NAT'L INST. OF STAND & TECH R.I.C.
A11104 593233
NIST PUBLICATIONS

NIST Special Publication 500–227

Computer Systems Technology

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards and Technology

NIST

ELECTRONIC ACCESS: BLUEPRINT for the National Archives and Records Administration

Judi Moline
Steve Otto

Dear Sir — Washington. U.S. of America. July 4. 1803.

In the journey which you are about to undertake for the discovery of the cour and source of the Mississippi, and of the most convenient water communication fro thence to the Pacific ocean, your party being small, it is to be expected that you will encounter considerable dangers from the Indian inhabitants. should you escape those dangers and reach the Pacific ocean, you may find it imprudent to hazard a return the same way, and be forced to seek a passage round by sea in such vessels as you may find on the Western coast. but you will be without money, without clothes & other necessaries; as a sufficient supply cannot be carr with you from hence. your resource in that case can only be in the credit of the US. for which purpose I hereby authorise you to draw on the Secretary of State, of the Treasury, of War & of the Navy of the US. according as you may fi your draughts will be most negociable, for the purpose of obtaining money or necessaries for yourself & your men: and I solemnly pledge the faith of the United States that these draughts shall be paid punctually at the date they are made payable. I also ask of the Consuls, agents, merchants & citizens of any nation with which we have intercourse or amity to furnish you with those supplies which your necessities may call for, assuring them of honorable and prom retribution. and our own Consuls in foreign parts where you may happen to be, are hereby instructed & required to be aiding & assisting to you in whatsoever may b necessary for procuring your return back to the United States. And to give mor entire satisfaction & confidence to those who may be disposed to aid you, I Thomas Jefferson, President of the United States of America, have written this Letter of general credit for you with my own hand, and signed it with my name.

Th: Jefferson

To Capt. Meriwether Lewis.

QC
100
U57
.500-227
1995

The National Institute of Standards and Technology was established in 1988 by Congress to "assist industry in the development of technology . . . needed to improve product quality, to modernize manufacturing processes, to ensure product reliability . . . and to facilitate rapid commercialization . . . of products based on new scientific discoveries."

NIST, originally founded as the National Bureau of Standards in 1901, works to strengthen U.S. industry's competitiveness; advance science and engineering; and improve public health, safety, and the environment. One of the agency's basic functions is to develop, maintain, and retain custody of the national standards of measurement, and provide the means and methods for comparing standards used in science, engineering, manufacturing, commerce, industry, and education with the standards adopted or recognized by the Federal Government.

As an agency of the U.S. Commerce Department's Technology Administration, NIST conducts basic and applied research in the physical sciences and engineering, and develops measurement techniques, test methods, standards, and related services. The Institute does generic and precompetitive work on new and advanced technologies. NIST's research facilities are located at Gaithersburg, MD 20899, and at Boulder, CO 80303. Major technical operating units and their principal activities are listed below. For more information contact the Public Inquiries Desk, 301-975-3058.

Office of the Director
- Advanced Technology Program
- Quality Programs
- International and Academic Affairs

Technology Services
- Manufacturing Extension Partnership
- Standards Services
- Technology Commercialization
- Measurement Services
- Technology Evaluation and Assessment
- Information Services

Materials Science and Engineering Laboratory
- Intelligent Processing of Materials
- Ceramics
- Materials Reliability[1]
- Polymers
- Metallurgy
- Reactor Radiation

Chemical Science and Technology Laboratory
- Biotechnology
- Chemical Kinetics and Thermodynamics
- Analytical Chemical Research
- Process Measurements[2]
- Surface and Microanalysis Science
- Thermophysics[2]

Physics Laboratory
- Electron and Optical Physics
- Atomic Physics
- Molecular Physics
- Radiometric Physics
- Quantum Metrology
- Ionizing Radiation
- Time and Frequency[1]
- Quantum Physics[1]

Manufacturing Engineering Laboratory
- Precision Engineering
- Automated Production Technology
- Intelligent Systems
- Manufacturing Systems Integration
- Fabrication Technology

