the Foundation to hire a professional information specialist to serve in its headquarters. A full-time staff person would be desirable, but some useful planning and operations could be achieved by a half-time person. Such a person could be a (special) librarian with experience in reference services. Knowledge of and experience with (or, at least, interest in) bibliographic database retrieval systems (Dialog, etc.), microcomputer

operations, and, of course, environmental subject matter would be desirable. The development of PAIS and/or SDIMS files could be aided by the use of standard microcomputer software (e.g., dBASE-3) on the Foundation PC AT microcomputer.

To the extent that the acquisition of additional Foundation staff should prove infeasible, one would need to consider either restricted operations with existing staff and/or more emphasis on outside support -- e.g., for newsletter development and an information specialist intermediary searcher (perhaps located at one of the associated organizations) to provide assistance for those needing help in searching online bibliographic databases either through CONIT or through the retrieval systems directly.

Acknowledgments

We wish to acknowledge the cooperation extended to us in our investigations by the Hudson River Foundation Science Officer, Jon C. Cooper, as well as by other members of its Information Management Committee and all those who participated in the experiments.

The author would like to acknowledge those at M.I.T. who worked on the CONIT project including the following students: Ricardo A. Cardenas, Louis Hing-Fai Chong, Monica R. Gerber, Aleks Gollu, William Tiger Lee, Steven S. Schwartz, and Colina Man-Wah Yip. In particular, Mr. Gollu was responsible for the CONIT interface to RDMS and Ms. Yip assisted in the comparative analysis of databases and bibliographies.

References

[GOLD85] Goldman, Morton F.; and Bergman, Rita F. Approach for the Development of a Scientific Database for the Hudson River Foundation. Computer Corporation of America Report. March 1, 1985.

[HORV84] Horvath, R.F.; Carroll, G.E.; Covell, S.A.; Evjen, A.J.; and Schoof, J. Annotated Bibliography of the New York Bight, Hudson-Raritan Estuarine System and Contiguous Waters: 1973-1981 (2 volumes). Marine Sciences and Research Center, State University of New York, Stony Brook. June, 1984.

[HRF85] Hudson River Foundation Annual Program Plan and Solicitation of Proposals for 1985.

[KEIS85] Keiser, Richard K., Jr. and Klauda, Ronald J. Users Guide to the Hudson River Data Base, Volume I - Overview. Environmental Information Systems, Cockeysville, MD. November, 1985.

[LEFK82] Lefkovitz, David "Primate: A Microcomputer System for Host Mediation and Private File Text Search." *Proceedings of the 45th ASIS Annual Meeting* 19: 162-164. October, 1982.

[LIMB85] Limburg, Karin E.; Moran, Mary Ann; and McDowell, William H. Environmental Impact Assessment of the Hudson River Ecosystem: Multiple Case Study and Data Base Review. Ecosystems Research Center, Cornell University. February, 1985.

[LISA84] Lisanti, Suzana. "The On-line Search." *BYTE*. 9(13):215-230; December, 1984.

[MARC83a] Marcus, Richard S. "Computer-Assisted Search Planning and Evaluation." *Proceedings of the 46th Annual Meeting of the American Society for Information Science.* 20:19-21; October, 1983.

[MARC83b] Marcus, Richard S. "An Experimental Comparison of the Effectiveness of Computers and Humans as Search Intermediaries." *Journal of the American Society for Information Science* 34(6):381-404. November, 1983.

[MARC85] Marcus, R.S. "Development and Testing of Expert Systems for Retrieval Assistance". *Proceedings of the 48th ASIS Annual Meeting.* 22:289-292; October, 1985.

[MCGI83] McGill, Michael J.; Brown, Gerhard E.; and Siegel, Sidney. "The Chemical Substances Information Network: Its Evaluation and Evolution" *Proceedings of the 1983 National Online Meeting* Medford, NJ: Learned Information, Inc. pp 367-376; April, 1983.

[NTIS67] TEST: Thesaurus of Engineering and Scientific Terms. Office of Naval Research, Washington, DC and Engineers Joint Council, New York, NY. 1967. NTIS order no.: AD-672 000.

