



NIST
PUBLICATIONS

NISTIR 5395

Preliminary Functional Specifications of a Prototype Electronic Research Notebook for NIST

**Shu-jen Chang
Elizabeth Fong
James Foti
Bruce Rosen**

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Computer Systems Laboratory
Gaithersburg, MD 20899

QC
100
.U56
N0.5395
1994

NIST

Preliminary Functional Specifications of a Prototype Electronic Research Notebook for NIST

Shu-jen Chang
Elizabeth Fong
James Foti
Bruce Rosen

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Computer Systems Laboratory
Gaithersburg, MD 20899

April 1994



U.S. DEPARTMENT OF COMMERCE
Ronald H. Brown, Secretary

TECHNOLOGY ADMINISTRATION
Mary L. Good, Under Secretary for Technology

NATIONAL INSTITUTE OF STANDARDS
AND TECHNOLOGY
Arati Prabhakar, Director

ABSTRACT

This report is a preliminary study on the feasibility and possible use of electronic research notebooks (ERNs) at the National Institute of Standards and Technology (NIST). The goal of this project is to determine the requirements for ERN and to assess current technologies for the design of a prototype ERN for NIST scientists. The project involves the determination and specifications of functional requirements of the ERN. NIST scientists were interviewed to survey current notekeeping practices and identify specific needs for the ERN. The survey identified a set of basic and enhanced ERN features. Current technologies and products were also assessed in order to see how these requirements may be met. A potential system configuration is proposed where functional specifications for a basic ERN are defined.

Keywords: Electronic research notebook, multimedia, information security, digital signature, digital time-stamp, information storage and retrieval.

ACKNOWLEDGEMENTS

We like to thank those NIST scientists named in Appendix C who graciously accepted our interview requests. Their insights and comments are crucial for our requirements analysis, upon which the design of the ERN is based. We are also very grateful for Dr. David Jefferson, Chief of Information Systems Engineering Division, CSL, who originally suggested this proposal. Dr. Stuart Katzke, Chief of Computer Security Division, CSL, also made significant contribution to the proposal. Many people assisted us in our assessment of relevant technologies and products including Gertrude Sherwood, Miles Smid, Bill Higgins, Charles Sheppard, and Kathryn Harvill, to whom we are deeply indebted.

DISCLAIMER

Several vendor products and companies are mentioned in this document. Mention of these products is for descriptive purposes only and does not constitute a NIST endorsement nor does it imply that these products are necessarily the best available for that purpose.

EXECUTIVE SUMMARY

This report is the preliminary study on the feasibility and possible use of electronic research notebooks (ERN) for NIST scientists.

In a memorandum for NIST staff from Raymond G. Kammer dated June 7, 1993, Mr. Kammer instructed each laboratory director to develop a policy for research notebooks in his operating unit. As stated in a July 24, 1992 memorandum on the general policy on research notebooks at NIST, all NIST employees engaged in research are responsible for recording and maintaining a thorough and accurate record of their work by maintaining research notebooks. The general policy also stated that notebooks using electronic media need to be retrievable as a chronological record and as an index to electronic work files.

In today's workplace where more and more professionals are using computers in performing their tasks, keeping research notes in a separate handwritten notebook adds unnecessary overhead. This project is to investigate how research notebooks can be maintained, retrieved, cross-referenced, archived and legalized electronically. Using contemporary computer security technology, electronic notes can also provide unforgeable time-stamping, sequencing, and signing of entries, as well as protection against unauthorized disclosure and modification of contents. The purposes of the ERN project are:

- (1) to survey the notekeeping practices of NIST scientists;
- (2) to determine the requirements for electronic notekeeping;
- (3) to assess current technologies for the design of the ERN; and
- (4) to develop functional specifications of the ERN.

Thus, this project explores the technical means, without establishing the policies, for notekeeping in laboratory and technical units at NIST. It needs to be complemented with appropriate policies for electronic research notebooks. The responsibility for developing and enforcing these policies lies with each laboratory director.

Survey Findings

Although all of the scientists interviewed keep some form of research notebooks, their notekeeping practices and styles vary. The majority of them keep a combination of handwritten notes and electronic notes. Most scientists consider notekeeping to be solely for documenting important research progress, procedures and results. Presently, security and legal concerns are not considered a major issue, although this may change in the future. The most important requirement that was expressed for an ERN is **ease of use**. Information must be entered in an effortless manner and must be cost-effective in order for the ERN to be acceptable to the scientists and managers.

Basically, the ERN must do everything that a pencil and paper can do, with value-added. Writing and drawing in a paper-based notebook will be replaced by an ERN system with text-and image-processing capabilities, and possibly with a scanner as well. The tools for making different types of notebook entries must be provided and integrated so that it is possible to link different objects together and still allow the retrieval of a specific object if desired. Each committed entry should be signed by the author using digital signature technology. The signed entry should also be sent to an independent time-stamping server in order to obtain an official time-stamp.

Technology Assessment

Based on the requirements gathered from the survey, a set of enabling technologies were identified and assessed. These include:

- . the electronic notepad and PDA technology;
- . the handwriting recognition technology;
- . the optical character recognition (OCR) technology;
- . the multimedia technology;
- . the object-oriented technology;
- . the database management technology;
- . the digital signature technology; and
- . the time-stamping technology.

Commercial products that are relevant for the design of the ERN were also reviewed.

Functional Specifications

Instead of prescribing a single system configuration for all scientists at NIST, we recommend that the ERN be implemented as a software application and be portable to different platforms. If there is a need to carry the ERN around, a light-weight notebook computer may be considered for installing the ERN. Regardless of the platform to be used, an operating system that supports Graphical User Interface (GUI) is essential due to the various types of input possibilities. As an option, a scanner with associated OCR software may be added to provide an alternative for making ERN entries. We also recommend that an independent time-stamping service be installed. The computer hosting the ERN must be able to access the time-stamping server, preferably through a network connection.

The acceptance of the ERN depends greatly on the tools provided for its users. Not only must the appropriate tools be provided and easy to use, but they need to be integrated as a package. An ERN can be designed to be a simple application, or it can be built to have a powerful data tracking capability, in which case an indexing scheme and database management functions can be added to the basic ERN.

Building indices is essential for context retrieval. Information that is crucial for searching should be included in the indexing system. Indices may be built on ERN contents or on the information that the ERN uses to provide a data tracking capability. A good design of the ERN should consider the needs of its users, and provide a means for its users to customize functionalities suitable for them.

Future Tasks

Since the majority of the NIST scientists use computers, it is feasible to have their research notes maintained electronically. Many products that were reviewed seemed promising for the design of the ERN. These include pen-based notebook computers, multimedia stations, and elaborate database management tools. However, none of these products presents an integrated solution for immediate use. Specific applications have yet to be developed.

There are other technical issues that need to be addressed, since many of the technologies reviewed are still new and developing. The use of any one of these in a single project, let alone using all of them, involves and raises many design issues. Obviously there is still much research to be done.

Additional research and a prototype implementation are necessary. This includes the investigation of full multimedia support for ERN entries, the linking and signing of different types of objects, the logical storage of ERN contents, and the indexing scheme to be used.

TABLE OF CONTENTS

| | |
|--|----|
| 1. INTRODUCTION | 1 |
| 1.1 Background | 1 |
| 1.2 Research Approach | 1 |
| 2. DETERMINATION OF ERN REQUIREMENTS | 3 |
| 2.1 NIST's Policy on Research Notebooks | 3 |
| 2.2 Survey of Current Notekeeping Practices | 3 |
| 2.2.1 Survey Questionnaire | 4 |
| 2.2.2 General Survey Findings | 4 |
| 2.2.3 Specific Needs and Concerns for ERN | 7 |
| 2.3 Required ERN Features | 9 |
| 2.3.1 Basic ERN Features | 9 |
| 2.3.2 Enhanced ERN Features | 12 |
| 3. ASSESSMENT OF MARKET TECHNOLOGY | 14 |
| 3.1 Electronic Notepad and PDA Technology | 14 |
| 3.1.1 Digitizer Tablets | 15 |
| 3.1.2 Tablet Computers | 15 |
| 3.1.3 Pentop Notebook Computers | 16 |
| 3.1.4 Personal Digital Assistants (PDAs) | 17 |
| 3.2 Handwriting Recognition Technology | 18 |
| 3.2.1 Printed Handwriting Recognition | 18 |
| 3.2.2 Cursive Handwriting Recognition | 18 |
| 3.3 Optical Character Recognition (OCR) Technology | 19 |
| 3.4 Multimedia Technology | 20 |
| 3.4.1 Multimedia Hardware | 20 |
| 3.4.2 Multimedia Software | 21 |
| 3.5 Object-Oriented Technology | 22 |
| 3.5.1 Overview of Object-Oriented Technology | 22 |
| 3.5.2 Applicability to ERN | 23 |
| 3.6 Database Management Technology | 24 |
| 3.7 Digital Signature Technology | 25 |
| 3.7.1 Overview of Secret-Key and Public-Key Cryptography | 26 |
| 3.7.2 Implementation of Digital Signature Algorithm | 27 |
| 3.8 Time-Stamping Technology | 28 |
| 4. FUNCTIONAL SPECIFICATIONS OF THE ERN | 30 |
| 4.1 System Configuration | 30 |
| 4.2 Functional Specifications | 31 |
| 4.2.1 ERN Input Capability | 31 |
| 4.2.2 Storage and Retrieval of ERN Contents | 32 |
| 4.2.3 Time-stamping Service | 34 |

| | |
|---|----|
| 5. FUTURE WORK | 36 |
| APPENDIX A - NIST RESEARCH NOTEBOOKS POLICY | 37 |
| APPENDIX B - ERN QUESTIONNAIRE | 38 |
| APPENDIX C - INTERVIEWED SCIENTISTS | 42 |
| APPENDIX D - PRODUCTS REVIEW | 43 |
| LIST OF ACRONYMS | 54 |
| REFERENCES | 55 |

1. INTRODUCTION

1.1 Background

This report documents the preliminary study of the feasibility and possible use of electronic research notebook (ERN) at the National Institute of Standards and Technology (NIST). The project started in January 1993 after the approval of a proposal for the NIST Director's reserve funds.

The proposal was initiated in a response to a NIST memorandum (dated July 29, 1992) on the general policy of research notebooks. The policy stated that all NIST employees engaged in research must record and maintain a thorough and accurate research notebook. The general policy also stated that notebooks using electronic media must be retrievable in chronological order and must also provide indices to electronic work files. The policy was reiterated in another memorandum from the acting Director Raymond Kammer on June 7, 1993. In that memorandum, Mr. Kammer instructed each laboratory director to develop a policy for research notebooks in his operating unit.

The ERN project was proposed to apply current information management and computer security technologies to the design and development of an electronic analogue of a research notebook, and to determine if the benefits of an ERN outweigh those of traditional paper-based notebooks. In addition to fulfilling the record-keeping requirements of hardbound notebooks, the ERN will allow automated storage, cross-referencing, retrieval, and archiving of research related material including large volumes of raw data. Using contemporary computer security technology, the ERN could provide unforgeable time-stamping, sequencing, and signing of entries; and protection against unauthorized disclosure and modification of the ERN contents.

The approach to the project was as follows: The requirements of hardbound notebooks were reviewed where required properties and desirable features were determined. In order to better understand specific electronic notebook requirements, nineteen NIST scientists from different laboratories were interviewed. The interviews resulted in a list of basic requirements and potential enhancements of the ERN. To meet these requirements, current technologies and commercial products were also assessed. Based on the information obtained, functional specifications of the prototype ERN that can meet basic ERN requirements were developed.

The Information Systems Engineering Division and the Computer Security Division are collaborating in this research effort by applying database management technology, object-oriented technology, and digital signature technology in the design of the ERN. Other collaborators also include NIST research scientists from different disciplines. Their insights and comments are essential for the successful design, development, and utilization of the prototype ERN.

1.2 Research Approach

To develop the functional specifications of the prototype ERN, the following steps were taken:

Step 1 - Determine the Requirements of the ERN

In this step, NIST's research notebook policy was reviewed and current notekeeping practices were assessed. A survey questionnaire was developed for the interviews with NIST scientists. The interview notes were analyzed and specific needs and concerns for ERNs were identified. From these findings, required ERN features were separated into two types: the basic requirements and the enhancement features.

Step 2 - Assess Enabling Technologies and Available Products

In this step, current technologies relevant for the design of the ERN were assessed. Specifically, these include: the electronic notepad and Personal Digital Assistant (PDA) technology, handwriting recognition technology, optical character recognition technology, multimedia technology, object-oriented technology, database management technology, digital signature technology, and time-stamping technology. Commercial products were also assessed for their feasibility for ERN implementation.

Step 3 - Specify System Configuration and Functional Specifications

This step specified a potential system configuration for the ERN which was considered suitable for the NIST environment, though flexibility in this configuration was possible depending on an individual user's needs. Functional specifications were also defined for an ERN that had the basic ERN capabilities as described in Step 1.