Electronics and Electrical Engineering Laboratory
- Microelectronics
- Law Enforcement Standards
- Electricity
- Semiconductor Electronics
- Electromagnetic Fields[1]
- Electromagnetic Technology[1]
- Optoelectronics[1]

Building and Fire Research Laboratory
- Structures
- Building Materials
- Building Environment
- Fire Safety
- Fire Science

Computer Systems Laboratory
- Office of Enterprise Integration
- Information Systems Engineering
- Systems and Software Technology
- Computer Security
- Systems and Network Architecture
- Advanced Systems

Computing and Applied Mathematics Laboratory
- Applied and Computational Mathematics[2]
- Statistical Engineering[2]
- Scientific Computing Environments[2]
- Computer Services
- Computer Systems and Communications[2]
- Information Systems

[1] At Boulder, CO 80303.
[2] Some elements at Boulder, CO 80303.

NIST Special Publication 500–227

ELECTRONIC ACCESS: BLUEPRINT for the National Archives and Records Administration

Judi Moline
Steve Otto

Systems and Software Technology Division
Computer Systems Laboratory
National Institute of Standards and Technology
Gaithersburg, MD 20899-0001

April 1995

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

Reports on Computer Systems Technology

The National Institute of Standards and Technology (NIST) has a unique responsibility for computer systems technology within the Federal government. NIST's Computer Systems Laboratory (CSL) develops standards and guidelines, provides technical assistance, and conducts research for computers and related telecommunications systems to achieve more effective utilization of Federal information technology resources. CSL's responsibilities include development of technical, management, physical, and administrative standards and guidelines for the cost-effective security and privacy of sensitive unclassified information processed in Federal computers. CSL assists agencies in developing security plans and in improving computer security awareness training. This Special Publication 500 series reports CSL research and guidelines to Federal agencies as well as to organizations in industry, government, and academia.

National Institute of Standards and Technology Special Publication 500-227
Natl. Inst. Stand. Technol. Spec. Publ. 500-227, 57 pages (April 1995)
CODEN: NSPUE2

U.S. GOVERNMENT PRINTING OFFICE
WASHINGTON: 1995

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, DC 20402

EXECUTIVE SUMMARY

The National Archives and Records Administration (NARA) contains a wealth of information that documents the history of the American people. This report provides a blueprint for an electronic access system that would allow citizens in remote areas to receive this information. The report also includes suggestions for providing electronic access to information and holdings in the short-term before the "ideal system" is funded and built.

The 1994 Electronic Access Study identified four high priority needs that users and potential users would like to see NARA address. Users would like information on getting started and locating information and services available through NARA, information about records and related materials held by NARA, the capability for on-line ordering of records or publications, and the ability to download copies of actual documents held by NARA. The blueprint explicitly deals with providing a way to meet each of these needs.

Long Range Plan for Electronic Access to NARA

The first component of a system to meet the users' needs is a base public access system that could present information on getting started and on locating information or records and services available through NARA. A base system that could provide the needed functionality and the backbone for the other components would consist of an electronic access server that includes a 10-megabit SMDS line and router, an ISDN access unit, a SUN server with maintenance, Hierarchical Storage Management with 100 GB of storage, and 1.5 system administrators. The estimated cost for the base system is $597.7 K.

The second component would be one that provides the public with information about records and related materials held by NARA. The labor and overhead costs for creating descriptions of 10,000 series at 2 hours each would be approximately $1,420.5 K. (Series refers to core record groups and collections.) Estimated costs for building a traditional data management system are given in the report. Although NARA may need such a system in order to comply with archival bibliographic standards, such a system is not necessary to meet the needs of the users. The users need series level descriptions and then item level information so that they can request specific holdings.

The third component would be one that provides an on-line ordering capability. An accounting/delivery system is discussed in the report. Such a system would be cost-effective for NARA but would require that a subscription service be implemented and/or that billing is an acceptable means for recovering funds until security on the Internet has reached a point where credit cards could be used. The cost for establishing an on-line accounting/delivery service would be approximately $525 K.