[NTIS77] Environmental Microthesaurus - A Hierarchical List of Indexing Terms Used by NTIS. National Technical Information Service. May, 1977. NTIS order no.: PB-265261.

[NTIS78] NTIS Subject Classification (Past and Present). National Technical Information Service. January, 1978. NTIS order no.: PB-270 575.

[NTIS83] How to Submit Your Information Items to NTIS: A Manual for Information Sources. U.S. Department of Commerce, National Technical Information Service. 1983-1984.

[SOUS85a] Sousa, Robert J; and Stephens, Claude E. (ed.) Current Federal Aid Research Report: Fish. Spring 1985. U.S. Fish and Wildlife Service, Division of Federal Aid.

[SOUS85b] Sousa, Robert J; and Stephens, Claude E. (ed.) Current Federal Aid Research Report: Wildlife. Spring 1985. U.S. Fish and Wildlife Service, Division of Federal Aid.

Appendix A

Instructions and Questionnaire for CONIT Experiment

HUDSON RIVER FOUNDATION
FOR SCIENCE AND ENVIRONMENTAL RESEARCH

HRF

122 East 42nd Street • Suite 1901 • New York, NY 10168 212 949-0028

July 17, 1984

One of the functions of the Foundation is to manage information about the Hudson River's environmental characteristics, resources, uses, and dynamics. The Foundation wishes to ensure ready access to all relevant scientific, economic, and public policy information, whether or not such information is generated by projects under its own sponsorship.

To assist in the fulfillment of the information management function we have given grants to the Massachusetts Institute of Technology to analyze how a new technological development, the intermediary information system, could be used to satisfy Foundation goals. Your response to the enclosed questionnaire and instructions for test usage of the experimental M.I.T. CONIT intermediary system will help us evaluate our needs and potential methods to achieve the goals. It would be most helpful if your response is sent in two weeks time.

For your further information concerning the Foundation's efforts in the information management area the rest of this letter provides a brief historical background.

During 1983, the Foundation issued two requests for proposals (RFPs) in the area of information management entitled "Scientific Database for the Hudson River" and "Public Access System for the Hudson River". Sixteen applicants to a Request for Qualifications were asked to respond to these RFPs.

The Foundation is now planning and implementing database systems which will serve the long-term activities and needs of the Foundation. In the Public Access System, the Foundation will be seeking to develop a system which has a summary, indexing, cataloging or inventory function and which provides an overview of information in the most general sense (who, what, where, when). The Public Access System will enable users to identify and access information in various indexes, summaries, catalogs, and inventories; will have remote-access capabilities; will be highly "user friendly"; and will be one that provides for easy access to information on the Hudson River. Users of the Public Access System are expected to include permit applicants, researchers for environmental groups, students, and others who may not necessarily be trained in the use of computers or have a knowledge of computer languages.

In the Scientific Database System, the Foundation seeks to develop the plans for organizing and storing scientific databases relating to the Hudson River and environs. The Scientific Database System is expected to be used by people with training and experience in

quantitative methods and the use of computers. However, since different disciplines may be accustomed to different computer languages, this system will also be "user-friendly" to the extent possible.

Two contracts were signed in early 1984 for the initial assessment of use and development of a computer structure (infrastructure) which would lead to the establishment of the Public Access System. We felt that the most cost effective approach to providing the capabilities of a Public Access System for the Foundation involves the utilization of existing information systems and databases. Not only will the shared use of such systems provide efficient access by the Hudson River Foundation community to a large portion of the information it may need from the world at large, but also the information generated by this community will be made readily available to the outside world as well as to the community itself. Recent research and development in intermediary computer assistant systems for information retrieval has demonstrated effective mechanisms for allowing users easy access to and operation of existing multiple retrieval systems. The first grant provides consulting support to investigate how these new intermediary systems could be adapted to provide this kind of effective networking and sharing of information resources within and outside the Hudson River Foundation community. Both analysis and demonstrations of existing systems will be part of the effort in conjunction with a grant to Massachusetts Institute of Technology.