These steps correspond one to one with Sections 2, 3, and 4 of this document.

2. DETERMINATION OF ERN REQUIREMENTS

This section presents the results of a survey of the notekeeping practices of NIST scientists in order to find out specific requirements that the scientists may have for electronic research notebooks (ERNs). In the process, NIST's policy on keeping research notebooks (Appendix A) and other references regarding the maintenance of laboratory notebooks [KANA85, NAS89, NBS83] were reviewed. A sample questionnaire was also developed and used in face-to-face interviews with scientists from NIST's seven organizational Laboratories. The basic requirements and major concerns pertaining to the ERNs are summarized later in this section.

2.1 NIST's Policy on Research Notebooks

According to NIST memoranda issued on July 29, 1992 and on June 7, 1993, all NIST employees engaged in research are responsible for recording and maintaining a research notebook. The policy, included in Appendix A, is summarized as follows:

- Notebooks can be in the form of written pages using hardcovered or spiral bound notebooks with sequentially numbered pages, such as the types available from the NIST storeroom.
- A notebook using electronic media needs to be retrievable as a chronological record and as an index to electronic work files.
- Each page of the notebook should be signed and dated by the author. Two co-workers must also sign and date each page to signify that they have read and understood the notebook entries.
- All research notebooks and electronic files pertaining to NIST activities are official records of the U.S. Government and, as such, are the property of the Government. These records are not to be removed from NIST without proper authorization.

The primary reasons for notekeeping are to accurately document scientific investigation, protect one's intellectual properties, and prove the integrity of work in the event of inventions and patent applications. Reference [KANA85] provides a detailed discussion of good notekeeping practices.

2.2 Survey of Current Notekeeping Practices

The first phase of the ERN project involved the collection of information on how NIST scientists use their notebooks and what scientists would like to see in the design of an ERN.

2.2.1 Survey Questionnaire

To determine the scientists' current notekeeping practices and their requirements for the ERN, a questionnaire was developed (as shown in Appendix B) and used in interviews with the NIST scientists. The questionnaire covered the following topics:

Current Notekeeping Practice

- Primary purpose of notekeeping
- Notebook issuance
- Types of notebook entries
- Notebook maintenance

Computer Familiarity and Preferences for the ERN

- Computer literacy
- Preferred platform for the ERN
- Software environment
- Programming languages and tools used

Visions of the ERN

- Minimum requirements
- Desirable features
- Specific concerns (including security concerns)

Managerial Questions

- Acceptable cost range
- Operation requirements

2.2.2 General Survey Findings

With the assistance of the upper management at NIST's organizational laboratories, a list of scientists was suggested for interviews (Appendix C). An important aspect of the interviews was to learn about current notekeeping practices and to determine how feasible it would be to automate those practices. The interview findings are summarized as follows.

Finding 1: Diverse notekeeping needs and styles

Most scientists have developed personal notekeeping styles, and their needs and practices vary. There is no one particular format which pleases all respondents. However, a unanimous requirement for the ERN is for **ease of use**. Even though many scientists realize that user-friendliness may be a demanding requirement with the current technology, an ERN that is restrictive and difficult to use is considered worthless for recording their work.

Finding 2: Different usage of notebooks

Most notebooks used by NIST scientists are hardcovered or spiral bound. Many scientists also use three-ring binders and file folders because it is easy to insert and file other papers, such as correspondence, loose-leaf meeting notes, etc. The usage of notebooks varies and can be classified as follows:

1) Public notebooks - These notebooks are frequently associated with a particular instrument or a specific material. Many scientists use public notebooks to record instrument configuration, control settings, service and repair log records, usage and location of samples, etc. A few characteristics of these public notebooks are:

- They are often kept close to the equipment. If the equipment is moved, so are the notebooks.
- The entries are normally quite legible since they are intended to be read by others.
- The data entries can be quite varied, with some entries consisting of a one-line entry, while others may contain textual information and pictures. There is not a "standard" format defined.
- Each entry is dated, but rarely signed. Most scientists claim that they recognize the handwriting of their colleagues.
- In laboratories that comply with the ISO/IEC GUIDE 25: 1990 ("General requirements for the competence of calibration and testing laboratories"), these public notebooks are very useful in providing documentary evidence and audit trails.

2) Project notebooks - A scientist may keep several project notebooks - one for each of their projects. These project notebooks are used to record procedural descriptions, observations, and calculations for a specific project. The characteristics of these notebooks are:

- The entries may contain text, equations, and sketches which are typically handwritten by the scientist.
- Photographs, instrument printouts, computer printouts, and spreadsheets are frequently pasted in the notebook.
- References are frequently made to external objects such as electronic files, sample materials, and spectrographic plates.

- Entries are referenced by date or by visual recognition of a particular page layout when browsing.
- Both completed and current notebooks are often kept close to where the experiments take place or in the scientist's office for easy retrieval.

3) Personal Notebooks - Some scientists also keep a personal notebook for recording telephone conversations, meeting notes, and actions items. Characteristics of such notebooks are:

- They are often carried around to meetings in different places.
- They are sometimes used as "chalkboards" during research interaction with peers.
- The entries are typically scribbled by hand in text or as diagrams.
- These notes are more for private use rather than for public review.

Finding 3: High computer literacy

Most scientists we interviewed are skilled computer users and have access to a personal desktop computer which generally is an IBM-PC (or PC compatible), a Macintosh, or a UNIX-based workstation. The majority of these machines are connected to a local area network. Many laboratory instruments are connected to computers and a network, but stand-alone systems are also numerous.

Software tools are abundant, word-processors are essential, and database management systems (dBASE, Paradox), spreadsheets (Lotus 1-2-3, Excel), Windows tools, statistical packages (SPSS), and other special-purpose tools are also used widely.

Finding 4: Notekeeping solely for recording purposes

Almost all of the scientists interviewed claimed that the purpose of maintaining research notebooks is solely for record keeping. Keeping notes helps them keep track of important research procedures, major research events, significant findings and discoveries. Notebooks are frequently used and referenced - even notebooks dated twenty years ago. However, none of the scientists interviewed used or intended to use notebook entries to support patent claims. Though the attitude may change in the future with respect to ERNs, the current consensus is that notekeeping is solely for record keeping.

Notebook security is not currently a major concern. Specifically, no scientists felt that there was a need to hide or "lock up" their notebooks unless the project was classified. They would be more concerned if a notebook was misplaced than if some notes were altered. If notes were altered, chances are that the alteration was done by a colleague,

and most scientists claimed that they had no problem recognizing the handwriting of their colleagues.

Most scientists review and modify notes. However, this practice does not involve erasing the previous entries, but rather striking out old text and dating the modification, often with a different color pen.

Finding 5: Need for notebook storage, archival and retrieval

Scientists generally store laboratory notebooks in easily accessible places since notes are typically taken daily. Project notebooks can be kept at the scientist's desk or in a location where research is performed. Most scientists allow their peers to have access to their notebooks. Public notebooks usually stay in the lab where experiments or testings take place. If a notebook is associated with a particular item of equipment, the notebook remains with that equipment.

When notebooks are completed, they may still be referenced regularly. Therefore, archiving a notebook usually does not involve more than placing it on an office shelf or in an easily-accessible filing cabinet.

Many of the scientists that were interviewed regularly backup their computer files. The procedure usually takes place over the network, and uses such storage media as magnetic tapes, 8mm digital tapes, or optical disks. Most scientists expressed the need for the proper storage of information in ERNs, such as automatic backup and the archival of ERN files.

2.2.3 Specific Needs and Concerns for ERN

Presented in the previous section were a few of the observations made during the interviews, which provided a glimpse of the computing environment accessible to scientists, and a general understanding of their current notekeeping practices. This section will focus on the scientists' expectations and concerns with regard to using an ERN. The discussion will start with the scientists' wish list, followed by their concerns with ERN use. The needs and concerns specified in the following are as stated by the scientists interviewed.

Specific Needs for the ERNs:

- **Ease of use** - this is the only unanimous requirement of the ERN that was expressed;
- Should have tools available to facilitate the input of text, equations, and drawings;
- Should be able to accept scanned-in documents which might include handwritten equations and other papers;

- Should have the capability to input and edit video and imagery;
- Should be able to integrate with applications already in use;
- Would be useful to have an electronic pen capability for entering data while using an instrument, such as a microscope;
- Should have a powerful indexing scheme for fast and easy retrieval;
- Should have an object linking capability to permit linkage to other electronic files;
- Should have the capability to modify earlier entries;
- Should have an automatic and transparent time-stamp;
- Backing up of the ERN should be as transparent as possible;
- Electronic bookmarks would be useful for visually browsing through entries;
- Should have a math coprocessor for complex calculations (note: this requirement stems from older PC chip architectures and is no longer valid under the new chip architectures for PCs.);
- Should have a large amount of storage space;
- Would be nice to have a powerful processing capability;
- Should have proper access control;
- Would be nice to have a large color screen and full-size keyboard;
- Would be nice to have an option to encrypt entries;

Concerns and Possible Difficulties:

- To justify the use of the ERN over the much less expensive paper-based notebook, the ERN should be easy to use and yet provides functionality which is not available in a paper-based notebook.
- How usable will the ERN be after five or ten years? Will its storage media be accessible, and will the ERN itself be human-readable?
- Having data stored on multiple machines which are not necessarily compatible may cause serious incompatibilities while transferring data to the ERN.

- Scientists are comfortable with particular computer environments and don't want to be forced to use different ones, so the ERN should be portable across various platforms.
- The ERN must ensure data integrity and provide protection from unauthorized access. The ERN must guarantee to laboratory auditors and scientists alike that the computer records are secure. In particular, much of a scientist's career work is stored in his notebooks.
- The ERN should be cost-effective for wide acceptance. The unit cost as well as the staff training cost should be very small.
- If the ERN involves new hardware, it must be small and light, due to space constraints in the laboratories and offices. The ERN must also be able to sustain harsh laboratory and field conditions.
- How will changing the path of a data file's location affect references to that file in the ERN?
- Is the ERN intended to help NIST scientists do their work more efficiently? If the ERN is implemented, will all NIST scientists be required to use it?

2.3 Required ERN Features

From the survey with NIST scientists, the requirements for the ERN were determined. These requirements are divided into two groups: basic and enhanced features. The basic features are essential capabilities to prompt scientists to convert to ERNs from today's paper-based notebooks. The interviewees also contributed ideas for enhanced features which would allow for incorporating some new emerging technologies in the design of the ERN.

2.3.1 Basic ERN Features

The basic requirements identified for an ERN are:

- Customizable format
- Word processing capability
- Sketching/drawing capability
- Ability to scan documents and images
- Searching capability
- Object linking capability
- Ability to redline or append data to an entry
- Time-stamping capability
- Digital signature capability

- Storage and archival facility
- Printing capability

Each of these is discussed in the following paragraphs.

Customizable format

Scientists like to take notes in an individual style which is convenient and comfortable to them. Of all the notebooks we saw during the interviews, no two were alike. The ERN will therefore have to allow each user to create his custom format for making notebook entries.

Word processing and sketching/drawing capabilities

For the scientists to record procedures, data, and observations, the ERN must have a word processing functionality. Creating sketches and graphs in the laboratory notebook is almost as important as entering text, so a sketching capability which is integrated with word processing is crucial to the success of the ERN. It would be ideal for the ERN user to be able to enter text and drawings on the same "page", so that the information can be kept together. That user should not be requested to enter separate applications to make each type of entry.

Ability to scan documents and images

Many ERN users have a need to enter equations or hand-drawn sketches. If this cannot be done using the word processing feature, then the scientist should have the option of scanning handwritten materials into the notebook from a piece of paper. Other paper documents, such as correspondence, memoranda, and publications, could also be scanned and placed in the main body of the notebook. It might be desirable to have the user to include a short caption each time a scanned image is placed in the notebook in case the image is not self-explanatory.

Searching capability

Most scientists commented that it is usually easy to find particular entries in a paper-based notebook by flipping through pages. Comparing multiple entries which are not necessarily on adjoining pages is also possible. With an electronic notebook, such a powerful referencing capability can be provided with hypertext, indexing, or a text search engine. This also gives the user an indexing tool which is not commonly found in paper-based notebooks. Scientists often use bookmarks or highlighting to quickly search for some important facts. These features should also be available in the ERN by using electronic bookmarks.

Object linking capability

To take advantage of the fact that most scientific work involves computers, the ERN should allow the user to create links to that work. For example, if a scientist makes an entry that refers to a particular data file from a laboratory computer, it should be possible to include a

pointer to that file in the notebook. The inclusion of an entire data file in the notebook would result in unnecessary replication of data, as well as an ERN of enormous size. If only specific parts of a file are needed for inclusion in an ERN entry, then the user should be able to copy that data and paste it in the notebook. Hypertext could provide object links.