The fourth component would be subsystems to allow the creation and delivery of downloadable holdings' surrogates (i.e., image files). The subsystems necessary would be production scanning stations and their accompanying CD-ROM recording stations, and reference room viewing and scanning stations. These subsystems and the required personnel, additional storage for the base electronic access server, and three additional distributed servers are estimated to cost $28,198.5 K.

This figure would provide for scanning and for the access to approximately 4.5 million images both over the Internet and at 50 NARA sites.

Recommended Short Term Actions to Meet the Users' Informational Needs

NARA must set up a prototype public access system that addresses the needs of the customers. The NARA gopher and WWW servers should contain explanatory files on how the users/customers can get started using the resources currently available, guidelines on locating the information or records and the services provided by NARA, and forms for ordering materials from NARA. Also, NARA must identify small sets of archival materials including photographs, maps, and textual documents and make them available on the gopher server.

Second, NARA must deal with the lack of a comprehensive catalogue in a pragmatic way that meets the needs of the public. We recommend that NARA make available descriptions of the 525 core NARA record groups, the Cartographic and Architectural series descriptions, the Still Pictures series descriptions, the descriptive records for the Presidential Library collections, and the descriptions of the other agency-wide systems (e.g., microfilm locator and the Master Location Register). Again, these materials should be placed on the gopher server.

Third, NARA must provide a system for users to place orders for information and materials electronically. To begin with, order forms should be made available on the gopher server. As NARA develops its electronic access system, an on-line ordering capability should be developed.

Fourth, NARA must identify sets of holdings that are of interest to large groups of customers, prioritize these materials, and actively seek cooperative partnerships with ongoing projects or seek funding for new projects. However these activities are funded, NARA must begin putting images on-line with accompanying item level ASCII text files containing descriptions. Prototypes are a good place to begin this activity but a large scale operational system is needed.

Summary

Besides planning and building the ideal system, there are interim steps that NARA could take to begin meeting their users' needs. NARA has plans for two prototypes that will help guide the future of electronic access. We recommend that NARA add a third prototype imitating the forms-based approach developed at the University of Victoria in cooperation with the British Columbia Archives. This electronic forms approach allows the structuring of descriptive data for management, and the inclusion of links for search and retrieval without the level of effort required for traditional database development. This approach has shown that an archive facility's digitization project could quickly and effectively meet the demands of its customers. The contents of this third prototype should include information about NARA, order forms that users could download for materials in traditional formats, and electronic copies of holdings. These materials should be made available on the existing NARA gopher server. Using the forms approach to structure the descriptive files and images will enable NARA to quickly begin meeting their customers' needs for electronic information. Providing access to a large group of images on this prototypical system will enable NARA to evolve a methodology that is appropriate for its particular infrastructure.

Table of Contents

List of Tables

Introduction

Goals of Project:

The Electronic Access Project of the National Archives and Records Administration had three main tasks. They were to design a methodology for exploring the informational needs of remote customers, to survey the targeted population, and to develop a blueprint for NARA's long-term information delivery systems that would enable NARA to meet their customers' needs. This report covers the last task.[1] It discusses the findings of the user study as they influence the design of an electronic access system for the National Archives and Records Administration and presents a blueprint for meeting the users' needs.

Review of the Findings Regarding Users' Needs:

The methodology developed in the first task of this study provided for the inclusion of a wide spectrum of users and potential users of NARA records and information. Data were collected in the second task from six individual interviews, meetings with 10 groups that focused on professional areas or avocations (141 attendees), meetings with 18 small groups based on geographical areas (122 attendees), electronic input (8 people), and a formal questionnaire (244 respondents). Categories of customers and potential customers surveyed included veterans, genealogists and others concerned with local history, post secondary educators, K-12 educators, government officials, law, business, and other professionals, information service providers, agriculturalists, environmentalists, and others in the general population of the State of Nebraska.