The second project will devise a prototype Public Access System based on the CONIT intermediary retrieval system (which has been recently developed on research projects at MIT to demonstrate new techniques for providing easy and effective access to the world's literature). The prototype system will be an expanded version of CONIT, incorporating files specifically formulated to meet Hudson River information needs not provided by the publicly available literature databases. The prototype system will be evaluated through analysis of its use to determine how well it serves the information needs of the Hudson River community. The analysis will also consider the wider questions of how the prototype Public Access System could be enhanced in ways that would greatly improve on the current state of information retrieval technology and on the vital societal question of whether such a system can serve as a model for information transfer on a national scale.

I hope the above information provides sufficient background to explain the importance of the experiments and accompanying questionnaire.

Sincerely,

Jon C. Cooper
Science Officer

Enclosures:
Questionnaire (4 pages, 2 copies)
with addressed, stamped return envelopes
(Simple) Instructions for Accessing and Using CONIT (2 pages)
CONIT Manual (54 pages)
List of Telephone Numbers for Telenet and Tymnet Networks (8 pages)
Instruction Booklets for Using Telenet and Tymnet (16 and 12 pages)

Questionnaire on Information Systems for Hudson River Foundation

A. Personal Data

A1. Your name _____

A2. Address (if not one on cover letter)

_____

A3. Telephone _____

A4. Relationship to Hudson River Foundation
Hudson River Panel _____
Current or past grantee _____
Other (explain) _____

B. Access to Computer Networks

One important requirement for an effective group use of a computer-based information system is that the individuals of the group have convenient access to the computers. We are seeking here to learn how convenient *your* access might be. In particular, what is your access to a computer or computer terminal with the necessary telecommunications capability to make telephone connections to remote computers. For each mode of access you might use we would like to know the following information:

B1. Maker and model of computer or computer terminal (e.g., Texas Instruments 745 printing terminal; IBM PC XT-370; DEC VAX 11-730)

_____

B2. Which items in (B1) have telecommunications capability (e.g., modem on terminal, telecommunications hardware and software on computer)?

_____

B3. What speed in characters per second (CPS) on items listed above? (Note, CPS is generally BAUD rate divided by 10 and is typically between 30 and 120 for printing terminals and in same range or higher for CRT terminals and computer telecommunication facilities.)

_____

B4. If access is through CRT, is there an associated printer for getting hard-copy output?

(Yes/no) _____; Type/speed (if known) _____

B5. CONIT and many other information systems are accessible via computer networks that have many local nodes. Please look at the accompanying list of local nodes of the Telenet

and Tymnet networks. Assuming you had the appropriate computer capabilities, which telephone number(s) would be most convenient for you to use?

_____

Would this be a local or long-distance call for you? Local ____; Long distance ____

B6. Even if you have *some* access to equipment for computer telecommunications, it may not necessarily be *convenient* for you to use. Please give an indication of how convenient such access is for you; e.g.,

a. I have a terminal on my desk and/or can easily make access any time of day ____
b. There is a terminal nearby and would be accessible on 10-minutes notice ____
c. I would have to go (or send someone) to a remote (greater than 10 minutes away) location and/or make special provisions to use a terminal or computer ____
d. Currently, no access at all ____
e. Other (explain) _____

C. Trial Use of CONIT

Proof positive of your computer telecommunications capability would be to make an actual connection to the M.I.T. CONIT system. It would be helpful to us if you -- or someone you designate in your organization -- would try to follow the enclosed instructions to make such a connection and use the CONIT system. We have set up an account on CONIT for this purpose. For these preliminary investigations you will be limited to $100 of computer costs (unless you can make arrangements to reimburse the Foundation for additional costs). (CONIT keeps track of costs for you.) Beyond these preliminary tests the Foundation is studying various mechanisms for allocating "free" resources to users and/or establishing mechanisms for them to reimburse costs beyond a given amount.

It would be most helpful if you could try to use CONIT within the next two weeks and answer the rest of the questionnaire. If you cannot do so, please mark the appropriate item in question C1, skip over to section D, complete the questionnaire and return it. If you make subsequent attempts to access CONIT after sending in the questionnaire, please fill out a second copy of the questionnaire (copy enclosed) for sections A and C -- and any changes in the other sections -- and return it.