Ability to redline or append data to an entry

An important aspect of the written notebook is that entries are written in ink, which prevents erasure. When a correction is made, the author strikes a line through the error so that the text remains readable, adds the correction, and dates the change (the author may or may not initial the change). A similar feature is needed in the ERN that would allow for dated and signed addenda, which would also prevent the deletion of previously signed material.

Time-stamping capability

The NIST policy on keeping research notebooks requires each notebook entry to be dated by the author. To meet this requirement as well as to prevent forgery, the ERN should be automatically time-stamped by an accurate and trusted Time Stamping Service. This procedure should be as transparent as possible to the user and should be used each time the user completes an entry.

Digital signature capability

Officially, a notebook must contain the signature of the author after each entry. In the case of the ERN, a digital signature can be used as a replacement of the handwritten signature. The topic of digital signatures is addressed in Section 3. Notebook entries should be signed automatically by a trusted third party, much like the trusted time-stamping service. The author may want to sign his/her notebook entry before sending it to the trusted signing authority. Not only can a digital signature be used to authenticate the author of an entry, but the digital signature can also be used to check the integrity of the entry. Precautions must be taken to ensure that a previously signed entry cannot be modified and then re-signed.

Storage and archival facility

Electronic notebook entries should be easy to retrieve. For permanent archiving, ERNs should be stored on a medium which can be readable over a long period of time, but which cannot be overwritten. Cryptographic elements necessary for signature verification (e.g. keys) should be archived for the lifetime of the archived ERNs.

Printing capability

Although one goal of the ERN is to minimize the scientist's dependence on paper-based documents, allowing the user to print hard copies of ERN entries would help smooth the transition from a paper-based to a fully automated notekeeping system. Having a hard copy is

comforting and familiar to most computer users. In addition, a scientist may wish to either show an entry to another person who cannot electronically access the ERN, or take notebook information into an environment which is not suitable for the ERN.

2.3.2 Enhanced ERN Features

Based on current technology, the previous section showed the basic features of the ERN. These features seem to be more achievable today than the features addressed below. In this section, enhanced features that can make the ERN more powerful and easier to use are addressed. However, the enabling technologies for some of these features may not exist or need to evolve further. Section 3 contains an assessment of current technology related to ERN. The desirable ERN features for enhancement include:

- Electronic stylus for entering text and drawings
- Multimedia capabilities
- Central database for ERN storage
- Integration with external software
- Visual browsing capability

Electronic stylus capability

In a laboratory environment, scientists must often record data when it is inconvenient to use a keyboard. For example, a scientist may be working at a microscope or adjusting instrument settings. Using a keyboard, which normally requires both hands, may not be desirable at such times. It might be easier if the scientist has the option of using an electronic stylus to make entries. By using a pen-based application within the ERN, one could use handwriting recognition software to convert stylus-entered electronic ink into text, or simply save the electronic data as an ink type (bit map). In this way, textual data and sketches can be entered side by side in the ERN with little effort on the part of the scientist. However, it should be noted that handwriting recognition technology is still developing, and the capability of recognizing cursive handwriting is quite limited.

Multimedia capabilities

Multimedia features would add flexibility to the ERN. Many scientists currently include photographs in their notebooks on which they often make marks and notations. These photographs could be scanned into the computer and represented as a bit map in the ERN entry. The entry could then be edited or annotated by the user. A similar capability with video input would be desirable which not only allows video input, but also allows the editing of the video imagery. With paper-based notebooks, there is simply no way of including video imagery other than by making a reference to a video file.

Voice input, another multimedia feature, would allow the user to enter data by simply speaking into a microphone. Enabling lab scientists to enter data while doing something else

could be an extremely valuable feature of the ERN. Though speech recognition technology, which converts spoken language into text, is fairly primitive at the moment, voice objects can always be represented as digital data and adds another multimedia capability to the ERN.

Central database for ERN storage

Collaborative research efforts would benefit greatly from the maintenance of an ERN database on a local area network. This would be ideal for sharing public notebooks in which multiple scientists make entries. Once an entry is signed, permissions can be automatically set to make the entry read-only for all users. However, the original author should be able to add a signed addendum to the entry if the need arises. It may also be beneficial to include single-author project notebooks in this database. Scientists often find it useful to share information by consulting a peer's notebook during information searches and work flow coordination. In an electronic environment, a centralized database for the ERN that permits simultaneous access over a network would be most convenient.

Another important reason for storing the ERNs in a central database is that the central database management system can provide the crucial time-stamping and digital signature capabilities for the ERNs.

Integration with external software

Most scientists use a variety of software tools such as spreadsheets, statistical packages, finite element analysis packages, etc. Ideally, these should be linked to the ERN in order to integrate important data and prevent needless repetition.

Visual browsing capability

Since many scientists use visual browsing to search for particular notebook entries, it would be useful for the ERN to provide a similar capability. This may be done by placing a series of thumbnail sketches of consecutive entries on the screen. Using visual recognition as a search mechanism might prove to be faster in some cases than using hypertext to retrieve important data.

3. ASSESSMENT OF MARKET TECHNOLOGY

In the previous section, the required ERN features were presented. In this section, current technologies as well as commercial products that are relevant for the design of the ERN are reviewed. The enabling technologies to support the ERN design include:

- Electronic notepad and PDA technology;
- Handwriting recognition technology;
- Optical Character Recognition (OCR) technology;
- Multimedia technology;
- Object-oriented technology;
- Database management technology;
- Digital signature technology;
- Time-stamping technology.

For each technology reviewed, a brief overview is provided, and commercial products are identified. Detailed product descriptions are provided in Appendix D. It is impossible to conduct an exhaustive survey and evaluation of all possible products due to the vast number in the marketplace and the rapid rate at which technology evolves. It should be noted that the mention of specific vendor products is for description purposes only and does not constitute a NIST endorsement.

3.1 Electronic Notepad and PDA Technology

The term "electronic notepad" refers to a computer or peripheral device which functions in a manner similar to a conventional notebook that involves paper and pen. In an electronic environment, a digitizing surface and an electronic stylus (pen) interact to create digital information, which is sent to the application being used. To make sense of the digital input, the application in use must be made to recognize interactions between the electronic stylus and the digitizing surface. This requires support from the operating system.

The most common configuration of digitizer and stylus is one in which a cordless pen sends digital data to the processor via the digitizer. The stylus has the look and feel of an ordinary pen, and is not attached to the computer with any type of cable. Some styluses are pressure-sensitive, and the digitizer can adjust the width of the electronic "ink" based upon how hard the pen is pressed against the digitizing surface.

Stylus functions may include dragging and dropping objects, selecting an object by pointing or tapping, or inputting data by writing or drawing as if the stylus was an ordinary writing device. Depending upon the application, the input data may either be stored as an image or converted to actual text.

The integration of the stylus, digitizing surface and computer is accomplished in various ways. Four such products are discussed in the following sections: digitizer tablets, tablet computers, personal digital assistants (PDAs), and pentop notebook computers. One type of technology which is essential to the latter three devices is the PCMCIA (Personal Computer Memory Card International Association) Type II card. The PCMCIA card is capable of storing large amounts of information and functionality on a very small token which is about the size of a credit card, but thicker. The card can be used by any machine with a PCMCIA Type II slot, in which the user plugs the card. It has many uses, including data storage, modem and network connectivity, and a system-level interface capability such as the Small Computer System Interface (SCSI). Since PCMCIA cards can currently hold up to 20 MB of information, they may be used in place of floppy disks to distribute applications. In addition, information stored on PCMCIA cards can be accessed much more quickly than data found on ordinary floppy disks.

3.1.1 Digitizer Tablets

Digitizer tablets are thin, flat screens which are similar to mouse pads. They range in size from under a square foot to well over three square feet. The larger ones are typically used in conjunction with Computer Aided Design (CAD) packages. Various drivers are shipped with the digitizer, depending on the platform in use. The digitizer is connected to the computer with a cable. The type of cable is also determined by the platform in use.

Most digitizer tablets are input-only devices. To see what effects the pen is having, the user must watch the computer's monitor while writing on the tablet. However, there are some digitizers which integrate the tablet and screen, and the user appears to be writing on the monitor. In fact, this is the technology used by each of the products discussed in the following three sections.

Smaller digitizer tablets weigh less than ten pounds, and the ones with the smallest footprint weigh less than five. Such devices are available for under a thousand dollars. These factors would allow ERN users to inexpensively add the pen capability to their existing systems without sacrificing a large amount of desk- or laboratory-bench space.

3.1.2 Tablet Computers

Tablet or notepad computers combine the digitizer tablet, display, and processor in a single device. Oriented primarily toward pen-based applications, tablets have external keyboards as options. They are less than a square foot in area and weigh under six pounds. Physical portability is a distinct advantage for these machines. However, reduced storage space is a trade-off for this increased mobility.

Currently, hard disk drives in tablets are less than 100 megabytes, and PCMCIA Type II cards can hold at most 20 megabytes of data. As technology advances, though, so do storage capacities (last-generation tablets typically had 20 or 40 megabyte hard disks). Most tablet users also have the option of connecting an external floppy disk drive. Other peripherals can be connected to the tablet by means of serial, parallel, external video, and mouse ports.

When using the stylus, the user writes directly on the screen and can immediately see the results of the pen's actions. The typical display is black and white, with at least 640 x 400 pixel resolution.

If the ERN is to include pen technology as a significant feature, then the powerful processors, optional keyboards, and PCMCIA support in tablet computers make these highly mobile computers prospective candidates for the ERN. A large capacity hard disk (at least 80 megabytes), a 486/20 MHz processor, and two PCMCIA slots (one for storage, another for establishing a network connection) should be the minimum requirements for an ERN tablet computer. Tablet computers do have a few drawbacks: the lack of color screens, relatively small screen size, and limited storage, which may be the limiting factors on the integration of multimedia in the ERN. However, as technology advances, these problems may be solved.

3.1.3 Pentop Notebook Computers

The pentop notebook is essentially a notebook computer with a screen that is also a digitizing surface. This pentop notebook can be used as an ordinary notebook computer and becomes a tablet by simply sliding the screen down onto the keyboard.

Current pentop computers include one or more slots for PCMCIA Type II cards which can be used for additional memory, network and modem connections, and possibly handwriting recognizer in the future. Reasonable storage space is also provided by a 120MB hard disk, at a minimum. A cordless stylus is included which functions both as a pen and a mouse. Serial, parallel, and external video ports are all standard features. A math coprocessor, an internal fax/modem, and RAM upgrades are available options.

The present generation of pentop computers has several drawbacks, including 386 processors and gray-scale displays. The next logical step in pentop evolution is the implementation of a 486 processor. This processor is estimated to have twice the processing power of its predecessor. Gray-scale displays are the current norm for pentops because of the difficulty in combining color matrix displays with digitizers, due to the thickness of each. To work in color, current pentops can be connected to an external color monitor, and used as an integrated digitizer tablet.

Future generations of pentop computers may prove to be ideal candidates for implementing the ERN, given certain improvements such as more powerful processors, color displays, additional storage space, and advances in handwriting recognition and PCMCIA technology. The fundamental tools for building a robust system are already in place on pentop

systems: keyboard input, ease of switching between textual and pen input, PCMCIA slots for various interfaces (ethernet, modem, SCSI), mobility, and durability.

3.1.4 Personal Digital Assistants (PDAs)

The newest and smallest type of pen computer is known as a Personal Digital Assistant (PDA). It is geared towards being a portable Personal Information Manager (PIM), which allows the user to collect and organize valuable information at almost any time or place. PCMCIA cards serve as storage and communication devices for PDAs, and so as PCMCIA technology advances, PDA capabilities will increase. PDAs also make use of wireless communications, pen technology, and specialized software and operating systems.

PDA users can write on a small Liquid Crystal Display (LCD) screen to sketch, take notes, send a fax, and maintain an address book. A handwriting recognizer, to be discussed later, can convert printed characters to text, and PDAs can use a graphics recognizer to interpret and neatens sketches.

Pen-based computing is a new technology that has some drawbacks. The extremely small black-and-white screens used by PDAs make using a keyboard impractical, if not impossible. Therefore, instead of connecting a keyboard to the PDA to enter typed text, the PDA's contents (which are stored on a PCMCIA card) must be downloaded to a desktop personal computer for editing.

An ERN using PDA technology would be very portable and easy to use in the field or in the laboratory. However, there are concerns with durability and limited battery life.

It is still undetermined how easy it will be to exchange information between a PDA and a computer attached to a laboratory instrument, or if it is even possible. A scientist might use the PDA to take notes while on travel or in the lab. These notes can later be downloaded from the PDA to the scientist's desktop computer, and integrated with other files or notes.

Current pen-based computing with PDAs is mostly being used to assist in the management of a person's private information. It is especially useful for people who are on travel and need to retrieve information, access personal electronic mail, take notes or do simple calculations. These PDAs may not be adequate to provide several basic ERN features. Some PDAs may not have sufficient storage space to perform lengthy text or graphics retrieval. It is also uncertain how a central database would be maintained and accessed by pen-based computing devices. This technology is still in its infancy, which makes it unclear how useful it will be for the ERN development.