While different categories of customers had different needs and expectations (documented in NIST Special Publication 500-221), certain generalizations about the findings can be made. People are eager for electronic access to information and records. People want to access government information. Specifically, people want to know what information the National Archives has and how to access that information. They want to access that information electronically by searching by subjects and events, personal names and titles, and place names and geographical areas. Once they find the information that they want, the users want to download that information electronically, or at least be able to order it electronically. If they cannot receive it electronically, they want to receive it by first class mail. The general expectation is that turn around time for receiving an order would be a few days to 10 days (allowing a week for first class mail to reach remote regions). However, in spite of the willingness to wait, the idea expressed was, "If the material is in electronic form, then we expect it immediately."

The public library was the place most frequently cited as convenient for electronic access although substantial numbers of participants in the study also cited the office, home, and school. With on-

---

[1] The results of the first two tasks are reported in NIST Special Publication 500-221 entitled A User Study: Informational Needs of Remote National Archives and Records Administration Customers available from the authors or GPO.

line access, the place is not crucial from the point of view of the National Archives except to make sure that the public has equitable access.

Recommendations from NIST Special Publication 500-221:

Based on the survey findings, NARA must provide electronic resources in order to maximize the opportunity for citizens and organizations to locate and receive needed information. The participants in the study expressed a strong interest in knowing what is available and requested that NARA produce an inventory of their holdings nationwide. To completely meet the users' and potential users' requirements, an inventory of all the holdings of NARA would have to be available electronically. Further, a good portion of the materials would need to be on-line to support the users' requests. Toward that goal we recommended a phased approach that first meets the informational needs of users that are not currently being met. Information providers, local and family historians, government employees, business and law professionals, and veterans are working with the present system and, at least to some degree, are getting their needs met. K-12 educators, agriculturalists, environmentalists, and perhaps post secondary educators, are potential new users. Materials that meet the needs of these groups should be identified and made available electronically first.

One of the major concerns expressed in public meetings was the need for a simple user interface. Users with specific informational needs want to get on the computer, key in or select their search terms, find what they are looking for, capture that information, and get off the system. This is in the interest of economy of time and money. There will also be users who do not have a clear idea of their informational needs for whom browsing must be available. Again the requirement is for a simple user interface.

Organization of this Report:

The goal of this report is to define/design an electronic access system that would meet the users' requirements and to lay out strategies for developing the proposed on-line electronic access system. In order to achieve this goal, this report is structured as follows.

First NARA's framework is discussed. This section briefly reviews the breadth of NARA's holdings, the current and budgeted capabilities that could contribute to an electronic access system, and the status of finding aids. The issue of preservation is mentioned as not being part of this set of recommendations.

The second section is a discussion of the users' requirements as gleaned from participants in the Nebraska study. This section contains a brief discussion of the following topics: getting started and locating information from and about NARA, finding aids, on-line ordering, downloading documents, user interface, access points and information retrieval techniques, formats for dissemination of information from and about NARA, and criteria for determining which documents should be digitized.

The third section is a discussion of how NARA could meet the users' requirements. The points included are strategies and costing on the following topics: providing information on getting started and locating information or records and services, generating finding aids, providing on-line ordering capabilities, and providing downloadable documents to the public. In addition the section contains a summary of an approach for meeting the needs of the Nebraska study participants, a discussion of the extent to which digitization of documents would meet the public's informational needs, recommendations for achieving the widest possible dissemination, and recommendations regarding NARA's planned and existing information systems.

The fourth section is a proposed computer architecture for serving the public. It assumes that there is money available to buy the necessary equipment and hire the necessary people to get a system up and running in one year that would meet the stated needs of the users and provide access to 4.5 million scanned pages. The user requirements are drawn from the first four parts of the previous section. In this section the components are grouped as subsystems: the subsystem to provide electronic access to the National Archives, the subsystem for the production of files for public access, and the subsystem for NARA's operational environment. The listing of components is followed by a discussion and justification of the proposed computer components and architecture.