C1. Didn't try ...

a. No access equipment _____
b. No time to try _____
c. No interest _____
d. Other _____

C2. I couldn't get connected. Difficulty appeared at the following step (enclose copy of your terminal typescript, if you have one available):

_____

C3. I did get connected (date: _____)
User name (if other than yourself) _____
Any difficulties or other comments (positive or negative) you care to make about your use of the system that you did not already give to CONIT while connected.

_____

_____

_____

D. Potential Use of Public Access System

We would like to know how you might use, or like to use, the Public Access system. (Please answer for your organization as a whole.) It might be helpful for you to look over the enclosed CONIT reference manual before answering the following questions.

D1. How many potential users of a Public Access system are there at your organization? _____

D2. Which computerized databases would you be interested in searching (you may use CONIT code names -- e.g., GWQ [for the WATER QUALITY database -- see page 17]).

_____

D3. How many literature searches would you be interested in doing (e.g., one a year/month/week/day) _____

D4. Can you describe one or two particular searches you might like to perform

_____

_____

_____

D5. What kind of information might you like to have available in the Public Access system that may not be (easily) obtainable from the current database access through CONIT? This could include other "general purpose" type information, like that available in the public files of the systems CONIT connects to, or information highly specific to the Hudson River Foundation (e.g., lists of Hudson River Foundation grants, grantees, of organizations concerned with the Hudson River, etc.). It might include numerical as well as textual information. (Although large-scale numerical databases *per se* are likely to be included in the Scientific Database system rather than the Public Access system, the Public Access system would contain at least some information about what databases of the type were available and how to access them and, possibly, moderate-sized databases with mixed textual and numeric data in tabular form.) Also, please describe any databases, files, or other information in your possession (currently or planned) that might be useful if added to Public Access or Scientific Database systems, or to external systems.

D6. What kind of functions other than retrieval would you be interested in seeing in a Public Access system (e.g., message sending, conferencing, text preparation and printing, computation)

E. Other Comments

If you have other comments about this questionnaire or about information systems for the Hudson River Foundation, feel free to express them here or in a separate communication.

Instructions for Accessing and Using CONIT for Hudson River Foundation Experiments

General instructions for accessing CONIT appear on page 7 of the CONIT manual. Additional particulars are given here.

Computer/terminal communication mode settings that usually work best are full duplex (as opposed to half duplex), even or no parity (as opposed to odd parity), and 120 CPS (1200 BAUD) or the slower 30 CPS (300 BAUD).

General instructions are enclosed for using the Telenet and Tymnet networks for connecting to the M.I.T. Multics computer on which CONIT resides. A separate listing includes the telephone numbers that you may use to call each network. Note, that the legend at the beginning of the list explains the tag letters which identify which phone numbers can be used with which modems and BAUD rates (as with other mode settings, you may have to experiment to determine which works [best] for you).

In brief, for Telenet, when the telephone connection is established (as indicated, for example, by the terminal/modem "online" light going on), type the RETURN key twice. After Telenet responds with four lines ending with "TERMINAL=" type (for most connections) the RETURN key once more. When Telenet prompts you with a line with just "@" type:

C 617138

to request connection to Multics.

When connected to Tymnet you will get the message "please type your terminal identifier". You then type "a" for most connections ("e" for slow terminals). When Tymnet then says "please log in:" you type:

mitmul

Then type just a RETURN after it asks for the (Tymnet) password.

The Multics ID we are currently using for these experiments is Conitd and the password is: [redacted] Thus when Multics prints its two-line herald (last line begins "Load= ...) type:

login Conitd

and when Multics asks for the password type:

[redacted]

(Don't be dismayed by all this login protocol nonsense; it is the most complicated part of talking to the computer and once you get the hang of it it will be second nature. It may be of interest to note that "smart" modems or computer telecommunications packages allow programming of protocols so that all the above can be largely automated. Once you are connected to CONIT all further connections to other systems *are* automated.)