3.2 Handwriting Recognition Technology

For most people, entering text using a keyboard or typewriter is quite easy. However, when a scientist needs to enter an equation or draw some diagrams, one soon realizes that the easiest way to do such things is to use a pen and paper. Though there are software tools that allow a user to draw graphics or enter mathematical equations, it takes time to learn and use these tools. Moreover, for certain tasks, even if the tools are easy to use, these tools may just not be as convenient to work with as pen and paper. Mathematical equations are one example. While one can always use a powerful word processor to enter an equation, entering the superscripts, subscripts and unique mathematical symbols in the equation can be quite tedious. For this reason, handwriting input would be very desirable in ERNs, since it allows the user to interact with the computer in a natural way.

Pen-based computer systems have been reviewed in the previous section. The technology of handwriting recognition is assessed in this section, followed by an assessment of the optical scanner and optical character recognition (OCR) technology. The scanner and the OCR software are also important for the ERN because, if it is decided that the easiest way to input data is by means of pen and paper, then one may want to take notes by hand and scan the finished notes into the computer. The electronic version of these notes could then be edited using the extensive power available under word processing technology.

Research efforts in pen input recognition have been underway since the 1970's, particularly in Japan and China because of the non-alphabetic nature of their handwritten characters. Much of the delay in implementing pen input has been attributed to the lack of an adequate handwriting recognizer. Due to the complexity involved in electronically recognizing different types of handwriting, separate recognizers have been developed for printed and cursive handwriting. Recognizers for cursive writing are much more difficult to develop than for printed characters.

3.2.1 Printed Handwriting Recognition

Promising advances have been made in printed handwriting recognition systems. Particularly useful for form-filling applications in a pen-based system, a printed handwriting recognizer can be trained to convert printed characters into text. While a form is displayed on the screen, the user prints the data and fills out the form using an electronic stylus. The entered data is sent to the recognizer and converted to text or numbers.

3.2.2 Cursive Handwriting Recognition

There has been very limited success in cursive handwriting recognition. The recognition process is still in the research stage using advanced mathematical theories and artificial intelligence techniques. Cursive handwriting recognition software normally requires users to register their handwriting on the computer for later reference. Therefore, each individual can

have a customized library of handwritten characters which improves the accuracy and speed of character recognition. Some software also requires a domain-specific vocabulary so that the cursive handwriting can be semantically understood. Example applications of cursive handwriting recognition have a limited vocabulary on a simple subject matter such as an appointment schedule application.

3.3 Optical Character Recognition (OCR) Technology

It may be desirable at times to take notes by hand and have the handwritten notes scanned into the computer. Scanners are peripheral devices capable of optically scanning a printed document page by page into a computer and capturing the scanned document as images. Recently, color scanners capable of scanning colored documents are also appearing in the marketplace. Some color scanners have a resolution of 600 dots per inch (dpi) and accept 24-bit color images. Scanners only capture the scanned document's images. OCR software is needed to convert the textual images from bitmaps to a more useful format.

OCR is the process of converting hardcopy analog information into a digital bitmap form and then into a text file. Most OCR programs locate text blocks in a scan, align them if they are off baseline, then build character models. Algorithms and techniques used in character recognition include feature extraction, pattern matching, mathematical modeling and artificial intelligence. The foremost criteria of the OCR software is its accuracy rate, which is claimed to be about 95% by most of today's OCR vendors, depending on the quality and complexity of the original input document.

OCR software can offer many powerful features. For example, the ability to differentiate between different fonts, the ability to identify smudged or broken characters, dirt, staples, binder holes, etc. Different tools may be provided to process text and graphics images, and to store the resulting files separately. Bitmapped images of scanned documents usually consume a great amount of storage, yet once through the OCR process, a 1-MB image can be reduced to an average length of 10KB of text. This is a reduction of 99 percent. The resulting text file can then be processed by a variety of word processors. Indexing tools can also be built into and integrated with the OCR package to provide full-text search capability. This capability allows a user to search for a particular word or phrase using a variety of ways to express the search query.

Graphics images are usually compressed before they are stored in files. Some packages provide a limited indexing capability by allowing captions to be added to graphics images for search and retrieval. Operations for zooming and rotating the imagery can be made available. Images can also be cut and pasted to a clipboard, or linked to an object or word, thus providing a hypertext capability.

3.4 Multimedia Technology

A multimedia computer system is one that is capable of handling the input or output of more than one medium. Typically, a multimedia system can take input from a keyboard, mouse, electronic stylus, microphone, video camera, or VCR; and display output on the monitor or generate high quality audio and video output. A multimedia system would greatly enhance the functionality of an ERN, because it would include many different object types that are not found in current paper-based notebooks. Also, it would facilitate the entry of some data which may be tedious to use on a computer which does not have multimedia capabilities. For example, a scientist may want to dictate a research procedure into a recorder during an experiment, or the scientist may wish to record an invention and explain how it works on a videotape. Multimedia technology can provide the necessary tools for these undertakings.

In a multimedia system, each media represents information in a unique manner. In the case of audio or video, a multimedia system requires a series of data samples which are manipulated in succession to recreate a sound or moving picture. The integration of text with images and audio is an important requirement for an ERN.

3.4.1 Multimedia Hardware

Multimedia, as the term implies, requires a group of hardware integrated under the control of a computer. Although multimedia may pose great demands on computing power, today's PCs are adequate for general multimedia usage. Because of the wide range of equipment available on the market, building a multimedia station takes some planning and careful selection. Some of the factors that should be considered are discussed in the following paragraphs.

System CPU

The CPU chip determines the character and speed of the entire machine. A multimedia system requires a great amount of processing power, therefore, a powerful microprocessor with high clock speed is a requirement.

Random Access Memory (RAM) and Mass Storage

A multimedia system also requires vast storage, not only for storing generally huge files from some of the media, but also for the temporary storage required by the high demand in processing and performance. An important consideration is not only the amount of storage required but also the speed at which the storage can be accessed by the CPU to retrieve the stored data.

Video Display

Video displays in most PCs use a memory-mapped approach in which a region of memory stores all the pixel values to be displayed. To create the display, the video display adapter reads

through the memory at a rapid rate and converts the pixel values to signals that drive the display device. This process requires a massive amount of memory access to move data to refresh the screen. Multimedia systems need a dynamic video presentation, which requires the capability of rapidly changing the images. Therefore, the system must be able to rapidly retrieve video display information from the mass storage, process it, and pass the resulting information to the video display subsystem for presentation.

Another important consideration for a video display is the number of colors that can be displayed. This is directly related to the number of bits of data that can be allocated for each pixel. Most computers today only support 4 or 8 bits per pixel color displays. However, the display of realistic, high-resolution images requires at least 16 bits per pixel, preferably 24 bits.

Audio System

One important consideration for an audio system is the audio quality desired. To reproduce high quality sound, a large number of samples must be taken of a sound wave. Each sample uses either eight or sixteen bits to store information on a part of a sound wave. At least 20,000 samples should be taken each second (the sampling rate) to produce high-quality playback. Larger sample sizes and higher sampling rates produce digitized sound with finer resolution.

3.4.2 Multimedia Software

The quality of a multimedia system also depends on having software which allows a user to take full advantage of the multimedia hardware. The combination of both software and hardware creates a complete multimedia environment for the user, so that all multimedia elements can be accessed from a single application. The development of a multimedia project requires both editing tools that allow the user to create and edit individual multimedia objects, as well as authoring software which combines those multimedia objects into a project.

Editing Tools

Before organizing assorted multimedia objects, one may want to edit them. For example, the initial window of a multimedia project may introduce the user to the program with both animation and sound. The application's creator could customize these objects using editing tools. A drawing tool or image editor could be used to create the animated object, and a sound editor could manipulate the waveforms of a recorded sound to create the desired sounds.

Authoring Software

Multimedia elements are typically sewn together into a project using *authoring tools*. These software tools are designed to manage multimedia elements and provide user interaction. In addition to providing a method for users to interact with the project, most authoring tools also offer facilities for creating and editing text and images, and they have extensions to drive videodisc players, videotape players, and other relevant hardware peripherals [VAU93].

Generally speaking, there are three categories of authoring software: page-based, icon-based, and time-based tools. Page-based authoring systems are best-suited for applications in which the data elements can be viewed separately. A series of "pages" can be linked together in a desired order so that different events that take place on a page move the user to other pages. Icon-based systems let the author link multimedia objects in a flow chart format that displays the program's logic. The flow chart is a series of icons, each of which represents an event, that are drawn from libraries. Time-based authoring tools allow the user to set the duration and speed of different multimedia events. For example, still images can be played at a certain speed, and specific actions may activate audio events with predetermined durations.

3.5 Object-Oriented Technology

3.5.1 Overview of Object-Oriented Technology

Object-Oriented technology is a recent development in computing. It started as a programming language concept and was gradually adopted into the database world as an alternative to the relational data model. More recently, there are object-oriented approaches for performing analysis and attacking design problems. To add to the confusion, object computing is now being introduced to multimedia services, graphical user interfaces, and object linking environments for application development tools.

The basic premise with object-oriented technology is that computer systems and applications are constructed from so-called objects. Conceptually, an object is something which is perceived as an identifiable, self-contained unit, which is distinguishable from its surroundings. An object may be described by a set of attributes that constitute an internal state, and a set of externally visible operations, called *methods*, which define its behavior.

Some of the characteristics typically associated with object technology are briefly discussed below.

Message sending

In a pure "object" system, the predominant mode of computation is not with the conventional top-down program control, but is replaced by a collection of objects which send and receive messages. Communicating with an object is accomplished by sending a message to it.

An object has a set of messages to which it responds. When an object receives a message, one of its procedures (methods) is invoked, and an operation is performed which may return a result.

Encapsulation

The details of the operation performed are not specified in the message and are not visible to the message sender. An object's internal state is not accessible by an external agent, but can only be affected indirectly through message passing and method invocation. This hiding of the object's internal state is known as object *encapsulation*. The advantage of the encapsulation feature is that the implementation part of the object's internal state can be hidden while allowing the results of the computation to be performed. For example, one only needs to know when a computation is done, and not how to perform the computation.

Polymorphic Behavior

Polymorphism refers to the ability to hide different implementations behind a common interface, thus simplifying the communications among objects [TAY90]. For instance, several objects may receive the same message, but depending upon the parameters, an object may respond to the message differently and exhibit different behavior. Such binding may occur at compile time or at run time. Run time binding means that the selection of the method to respond to the message is made when the program is executed.

Classes

Objects having the same data structures and behavior can be considered as belonging to the same type, or *class*. There can be many different classes. Some objects are more specific kinds of other objects. For example, the "employee" class is a specialized class of the "person" object. It is possible to take advantage of natural relationships found in real world domains to create class hierarchies. The advantage of defining subclass (specialization) from a superclass (generalization) object is that all the class definitions in terms of attributes and methods are inherited. By extensive inheritance of methods, the amount of code needed to implement an application can be significantly reduced. For this reason, object programming is said to promote code *reuse*.

3.5.2 Applicability to ERN

The ideas that form the basis of object-oriented technology can be applied to the implementation of the ERN. For example, the ability to encapsulate and hide implementation details allows users to use and reuse object software components without having to know the details for data and operations. In particular, enhanced notebook features may be added in the future without having to modify any of the original objects. Encapsulation also permits better security for protection against alteration of data.

Another object feature applicable to ERN is the object polymorphic behavior which means software may respond to the same request, but depending upon the parameters, will perform different tasks. For example, the ERN will initially have the ability to do object linking and binding for an object data type in the "TEXT" class. At a later time, as new object data types in the form of pictures, graphs, and bitmap images are included, enhanced notebook features may support the storage and retrieval of these new data types in the form of voice and video clippings without having to reprogram the software.

3.6 Database Management Technology

The primary functions of a database management system are to store data and provide operations on the database. The operations usually include the creation, deletion, updating, and searching of data.

There are many commercially available Database Management Systems (DBMSs). Most DBMSs today are relational, which means the logical view of data to the users consists of tables with rows and columns. Queries to the relational DBMS are about sets of homogeneous types of data.

Recently, there have been a few commercial offerings of object DBMSs. An object DBMS generally supports most of the relational DBMS features, including persistence, concurrency control, recovery and database language. In addition, an object DBMS may support some or all of the following features: polymorphism, encapsulation, dynamic integrity checking, versioning, schema evolution, and change management.

Characteristics of DBMSs

Commercial DBMSs now come in all sizes and shapes. In order to choose one or more DBMSs for an ERN it is important to understand the different characteristics that can be affected by DBMS packages. Some important characteristics of DBMSs are summarized here:

DBMS for Private Use

Most PC users employ DBMS software to establish databases for their own private use. Typical features of such DBMSs are their ability to create data definitions, define a form for ease of entering data values, query and retrieve data with simple Boolean conditional statements, and produce simple reports. These DBMSs do not support multiple simultaneous accesses, and therefore do not have a concurrency control mechanism. Security is provided only at the database level with a password mechanism. Some DBMSs offer a fourth generation language to permit non-programmers to quickly establish a database for use almost immediately.