When you get connected you will be in an account which is listed under the name Tibor Polgar. When CONIT asks if you are he, say 'NO' and give your own name. (For those who have need for extensive use of CONIT we shall arrange having accounts with your own individual names.)

It is not *necessary* to read the CONIT manual before you use the system; CONIT explains what you can do, and how to do it, once you are connected. However, many users do like at least to browse through the manual before getting connected to see what kind of commands are available (see page 9 and the Table of Contents) and to see what files may be

searched (see pages 12-23 — especially pages 17 and 13). Also, the manual may be helpful for reference once connected as an alternative — and possible quicker — way to finding and reading instructional explanations as compared to using the EXPLAIN command.

Please remember that CONIT is an *experimental* system and that your current use is part of a series of experiments to determine how operational systems should be designed and operated. Your indulgence with any problems and your suggestions for enhancements will be greatly appreciated.

If you have any questions or comments about this experiment, feel free to contact Richard Marcus directly at M.I.T. (617/253-2340) in addition to the questionnaire and CONIT comment channels.

Appendix B

Table Used for Data Retrieval Demonstration

The following table is extracted from the Cornell report [LIMB85]: Environmental Impact Assessment of the Hudson River Ecosystem.

Table 7.16. Hudson River Open Literature Highlights: Algae and Macrophytes. (p.261)

REFERENCE	DESCRIPTION	YEAR(S) OF STUDY	SAMPLING INTENSITY	SAMPLE LOCATION (0 25 50 75 100 125 / BT YK TZ CH IP WP CW PK HP KQ SQ CB AI)	UPPER HUDSON	CONCURRENT PHYS/CHEM	METHODS GIVEN	SPECIES	ABUNDANCE	DISTRIBUTIONAL: JOURNAL	SEASONAL	SPATIAL	GROWTH/PROD PROD	FEEDING	REPRODUCTION: MATURITY	FECUNDITY	BEHAVIOR	TIME	EXPLOITATION: SPORT	COMM	SURVIVAL: PRED/PREY	DISEASE	ENT. IMP	TOXICITY
Neale et al. 1981	effects of f.water flow on phytoplankton and salinity	1977-1978	irreg. weekly, Feb-Sept	Lower New York Bay		flow, salinity, depth	X	phytopl.					prod.											
Carrella and Duever 1969	phytoplankton communities at Indian Point and further north	1968-1969	monthly, Apr '68-Aug '69	XX XX XX		N, PO_4, temp, element.	X	phytopl.	X		X	X											X	X
Storm and Heffner 1976	phytoplankton abundance, chlorophyll a, water quality	1973-1974	seasonal	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	X	detailed	X	phytopl.	X		X	X	X											X
Heffner et al. 1976	phytoplankton sampling methodologies encouraged for H.R.	general background	no data					phytopl.																
O'Reilly et al. 1976	annual primary productivity	1973-1975	monthly	Lower New York Bay		detailed	X	nannopl. netpl.	X		X	X	X											
Frederick et al. 1976	phytoplankton distribution and species list	1972-1976	summary of 5 studies	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX				algae				X												
Williams and Scott 1962	diatoms of major U.S. waterways	1960-1961	biweekly	XX			X	diatoms	X			X												
Malone 1977	environmental controls of phytoplankton productivity	1973-1974	monthly	XXX and south			X	phytopl.					X											
Weiss 1978	distribution of diatoms in H.R. sediments	1973-1975	monthly	XXXXXXXXXXXXXXXXXXXXXXXXXXXX and N.Y. Bight		temp, D.O., salinity		diatoms				X												
Lund 1976-1981	phytoplankton species lists at Green Island	1975-1980		X				phytopl.	X															
Riviat 1976	distribution on threatened plant, Goldenclub						X	Orontium aquaticum	X			X												
Buckley and Mattice 1976	distribution of rooted vegetation, comparison to earlier macrophyte surveys	1968-1974		XXXXXXXXXXXXX		salinity	X	macrophytes				X												
McVaugh 1936	partial list of aquatic plants	1936		XX				aquatic and upland plants	X			X												