DBMS for Workgroup

Some databases provide the capability for data to be shared and modified by a group of persons. Key features required are that the DBMS must be able to support careful control in access and modification of data, and also be able to manage multiple versions and copies of the data.

Workgroup software, or "groupware", is a cross between database software and a network application package. Typically, the database, the repository, and the database software can be kept centrally in a server computer with multiple clients accessing and manipulating the databases. The client side of the workgroup DBMS generally provides a graphical user interface to allow users to define data definitions by using many point and click options.

Client-Server Distributed DBMS

The client-server architecture is now commonly available in DBMS products. These DBMSs not only have all of the above basic features, but also have additional features to support distributed DBMSs. Examples of distributed features include location transparency and replication transparency. Location transparency means that there exist multiple database servers, but the user, when formulating a request, does not have to know where the database is located that will be used to satisfy the request. Replication transparency means that there exist multiple copies of the same data at multiple locations, and the DBMS software has mechanisms for ensuring that all of the multiple copies of data are consistent at the same point in time.

DBMS with Family of Adjunct Packages

Most of today's DBMS vendors offer not only basic data definition and manipulation capabilities, but also a family of compatible software tools that provide additional functionality. Examples of these families of adjunct packages include fourth generation languages for building forms and report writers, computer-aided software engineering (CASE) tools for application development with the DBMS, a data dictionary or information resource dictionary system for maintenance of meta data (data definitions), and database administration tools for performance tuning. In addition, some DBMSs allow the user to create links with the software of other vendors, such as spreadsheets or mathematical and statistical packages. In the case of multimedia stations, DBMSs may provide the underlying persistent, object-based engine for the storage and retrieval of varieties of data objects including pictures, graphs, bitmap images, voice and video.

3.7 Digital Signature Technology

One of the difficulties encountered in converting paper documents to electronic documents is the search for a viable replacement for handwritten signature. In a paper-based office environment, signatures are generated in every step of the work flow. To convert this to an automated, paperless environment, a suitable replacement for the handwritten signature is critical.

Such a replacement signature must be as legally binding as the handwritten signature. The digital signature technology provides an answer to this search.

Digital signature technology has become a reality due to developments in public key cryptography. An overview of cryptography is presented in the following section so that readers may have a better understanding of what a digital signature is, and how it actually works. Briefly, a digital signature can be considered as a computer checksum generated on an electronic message by using a specific signature generation algorithm. The signature is generated using the signer's private key which is known only to the signer. The digital signature can be attached to a message before the message is sent electronically, and the signature can be verified by the message recipient using the signer's public key.

This technology is important for the design of the ERN because without proper protection, electronic documents can be easily tampered with by unauthorized persons. The digital signature provides both a signing capability and a mechanism for checking the integrity of the signed message. Since an ERN may contain valuable information or significant scientific discoveries, proper access control of each individual ERN is extremely important. Moreover, the possibility exists that some entries in an ERN may later be used to support patent claims. Therefore, the integrity of the ERN contents must be protected. A brief overview of how digital signatures work is thus provided.

3.7.1 Overview of Secret-Key and Public-Key Cryptography

The prevalence of computers and computer networks has greatly influenced the way people work. Unfortunately, the lack of properly protected computer systems and networks has also contributed to a notable increase in computer-related crimes. Cryptography has been used for centuries to protect sensitive or secret information from viewing by unauthorized personnel while the information was delivered via unsecured channels.

When a message is encrypted, it is transformed into a form that is unreadable by anyone without a secret decryption key. Encryption is the only way known today that can protect the privacy of information traveling on unsecured networks. Cryptography may also be used to protect the integrity of information by calculating a computer checksum based on the message to be sent. The checksum is sent along with the message to a recipient, who may recompute the checksum and verify that the message was not modified in transit.

Two branches of cryptosystems prevail today: the secret-key based cryptosystem and the public-key based cryptosystem. In a secret-key cryptosystem (also known as a conventional cryptosystem), a secret key is established and shared between two individuals or parties. The same key is used to encrypt and decrypt messages. If the two parties are in different physical locations, they must trust a courier, or some transmission system to establish the secret key without disclosing or modifying its value. The generation, transmission, and storage of keys is called key management. Ensuring that key storage, exchange of new keys and destruction of old

keys are performed securely often creates complex key management requirements for secret-key cryptography [FAHN92].

In public-key cryptography [NECH91], a user makes use of a pair of keys: a public key and a private key. The two keys are uniquely related so that the public key of a user can be made public without revealing any information about the private key. The private key of a user is known only by its owner. Because the value of the private key is not shared, public-key cryptography is often considered to be more secure than secret-key cryptography. It is the specific feature that a private key is never shared with another party that permits the unique signing capability. Public-key cryptography is used more for signing than encryption. Few public-key algorithms provide an encryption capability. In the algorithm that does provide the encryption capability, no single key is used for both encryption and decryption. Thus, public-key cryptography is also known as asymmetric cryptography.

One of the major advantages of public-key cryptography is the digital signature capability. A user may sign a message using a private key while others may verify the signature with the corresponding public key. Since the private key never leaves the possession of its owner, nobody else can forge the signature unless the signer's private key is compromised. Due to this feature, public-key cryptography also provides a "non-repudiation" property which is not provided by secret-key cryptography. Non-repudiation is the ability of a third party to authenticate the source and content of a message.

Though public-key cryptography is more secure than secret-key cryptography in the sense that the private key is not shared, one disadvantage of the public-key cryptosystem is that it is not as fast as the secret-key cryptosystem. To compensate for this, a one-way hash function is typically applied to the message to be signed. The hashing produces a condensed version of the message called the *message digest*, which is then signed using the signer's private key.

The hash function takes every bit in the input message and hashes it to a bit string of fixed size. The message digest can be thought of as a "digital fingerprint" of the message from which it was computed. If a single bit in the message is altered, the message digest will be different, hence, a signature generated from the original message will not verify. Therefore, a digital signature based on public-key cryptography can be used to check and protect the integrity of the message signed. In summary, a digital signature based on public-key cryptography possesses the following attributes:

- . Unique to the signer
- . Signer's private key is under the signer's sole control
- . Signature is linked to the data being signed
- . Signature is verifiable
- . Provides non-repudiation property

3.7.2 Implementation of Digital Signature Algorithm

In 1991 NIST proposed a draft Digital Signature Standard (DSS) [DSS91], which is expected to be approved by the Secretary of Commerce. In addition, NIST also proposed a Secure Hash Algorithm (SHA) to be used along with the DSS. The Secure Hash Standard (SHS) has been adopted as a Federal Information Processing Standard [FIPS180]. The Security Technology Group at NIST has successfully implemented the Digital Signature Algorithm (DSA) in an in-house designed smart card.

Though the digital signature appears to be a promising replacement for a handwritten signature and was sanctioned in writing by GAO for certain government use [GAO91], there are specific technical concerns and issues that need to be resolved before the technology can be widely implemented. One of the issues relates to the impact of software version change on human readability and signature verification when numerous word processors are in use and software is frequently updated. How can one ensure that a signature generated on a document three years ago can still be verified five years from now? Furthermore, would the correct version of software be around to permit a user to read the document as it was signed? This raises a question on whether a document should be re-signed every time the software used to generate the document is upgraded.

Another issue is the generation of digital signatures on linked and embedded objects. At the present time, it is not clear how this should be done, nor is it clear that documents with code to link objects and with embedded objects are compatible with digital signature technology. These are areas that require further research.

3.8 Time-Stamping Technology

The provision of a reliable time-stamping service is particularly important for the adoption of the ERN. This service is necessary in order to prove that a certain research event did occur on or before a particular date, which is often a critical issue in patent disputes. The time-stamping service not only provides the correct date and time, but also certifies, by digitally signing the time-stamp, that the specific document has been received and time-stamped by the service. The resulting time-stamp is affixed to the message sent to the time-stamping server. At present, the United States Government has several time services which only provide correct date and time, none of these services provides time-stamping function. One of these time services is at the U.S. Naval Observatory and another is at NIST.

The NIST time service can be accessed by using Research Material 8101, Automated Computer Time Service (ACTS). ACTS consists of software that can track the differences between a computer's clock and the NIST clock, and synchronize the computer's clock with the NIST clock. This software can be used periodically to correct the computer clock's natural drift. The U.S. Naval Observatory provides a similar time service, although resetting the computer's clock requires using commercial or custom-made software.

NIST's Time and Frequency Division has submitted a proposal for establishing a time-stamping service at NIST which would integrate the ACTS time service, electronic mail, the Secure Hash Standard (SHS), and the draft Digital Signature Standard (DSS). Basically, the user would take the data to be signed, create a message digest using the algorithm specified in the SHS, and send this message digest to the NIST time server via electronic mail or telephone. The NIST server would then affix a time-stamp to the message digest and use the DSS to sign the combined data. This time-stamped document could then be transmitted back to the user. The signature on the time-stamp can be verified by using the time server's public key.

The proposed time-stamping service assumes the availability of a trusted time-stamp server. However, some people may prefer not to place so much trust in such a service. Haber and Stornetta of Bellcore have designed two time-stamping schemes which do not require a trusted server [HABE91]. Bayer, Haber, and Stornetta later published another paper to address ways to improve the efficiency and reliability of such time-stamping methods [BAYE93]. Bellcore has been developing a time-stamping service based on this work.

4. FUNCTIONAL SPECIFICATIONS OF THE ERN

This section presents the functional specifications of an ERN that can meet basic ERN requirements as discussed in Section 2.3.1. The functionalities are specified regardless of the maturity and availability of the enabling technologies. However, current technological limitations or restraints will be noted. The discussion starts with the system configuration that is considered suitable for the NIST environment. Due to the concerns over cost, it is recommended that NIST scientists use existing resources as much as possible, adding only the pieces of hardware/software that are absolutely necessary for their work. Consequently, there is no specific requirement for a particular platform. It should be noted that while functional specifications of the ERN are being addressed, there is still much research to be performed. Currently we do not have an implemented ERN application for the scientists to evaluate. It will become apparent in Section 4.2.3 that the development of an ERN with all the necessary security features is a very complicated task.

4.1 System Configuration

An ERN can be implemented as a software application which can run on a variety of platforms, much like a word processor. However, if an official time-stamp is desirable, an ERN will need to interact with an independent server in order to receive the official time-stamp. In this case, the computer hosting the ERN should be connected to a network, or should at least have a modem. To avoid unnecessary confusion, an ERN and a time-stamping service will be considered as two separate applications in the following discussion.

An ERN should be portable to different platforms so that one may install the ERN on whatever hardware is accessible. The survey revealed that most NIST scientists have their own desktop computers. The most common computers are PCs, Macintoshes (Macs), and workstations running a UNIX operating system. However, if there is a need for the ERN to be carried around like a paper-based notebook, then a light-weight notebook or subnotebook computer may be a better choice as the host for the ERN. This portability must be weighed against the negative factor of limited connection capabilities to a network or other devices.

In any case, an operating system that supports user-friendly Graphical User Interface (GUI) is essential for the ERN. The Macintosh and the PC, whether a desktop or a light-weight portable model, running in an environment with windows offer a compelling combination of affordability and software availability. If a multimedia capability is also a requirement for the ERN, these computers are the most desirable for the development and delivery of multimedia capabilities. This will be the model configuration, but not the required configuration, for the NIST ERN as illustrated in Figure 1. As an option, a scanner with associated OCR software may be added to this configuration to provide an alternative for making ERN entries in cases where software tools for ERN entries cannot adequately meet the users' needs.

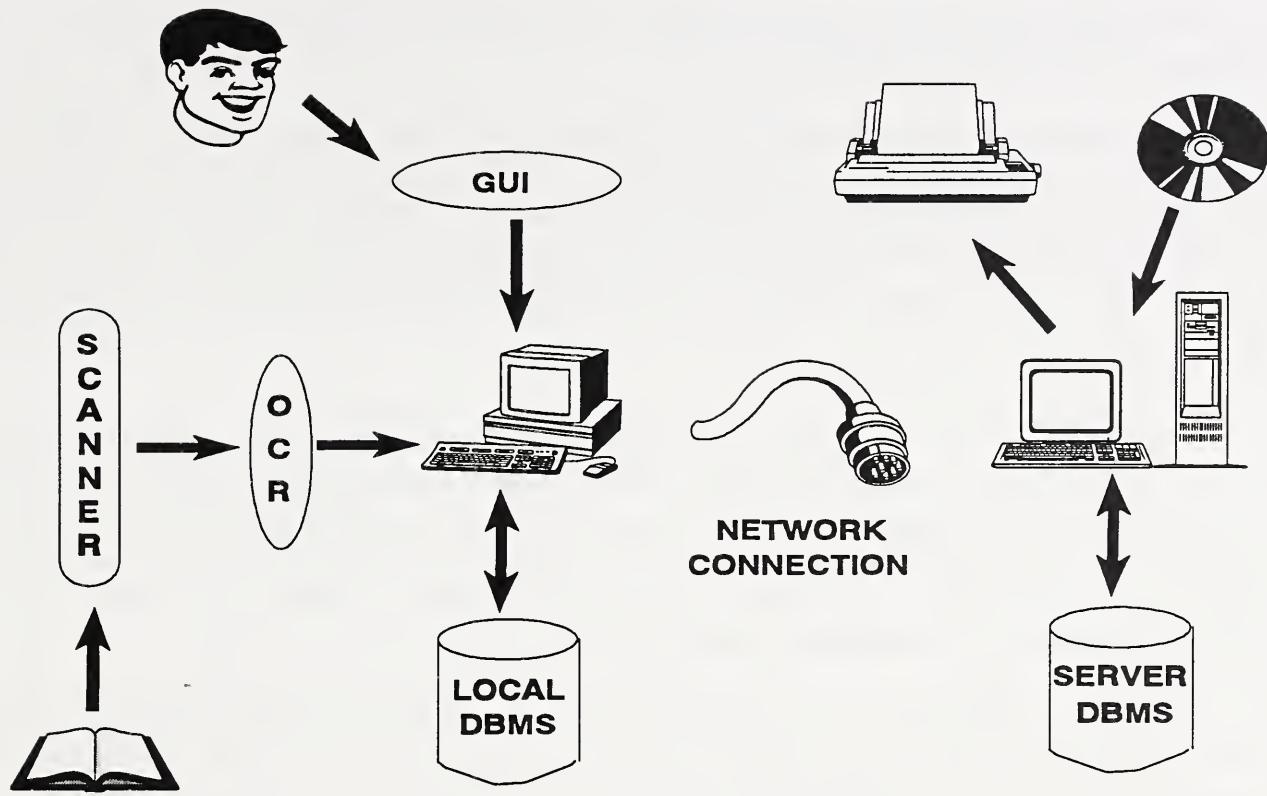


Figure 1. A System Configuration for the NIST ERN

4.2 Functional Specifications

This section presents the functional specifications of the basic ERN as described in Section 2.3.1. The specifications are provided in three subsections starting with the ERN input capability, followed by storage and retrieval considerations, and concluding with a discussion on security aspects of the ERN.

4.2.1 ERN Input Capability

The ERN should utilize a user-friendly interface. The acceptance of the ERN depends to a great extent on the tools that are provided for users to make notebook entries that include text, graphics, images, equations, etc. Not only must the appropriate tools be provided and easy to use, but they need to be integrated as a package. Due to the multiple input possibilities and different forms of storage, this integration of tools will not be a trivial task.

Since word processors are so common, textual input is easily supported. However, to sketch a diagram, a drawing tool with all the built-in functions that can be expected from a

commercial drawing package must be available. A capability to link a particular drawing to the associated text must also be provided.

Many scientists write equations in their paper notebooks. To simulate this function, a mathematical tool should be included in the ERN. The tool must be easy to use so that entering an equation does not take longer than writing the equation by hand. If such a tool is not currently available, an alternative is to write the equation on a piece of paper and scan the equation into the ERN using a scanner. The drawback of this, however, is that the scanned equation can only be stored as an image and cannot be edited as text due to the limitations of cursive handwriting recognition.

As indicated, the ERN should also permit the entry of imagery, whether from an image file or directly from a scanned image. The designer of the ERN should consider the compatibility issue and make sure the ERN can accept the type of image format that is used most frequently. Once images are incorporated into the ERN, the ERN should provide an image-editing capability. It is yet to be determined how well current commercially available image processing systems can be integrated into the ERN.

As multimedia PCs are becoming more and more popular, it may be desirable for the ERN to allow voice, video, or pen input and output. An ERN may also be designed to have the capability of retrieving reference information from a CD-ROM or a database and link the retrieved information to a notebook entry. The complexity of including multimedia capabilities in the ERN has not yet been explored since the enabling technologies are still evolving.

As indicated earlier, the integration of these input tools may not be an easy task. Some word processors already provide text and graphics tools that can easily be integrated. Equations can also be entered as text. In addition, some OCR software has built-in functionality to convert scanned imagery into specific formats that are acceptable by popular word processors. By combining the use of a powerful word processor and a scanner with OCR software, a plausible short-term solution for NIST scientists is provided.

To make a workable ERN, more functionalities need to be built into the package. These functionalities will be discussed in the following subsections. Three products recently entering the market seem to provide some of the ERN capabilities that are needed, though complete details are still lacking. Generally speaking, these products are electronic notebook systems which use a central database to manage notebooks. Users can work on various platforms and can include many different types of objects on the same page of the notebook. The system keeps a record of changes made to those objects and to the notebooks as a whole. Developments in this area will be closely watched, as the long-term solution for the ERN is being researched.

4.2.2 Storage and Retrieval of ERN Contents

The storage and retrieval of ERN contents is a complex topic that requires clarification. The storage of ERN contents may refer to the physical location *where* the ERN files reside, or storage may refer to the logical aspect of *how* the various ERN objects are stored. The latter is the focus of this section, namely, to consider how the objects are integrated and linked, and how to build an indexing system to permit fast and easy object retrieval. For physical storage, it is assumed that the ERN contents are stored either as files or in a database, and can reside locally on a desktop, or remotely on a designated server. For security considerations, those files or the database must be backed up periodically, and for certain information it may even be wise to keep a hard copy.

An ERN can be designed to be a simple application, or can be implemented with a powerful data tracking capability. At a minimum, an ERN should have the ability to handle textual and graphical input, capabilities which are already available in many word processing packages. Multimedia input can then be added, and each input data type can be an *object* in the ERN and stored in the proper format specific to that object. The storage of multimedia objects is an area that needs further research. However, a mechanism must be provided by the ERN application to link all these objects together and still allow the retrieval of a specific object if desired. Therefore, control information on these objects must be collected and maintained by the ERN application. Information such as the following should be considered for retention as control information:

- the type of information/object
- the creator of the object
- the date and time the object was created
- the description of the information
- security control of the object
- associated keyword or caption
- list of association or linkage information
- modifications of the object
- file name if applicable
- author's signature if applicable
- time-stamp if applicable
- version number if applicable

It is possible to create an ERN application using only these tracking functionalities. Under such an ERN application, it would be up to the user to remember what file name was given to a particular project or notebook entry for later retrieval. On the other hand, a more powerful ERN program can be built which would not only provide initial entry functionality for the user, but would also provide various tracking capabilities to assist the user in retrieving and organizing information in the ERN. An ERN application can log all notebooks maintained by a single user or by all users, can track the user's projects, or track the equipment or resources used for a particular project. To provide these tracking capabilities, the ERN is essentially building indices and adding database management functions to its basic functions.

Building indices is essential for context retrieval. Information that is crucial for searching should be included in the indexing system. For example, if an ERN is to have a project tracking capability, then the project title, project description, research approach, and project conclusion can all use specific keywords to permit a fast and easy search and retrieval. As indices can be built on the tracking or logging information, so too can they be built on the ERN contents. For example, a textual notebook entry may use a few keywords for searching, or a full-text searching capability may be provided by including every word, except for noise words, in the indexing system. A full-text indexing system would consume a tremendous amount of storage just for the indices. Therefore, a good design of the ERN should consider the needs of its users, and provide a capability for the users to customize functionalities as well as tracking capabilities in order to create an ERN that is best suited for them.

4.2.3 Time-stamping Service

As indicated earlier, a creditable time-stamping service from an independent source should be made available to ERN users. A time-stamp signed by an independent server and attached to an ERN entry may provide strong evidence of when the entry was signed. At the current time, legal questions concerning the validity of such time-stamps will have to be resolved by the legal experts. Since a great amount of trust is placed in the time-stamping server, physical security for the server should be provided and other vulnerabilities must be assessed and corrected, if any exists.

To protect the integrity of ERN contents, notebook entries, once finalized, should be signed by the author and sent to be time-stamped. This requires that the computer on which the ERN program resides be connected to a network, whether it is a Local Area Network (LAN) or a Wide Area Network (WAN).

Digital signature technology is currently available, but the infrastructure to support public key certificate management is not yet in place. Key management has always been a complex issue for any type of cryptosystem. To provide a digital signature capability for the ERN users, detailed key management issues need to be addressed and resolved. Included are issues such as the key generation, the determination of a suitable cryptographic module to be used for various cryptographic functions, the initialization of the cryptographic module, the generation and distribution of public key certificates, and the entire arena of certificate management. Therefore, even though the technology exists, the implementation of this technology is a non-trivial task.

The issues mentioned above also apply to the provision of a time-stamping service, since a time-stamping server uses the digital signature technology to generate the time-stamp. The issues raised here are not meant to discourage potential ERN implementors. Rather, the point is to forewarn them not to overlook some crucial security issues, especially since designing a secure system is never easy and needs to be done as an integral part of the initial system design.

Once the key management issues are resolved, an ERN user can use his/her private key to sign a notebook entry, possibly using a NIST-developed smart card for key storage and

signature generation. A smart card is a credit card size cryptographic module which not only provides protection for the information stored on the card, but also provides the implementations of cryptographic functions. A signed notebook entry can be sent to the time-stamping server over the network. The server stamps the ERN entry with the correct date and time, and signs the time-stamped entry with the server's private key. If a need arises to prove that an entry containing certain information was made at a specific time, the signature on the time-stamp can be verified by anyone who has access to the time-stamp server's public key. If even a single bit of the information on the notebook entry or the time-stamp was modified, the signature will not verify.

5. FUTURE WORK

Based upon the assessment of the ERN requirements and the enabling technologies, functional specifications for a basic ERN are presented in Section 4. During the survey, many NIST scientists expressed interest in maintaining research notebooks on the computer. Some have actually done that. However, in order for the ERN to be widely accepted and used regularly, more functions would need to be built into the basic system. Many of the current technologies and commercial products that were reviewed seem promising for the design of the ERN. These range from the specific hardware that can be used to host the ERN to various software tools that can be integrated to provide useful ERN functions. However, none of these products currently present an integrated solution for immediate use. Specific applications have yet to be developed.

To use these state-of-the-art technologies, many technical issues would need to be resolved since many of these technologies are still new and developing. The utilization of any one of these in a single project, let alone using all of them, will raise many design issues. Obviously there is still much research to be done.

Additional research and a prototype implementation are necessary. This includes the investigation of full multimedia support for ERN entries, the linking and signing of different types of objects, the logical storage of ERN contents, and the indexing scheme to be used. Once these issues are resolved, the current ERN functional specifications will be augmented, and the implementation of a prototype ERN will be attempted accordingly. The prototype ERN, once implemented, will be demonstrated to, and tested by, participating research scientists for possible improvement and enhancement.

APPENDIX A - NIST RESEARCH NOTEBOOKS POLICY

July 29, 1992

All NIST employees engaged in research and development activities are responsible for maintaining a thorough and accurate record of their work by keeping a research notebook. Employees conducting research using electronic media are responsible for maintaining a notebook that chronologically documents the progress of their research and indexes electronic work files so that primary research results may be retrieved.

Research notebooks are to be hard covered or spiral bound with pre-printed sequentially numbered pages, such as the types available from the NIST storeroom. Research records must be permanent, contemporaneous with the research, accurate, and reasonably protected from compromise. Record-keeping procedures discussed in Subchapter 2.10.06 of the NIST Administrative Manual should be followed when intellectual property issues, such as patents, copyrights, etc., are likely to be important since these records establish the legal foundation for future claims.

All NIST laboratory and technical unit managers are responsible for insuring that the technical activities of their staff are fully documented, that appropriate control measures are in place so that either paper or electronic records of research results are retrievable, and that all staff engaged in research are properly instructed. NIST laboratories that conduct research predominately using electronic media may be allowed to implement procedures that accomplish the same goals but use alternatives to bound notebooks. Prior approval for such exceptions must be obtained from NIST Director.

All technical records, including research notebooks, journals, electronic records, data, calculations, etc. pertaining to NIST activities are official files of the U.S. government and, as such, are the property of the Government, not the employee. Laboratory and technical unit directors are responsible for insuring that these records are not removed from NIST without proper authority even when an employee transfers, retires, or otherwise separates from NIST.

APPENDIX B - ERN QUESTIONNAIRE

I. Questions for all (scientists and managers)

A. Current Notebooks

-Issuance

- . Do you currently use a notebook to keep records of your research? Explain why or why not. If you do, what form does your notebook take (e.g. paper-based, computer-based, etc.) ?
- . Are notebooks automatically provided to you, or must you request them from supply? Do you fill out an issuance form which binds you to that particular notebook?
- . Upon starting employment at NIST, did you receive specific instructions for the maintenance of a laboratory notebook? From whom? Are these requirements NIST-wide, or particular to your organization?

-If you have organization-specific requirements, describe them briefly?

- . How closely do you adhere to the rules for maintaining your notebook? What regulations don't you follow?
- . What reason for keeping a notebook is most important to you:
 - a) recording data for easy future reference
 - b) recognition from peers in scientific community
 - c) potential support for patent claims
 - d) other (specify) _____

-Nature of notebook information

- . What form do your notebook entries take (e.g. text, numerical data, diagrams, references to other materials, attached printouts, etc.) ? Estimate what percentage of your notebook is in each format.
- . What type of attachments do you place in your notebook (e.g. photographs, instrument printouts, etc.) ?
- . What kind of line rule do you use in your notebook (e.g. horizontal, square grid, columnar) ? What kind is required for your work?
- . Do you make an effort to write legibly enough for others to read and understand your work?

- . What methods do you use to reference particular data/entries?

-Maintenance of the notebook

- . How often do you use your notebook?
- . Do you sign and date your entries? If so, how often?
- . Do you have witnesses sign and date each page of your notebook? How many witnesses? Who are they? Do they really read and understand your work, or do they just sign the page?
- . Does your supervisor review your notebook on a regular basis? If so, how often and how thorough are the reviews?
- . How many other people read your notebook each week? How do they obtain it? How concerned would you be if your notebook was disclosed to others without your knowledge?
- . How concerned would you be if you found out that information in your notebook had been altered?
- . How/Where do you keep your notebook? Do you keep it locked up? Is it easily accessible to others?
- . How often do you use your notebook *outside of your lab at NIST*? In what type of environment? Do you make entries or just review the notebook?
- . How long do you normally use an individual notebook? How many do you use while working on a project? (How many pages are in these notebooks?)
- . What do you do with completed notebooks? Are they archived? By whom?
- . What other requirements would you suggest which could benefit you and NIST?

-Other comments:

B. Computer familiarity

- . Do you currently make frequent use of a computer for your lab work? If so, what kind is it? Is it attached to the NIST Local Area Network?
- . Would you feel comfortable using a computer-based electronic notebook?

. Are you comfortable working in a Windows environment on the computer? If not, please explain why.

C. ERN

. What type of computer would be most convenient for your use of the ERN (e.g. desktop, laptop, hand-held) ?

. Would you like to have the capability of making handwritten entries into a computer as well as typed entries? Estimate what percentage of the time you would use each.

. Would the ability to choose what type of line rule is on a particular "page" in the ERN be convenient to you?

. What kind of electronic "attachments" (e.g. video, audio, instrument results) would you like to be able to place in the ERN? Which ones are required for your work?

. What cross-referencing capabilities would you like in the ERN which you do not have now?

. Is there any special requirement for the ERN for the type of work you are doing (e.g. Any special-format notebook that you use regularly?) ? Why do you need this format?

. What other features would you like the ERN to have? Please be sure to list features of the ERN which are needed to meet your work requirements, and which have not been addressed in previous questions.

. What security concerns do you have with respect to the ERN (possible issues: data integrity, accessibility(by yourself and others), loss of notebook)?

II. Questions for managers

. How do you enforce the NIST policy of requiring your staff to keep research notebooks?

. Does your staff adhere to the procedures for the distribution, collection, and control of notebooks?

. Do you require your staff to use a specific format for their notebooks (paper or electronic)? Do you require that they keep records of a specific nature?

. Are these notebooks reviewed periodically? If yes, by whom? How important is the role of the notebook during performance review time?

- . What kind of cost range will be acceptable for your organization to use ERNs on a regular basis?
- . What features must the ERN have, at a minimum, to meet your needs?
- . What *extra* features and capabilities would be nice to have on the ERN?
- . What concerns would you have about the security of ERNs in your organization?
- . What are the potential benefits of the ERN to your organization? Do you feel that implementation of the ERN is favorable to the system of research notebook maintenance which is now in use?
- . Suggestions:

APPENDIX C - INTERVIEWED SCIENTISTS

| <u>DATE</u> | <u>NAME</u> | <u>NIST LABORATORY</u> |
|-------------|--|--|
| 2/22/93 | Craig Sansonetti Joseph Reader | Physics Laboratory Physics Laboratory |
| 2/25/93 | Edward Kelley Bruce Field | Electronics and Electrical Engineering Laboratory Electronics and Electrical Engineering Laboratory |
| 2/25/93 | Thomas Larason | Physics Laboratory |
| 3/2/93 | Susannah Schiller James Blue | Computing and Applied Mathematics Laboratory Computing and Applied Mathematics Laboratory |
| 3/4/93 | Robert Saunders James Walker | Physics Laboratory Physics Laboratory |
| 3/5/93 | Robert Myklebust | Chemical Science and Technology Laboratory |
| 3/9/93 | Joseph Kinard | Electronics and Electrical Engineering Laboratory |
| 3/16/93 | Eric Steel Cynthia Zeissler Ryna Marinenko | Chemical Science and Technology Laboratory Chemical Science and Technology Laboratory Chemical Science and Technology Laboratory |
| 3/17/93 | Richard Lindstrom | Chemical Science and Technology Laboratory |
| 3/19/93 | Richard H. Harris Richard Peacock | Building and Fire Research Laboratory Building and Fire Research Laboratory |
| 3/26/93 | Donald Blomquist | Manufacturing Engineering Laboratory |
| 4/20/93 | Michael Gaitan | Electronics and Electrical Engineering Laboratory |

APPENDIX D - PRODUCTS REVIEW

Many off-the-shelf products including hardware devices and software tools were reviewed for potential use in an ERN implementation. Reviewed products are summarized using the following format:

1. A brief description of the product;
2. An assessment of how the product can provide the basic ERN features, and allow for the addition of enhanced features;
3. An identification of the software/hardware necessary to support the product;
4. The approximate cost.

It is impossible to survey and evaluate all the products in the marketplace due to the sheer number of available products and the pace at which new products are being introduced. It should also be noted that mention of vendor products is for descriptive purpose only and does not constitute a NIST endorsement.

Current Electronic Notebook Products

ResearchStation (from Helix Systems, Inc.)

1. This is a software environment which allows scientists to integrate currently-used applications and tools, existing computer systems, multimedia technology, and database systems. ResearchStation may be used to acquire, organize, search, and share information in various electronic ways in project notebooks or personal laboratory notebooks. Multiple data types can be manipulated on the same page by users working at two separate locations. Versions of notebooks and notebook objects can be tracked, and several visual browsing mechanisms are provided. An intelligent assistant can be used to monitor information changes and flows, and can automate certain tasks done by the scientist. ResearchStation can also be used to retrieve information from on-line catalogs and other services.
2. This is a very comprehensive software environment which appears to meet many of the basic requirements mentioned in Chapter 2 of this paper. It has not been thoroughly tested yet, but a demonstration of this product revealed that it has great potential for use in the development of a prototype ERN system.
3. ResearchStation will run in MS Windows 3.1, MS WindowsNT, and Macintosh environments. It may operate as a stand-alone system, or it may be connected to a network in a client-server configuration. Additional hardware and software

(scanner, OCR software, sound board, speakers, microphone) could be purchased to take advantage of scanning and audio capabilities offered with the software.

4. No pricing is available yet - this product is expected to be commercially available in 1994.

The Virtual Notebook System (from The ForeFront Group, Inc.)

1. The VNS is a software product which allows scientists to perform collaborative research. A VNS Applications Programming Interface (VAPI) can be used to integrate the VNS with other applications, such as digital signature software. Pages may be of varying size (they can even be scrolled), and may contain any number of information objects. Objects may be retrieved from another file, cut and pasted from another window, or created directly on the page, depending on the object type. Other features include an automatic table of contents, navigational links to other parts of the notebook, action links to external applications, support for audio and video objects, and a browser.
2. This system appears to provide many of the basic capabilities mentioned in Section 2. The system has not been tested as of the moment, but the Virtual Notebook System might help the development of a prototype ERN.
3. The VNS Server (Real-time Object Manager: RTOM) can be implemented on a computer with any of the following operating systems: Solaris, AIX, HP-UX, Ultrix, OSF/1, SGI, and Unixware/System V Release 4. The VNS Client can run on X Windows, Microsoft Windows, and in a Macintosh environment. Separate platforms of the VNS can communicate using TCP/IP, and a WindowsNT version is being developed. A VAPI should be used to integrate the VNS with other applications. A relational database system which is SQL-compatible may be used. Additional hardware and software may be purchased to take advantage of multimedia capabilities offered with the VNS.
4. MS Windows and Apple Macintosh clients are listed at \$595 each, while a UNIX client lists for \$795. The UNIX server portion of the VNS is priced at \$6000 for an unlimited number of users. Annual software maintenance costs run approximately \$1000 for each VNS server, and \$100-\$150 for each client. VNS API Programmer's Toolkits for clients cost between \$3750 and \$5000, with annual software maintenance costs between \$675 and \$900. Finally, VNS API Runtime Licenses are available for each client for under \$100.

FlowStream (from Consilium)

1. FlowStream was developed mainly for the pharmaceutical and related chemical industries, thus the compliance to regulations is much emphasized. FlowStream specifically addresses the needs of the chemical research and development facility. The researcher can define and document experiments and procedures, statistically compare results from different experiments, and automatically record notebook changes. FlowStream also maintains a detailed record of laboratory equipment, including its use and servicing history. A Document Vault permanently stores documents and provides a complete audit trail of their life cycle.
2. FlowStream appears to provide some of the basic requirements mentioned in Chapter 2 of this paper. It has not been tested as of the moment, but it might be helpful in the development of a prototype ERN system. One concern about using this product is that its format might be too industry-specific in the area of pharmaceutical, but this will have to be determined at a later time.
3. The Document Vault package is based on a client-server architecture, and requires either Microsoft Windows, XWindows, or a Macintosh platform, as well as network software and hardware. An SQL-compatible relational database could also be used. Tools from the Digital Equipment Corporation should allow the user to connect FlowStream to existing equipment, such as data acquisition devices and laboratory information systems.
4. To obtain *all* modules of FlowStream for twelve users, the estimated cost ranges between \$175,000 and \$200,000 (it is not clear if this figure includes the price of hardware).

Digitizer Tablets

Wacom SD-510C

1. A small digitizer (under one square foot and 5 lbs.) that is compatible with PCs, Macintoshes, and Sun, NeXT, and HP workstations. It comes with a cordless, pressure-sensitive pen.
2. This digitizer has a small footprint and would be a simple addition to the scientist's system. It might be awkward to use for writing or sketching since the user must write on the tablet while watching the monitor. Writing on the tablet does not feel as natural as writing on paper.
3. A pen operating system and pen-aware applications are required.

4. The cost of this digitizer tablet is under \$600.

CalComp NotePad (Model 3311)

1. This is another small digitizer (under one square foot) which weighs less than two pounds. It is designed for use with PCs, with an RS-232 interface and drivers for Microsoft Mouse and Windows.
2. In addition to its small size and possible awkwardness of use, it is unclear whether this digitizer can be connected to non-PC compatible computers.
3. Either Microsoft Windows for Pen Computing, CIC PenDOS, or Go Corporation PenPoint is required to allow applications to understand pen input.
4. The approximate cost is \$200.

Tablet Computers

NCR 3130

1. This is a light tablet (< 4.5 lbs) which has a last-generation processor (386SL), 60 MB hard disk, and additional memory which can be provided by a PCMCIA card. Four major pen environments are supported including the Microsoft Windows for Pen Computing, PenDOS, PenPoint, and PenRight!
2. The NCR 3130 is a rugged computer with a shock-mounted hard disk, light weight, 4 hours battery life, and PCMCIA capability, which make this tablet a good candidate for use in laboratory and field environments. However, the processor, clock (20MHz), and black-and-white screen are drawbacks for using the device very frequently for ERN work. The NCR 3130 system probably will not be able to support enhanced ERN features such as multimedia.
3. An optional keyboard and external video are recommended. PCMCIA memory, Ethernet/LAN, SCSI, and modem cards should also be considered for use.
4. The approximate cost is \$3500.

NEC UltraLite Autograph

1. The Autograph has a powerful 486SL processor, although its clock speed is only 20MHz. It also comes with a relatively large (80MB) hard disk drive, a pair of

PCMCIA Type II slots, and a choice of pre-installed Windows for Pen Computing or PenPoint. Next to the black-and-white screen are several HotZone buttons which allow for one-touch functionality.

2. Multiple PCMCIA slots would allow the ERN scientist to use a memory card and ethernet/LAN card simultaneously. Support of multimedia is questionable, considering the tablet's display and limited storage.
3. An optional keyboard and external video monitor are recommended. PCMCIA memory, Ethernet/LAN, SCSI, and modem cards should be considered for use.
4. No pricing information is available yet.

GRiD PalmPAD

1. PalmPAD is a wrist-held, 3 lb. computer with a very small screen and configurable communication module (either a short-range radio or data/fax). It holds a PCMCIA storage card, and the PenRight! environment is also provided.
2. The PalmPAD's size, weight, and radio link (up to 800 feet) are ideal for the lab and field environments. However, the small storage capacity (20MB) and 6.5" black-and-white screen make it unsuitable for implementing the entire ERN. Its orientation towards vertical, data collection applications does not allow for much variability - a factor which might detract from a scientist's willingness to use the PalmPAD.
3. An optional keyboard is recommended, and software and hardware for downloading data to a desktop computer are needed to use the radio capability.
4. The approximate cost is \$3395.

IBM ThinkPad 710T

1. Of all the tablet computers reviewed in this section, the ThinkPad has the most powerful processor (486SLC/25 MHz). It weighs about 6 lbs. The only available pen environments on the ThinkPad are PenDOS and PenPoint. Two different configurations are available: one PCMCIA Type II slot with a 60MB hard drive, or three PCMCIA Type II slots with an open bay.
2. The ThinkPad's powerful processor lends itself to relatively fast character recognition, but the limited storage space and monochrome LCD once again raise doubt about the possible incorporation of multimedia capabilities.

3. An optional keyboard and external video monitor are recommended. PCMCIA memory, Ethernet/LAN, SCSI, and modem cards should be considered.
4. The list price is \$3599.

Pentop Notebook Computers

GRiD Convertible / AST PenExec

1. This notebook computer combines the keyboard and digitizer in one small package. It weighs about 5.5 lbs. The Convertible/PenExec may be used as a regular notebook computer, but the screen can slide down over the keyboard for use as a tablet computer. It comes with a math coprocessor, one PCMCIA Type II slot, Windows for Pen Computing, and a 120MB hard disk drive.
2. The combination of keyboard and digitizer would allow the ERN user to easily switch between textual and sketching input. The notebook has a magnesium casing which is very rugged, and this factor combined with the computer's size and weight would make it very useful in laboratory and field environments. Storage space is reasonably large and sufficient to hold several large applications and ERN data. However, multimedia use is still questionable due to the black-and-white screen, 386 processor, and limited number of ports (not as many peripherals can be attached to a notebook computer as compared to a desktop machine). Later generations with larger storage space, more PCMCIA slots, more powerful processors, and color displays might be better-suited for use as a mobile ERN.
3. Recommended options include an internal modem, PCMCIA Ethernet and SCSI cards, and an external video monitor.
4. The approximate cost is \$3000.

Personal Digital Assistants (PDAs)

Apple Newton MessagePad

1. The Newton MessagePad is a hand-held, pen-based computer which weighs less than a pound and accepts a single PCMCIA Type II card for storage purposes. This computer is designed as a personal information manager (PIM) on which a person can jot down notes, send fax, mark important information, use a calendar, and do similar tasks. An Intelligent Assistant is a feature which anticipates some of the user's actions. Also included is a printed handwriting recognizer unique to

Newton products, and a graphics recognizer which straightens lines and smoothes curves. Newton products can communicate with one another through an infrared link, and a tool was developed for the Newton to easily download data to a PC or Mac.

2. If the Newton MessagePad were used in an implementation of the ERN, it would have to be paired with a desktop personal computer or workstation to provide all of the ERN's basic requirements. For example, word processing does not exist on the Notepad, and storage space is very limited. The lack of a hard disk necessitates the use of a PCMCIA card for storing applications and data, so all information used must not exceed 20 MB (which is the current storage limit on PCMCIA cards). Despite the advantage provided by the Newton's physical portability, PDA technology is extremely new and untested, and time will be required to pinpoint and correct the PDA's shortcomings.
3. A desktop system is needed to provide many of the ERN's basic requirements, and special software might be required to download PDA data to this system. PCMCIA cards are needed for storage.
4. The cost ranges between \$600 and \$800 depending on the configuration desired.

Scanners

Fujitsu 3093

1. The Fujitsu 3093 is a high performance, high-speed top of the line scanner. The output resolution is user definable from 50-600 dots per inch.
2. Printed documents or hand-drawn diagrams may be scanned. The scanner needs to be coupled with optical character recognition software for the capture, storage, and retrieval of scanned text and images.
3. The scanner operates with Sun SPARC workstations and PCs running Microsoft Windows.
4. The approximate cost is \$9,000.

Hewlett-Packard ScanJet Plus IIP and IIC

1. The Hewlett-Packard ScanJet Plus IIP and IIC are graphics and text scanners. The IIP model is a gray scale scanner, and the IIC model is a color scanner. The output resolution of both models is 400 dots per inch.

2. The scanners are useful for scanning and inserting graphics, text and handwritten equations into the ERN.
3. These scanners operate with PCs running Microsoft Windows.
4. The IIP model costs \$1,700 and the IIC model costs \$2,700.

Logitech Hand-operated Scanner

1. The Logitech scanner is a hand-operated scanner. It is capable of scanning both gray scales and colors up to 400 dots per inch.
2. This scanner needs a steady hand to scan the document.
3. The scanner requires software in the attached computer and will work with Microsoft Windows 3.0 and a VGA graphics card with 8-bit color.
4. The approximate cost is \$699.

OCR Software

WordScan Plus by Calera

1. WordScan is optical character recognition software which works in conjunction with a scanner. WordScan is capable of scanning text only, image only, text and image, or template mode. A verifier is provided to allow a user to edit the scanned document. Depending upon the quality of the input from the scanner, the WordScan claims to be 99% accurate.
2. The WordScan, after a scanned document is OCRed and verified, is capable of storing the text and image files in the formats of many popular word-processors and image-processing tools for further editing and processing.
3. The WordScan software operates with compatible brands of scanners and PCs running Microsoft Windows.
4. The approximate cost is \$600.

OmniPage Professional by Caere Corporation

1. OmniPage will reformat the scanned data, catalog it, and store it online. The OmniPage system documentation claims that this product can create databases, but more information is needed on this feature.
2. This product has most of the capability of WordScan.
3. The OmniPage system operates on PCs running MS-DOS.
4. No pricing information is available.

Document Storage and Retrieval Software

ZyIMAGE by Calera Recognition Systems

1. ZyIMAGE is designed specifically to work with the OCR products from the Calera Recognition Systems. It saves scanned text in the ASCII format, and saves images in TIFF files. The text database automatically maintains links between text files and the associated images.
2. The fully integrated text and image viewing capability of this product is useful for retrieval of ERN contents. The user interface of ZyIMAGES is easy to use, thanks to plain text searches and context-sensitive HELP support.
3. The ZyIMAGE system operates on PCs (or compatible computers) running DOS 3.1 or Windows 3.1. It needs a VGA monitor, and a Windows-compatible Mouse or pointing device.
4. The approximate cost is \$900.

Excalibur by Excalibur Technologies Corporation

1. Excalibur/Electronic Filing Software manages both text and images from multiple sources and locations. It offers automatic full-content indexing and content-based retrieval. It also supports relational DBMSs such as ORACLE, Informix, and Sybase.
2. This software has many document retrieval features, but its capabilities have not been fully tested as of the moment.
3. Excalibur operates on UNIX platforms as well as PCs running Microsoft Windows.

4. The approximate cost of the system is \$35,000.

Database Management Systems

InfoSelect from Micro Logic Corporation

1. InfoSelect is a personal information manager (PIM) which allows users to organize their personal information such as phone numbers, client lists, plans and projects, etc. It is not a generalized DBMS but rather a program that uses stacks and windows - like pieces of paper on a desk. It offers a number of predefined forms so that users do not have to create fields and enter information in a particular way.
2. InfoSelect has already been used by some NIST scientists to organize their information, which could be extended for ERN use. It is claimed to be very easy to use. The drawbacks are that it is strictly a personal information management tool and not designed to support sharable databases.
3. InfoSelect operates on PCs and compatible computers running Microsoft Windows.
4. The approximate cost is \$150.

Lotus Notes by Lotus Development Corporation

1. Lotus Notes is a cross between database software and network application packages for workgroup communications. Lotus Notes provides an easy user interface to allow users to create data definitions with many point and click options. The system has a simple macro language which allows users to define views so that selective queries may be displayed with a customized reporting style. All Windows applications can be "cut and pasted" into Notes, including data in the form of pictures, graphs, and bitmap images (In the future voice and video will be included.). Lotus Notes also incorporates public-key signature verification features.
2. Notes can be used to organize public ERN files and report and display different views of lab notebook contents for retrieval, browsing or reading. The key feature is that multiple clients may access a central server computer in order to manipulate multiple databases. It is also simple to design a database and display data, but Notes probably cannot handle very large database problems. There may also be performance issues.

3. Notes servers must operate under either Microsoft Windows or OS/2, on a dedicated 386 or 486 computer. Clients can run under Windows, OS/2 or in a Macintosh environment. The network software includes Novell, 3Com, Banyan VINES, Microsoft LAN Manager, and Microsoft Windows for Workgroup. TCP/IP will be available in the fall of 1993, along with UNIX and WindowsNT servers and clients.
4. Each copy of the client license for Notes Release 3.0 costs \$495. The Starter Pack with a Windows server and two clients retails for \$995.

Relational Database Management Systems

1. Many relational DBMSs complete with SQL query language are available commercially including the following:
 - . RDB by Digital Equipment Corporation
 - . Database 2 by IBM
 - . INFORMIX-Online by Informix Software Inc.
 - . ORACLE by Oracle Corporation
 - . Model 204 by Praxis International Inc.
 - . ADABAS SQL Server by Software AG
 - . Sybase System 10 by Sybase, Inc.
 - . WHITE CROSS 9000 by White Cross Systems Ltd.

These products were all validated with the NIST SQL test suite as conformant to Database Language SQL validation procedures. All of the above products offer a family of products such as SQL-forms, dictionary and client-server networking software.

2. Relational DBMSs are loaded with database functionality. An ERN may not need all the features provided by this DBMS, and the integration of an ERN with the DBMS will be quite complicated.
3. All products operate on most mainframe platforms as well as workstations and PCs.
4. Cost of the products depends on the specific DBMS and platform used.

LIST OF ACRONYMS

| | |
|--------|---|
| ACTS | - Automated Computer Time Service |
| CAD | - Computer Aided Design |
| CASE | - Computer-Aided Software Engineering |
| DBMS | - Database Management System |
| DPI | - Dots Per Inch |
| DSA | - Digital Signature Algorithm |
| DSS | - Digital Signature Standard |
| ERN | - Electronic Research Notebook |
| GAO | - Government Accounting Office |
| GUI | - Graphical User Interface |
| LAN | - Local Area Network |
| LCD | - Liquid Crystal Display |
| NIST | - National Institute of Standards and Technology |
| OCR | - Optical Character Recognition |
| PCMCIA | - Personal Computer Memory Card International Association |
| PDA | - Personal Digital Assistant |
| PIM | - Personal Information Manager |
| RAM | - Random Access Memory |
| RTI | - Research Triangle Institute |
| RTOM | - Real-time Object Manager |
| SCSI | - Small Computer System Interface |
| SHA | - Secure Hash Algorithm |
| SHS | - Secure Hash Standard |
| SRM | - Standard Reference Material |
| VAPI | - VNS Applications Programming Interface |
| VNS | - Virtual Notebook System |
| WAN | - Wide Area Network |

REFERENCES

- [BAYE93] D. Bayer, S. Haber, and W.S. Stornetta. *Improving the Efficiency and Reliability of Digital Time-stamping*. In Sequences II: Methods in Communication, Security, and Computer Science, Pages 329-334, Springer-Verlag, New York, 1993.
- [DSS91] National Institute of Standards and Technology (NIST). *FIPS Publication XX: Announcement and Specifications for a Digital Signature Standard (DSS)*. August 1991.
- [FAHN92] P. Fahn. *Answers to Frequently Asked Questions about Today's Cryptography*. RSA Laboratories. September 1992.
- [FIPS180] National Institute of Standards and Technology (NIST). *FIPS Publication 180: Secure Hash Standard*. May 1993.
- [GAO91] General Accounting Office. *Decision on Use of Electronic Data Interchange Technology to Create Valid Obligations*. Controller General Decision B-245714. December 13, 1991.
- [HABE91] S. Haber and W.S. Stornetta. *How to Time-stamp a Digital Document*. Journal of Cryptology, 3:99-111, 1991.
- [KANA85] H.M. Kanare. *Writing the Laboratory Notebook*. American Chemical Society, 1985.
- [NECH91] J. Nechvatal. *National Institute of Standards and Technology Special Publication 800-2: Public-Key Cryptography*. April 1991.
- [NBS83] National Bureau of Standards (NBS). *Inventions and Patents*. NBS Administrative Manual, Subchapter 2.10.06, May 20, 1983.
- [NAS89] Committee on the Conduct of Science, National Academy of Science. *On Being a Scientist*. National Academy Press. 1989.
- [TAY90] David A. Taylor. *Object-Oriented Technology: A Manager's Guide*. Servio Corporation, 1990.
- [VAU93] Tay Vaughan. *Multimedia: Making it Work*. Osborne McGraw-Hill, 1993.