The fifth section is the estimated cost of implementing NARA's electronic access system as proposed in section four. Again this draws on the figures and discussion in section three.

Section six presents alternative scenarios for implementing parts of the computer architecture. It presents a discussion on the use of a service bureau or contractor, suggestions on seeking grants or partnerships and working with ongoing projects, and includes an approach in which the first step uses current NARA resources to gradually meet the users' requirements and develop an empirical basis for costing.

Section seven is the conclusion and discusses where NARA goes from here. Specifically it suggests that NARA sets up a small prototype system as suggested in section six, that NARA uses a pragmatic approach to start building a comprehensive catalogue of holdings, that NARA identifies sets of holdings to make available on the Internet and actively seeks cooperative partnerships with ongoing projects or seeks new funding, and that NARA must be aware of the changing interests of the public and make available desired materials first.

Section 1: NARA's Framework

NARA's Information and Holdings:[2]

NARA accessions and preserves materials from all three branches of Government. NARA stores and makes available permanent Federal records, Presidential records, and related documentary materials through a national network of 15 archival repositories and ten presidential libraries

[2] Information from the NARA statement of work dated Spring 1994.

located in 15 states and the District of Columbia. NARA also administers special groups of records, including the John F. Kennedy Assassination collection, the Captured German Records collection, and the Richard M. Nixon Presidential Papers. NARA maintains descriptions of records held in six affiliated archives that hold collections of records for which NARA has legal custody (e.g., the Military Academy archives at West Point). NARA also operates a nationwide system of 14 Federal Records Centers located in 12 states. The records centers house and provide access to noncurrent Federal records for agencies, including both military and civilian personnel records. Additionally, NARA maintains inventories of records created or held by Federal agencies.

NARA's holdings included more than two million cubic feet of permanent Federal records at the start of FY 1992. The presidential libraries held over 259 million pages of additional permanent records, and over 17 million cubic feet of noncurrent Federal records were housed in the Federal Record Centers. Together, NARA's facilities hold close to 20 million cubic feet of original textual material documenting the activities of the Federal Government. In addition, NARA has extensive multimedia collections, including:

* 9 million aerial photographs,
* 12 million still pictures,
* more than 4 million maps, charts, and architectural and engineering plans,
* more than 310,000 motion picture, sound and video recordings,
* almost 300,000 microforms, and
* nearly 6,000 computer data sets.

Current and Budgeted Capabilities for Electronic Public Access:

Current and Planned On-line Access. NARA has a public gopher server and has plans for a public World Wide Web (WWW) server. In addition, the Clinton administration publications are available on a WWW server called PRESIDENT at the University of North Carolina. Other electronic access resources of NARA sites include a gopher server at the University of Texas for the Lyndon B. Johnson Library, the Federal Register Electronic News Delivery (FREND) accessed through "fedworld," and anonymous ftp access from ftp.nara.gov.

Current and Planned Equipment to Facilitate Preparation for Electronic Access. NARA has a variety of desktop scanners, four production scanners, a digital camera, and a CD-ROM recorder. This equipment would be appropriate as a small part of the support for preparing materials for electronic access. Planned activities include the installation of a SUN development workstation and a NARA-wide Novell LAN/WAN.[3]

---

[3] Systems with digital imaging capabilities include: 486 PC with Xerox Model 700 scanner that can handle documents up to approximately 12" x 17" (30.48 x 43.18 cm.); Blue Rose 486/66 PC with 32 MB of RAM, a Bernoulli drive for removable storage, a UMAX Model 8100 flat bed scanner, and an Ektron 1412 digital camera (4096 x 4096 line resolution) on a copy stand that can accommodate reflection originals up to 24" x 24" and transparent originals up to 8" x 10" (20.32 x

Status of Finding Aids:

Customers find it difficult to locate the wealth of records and information available throughout NARA because of the variety of indices and finding aids created by NARA's organizational units. NARA has few automated information systems currently in use, and those that exist are fragmented and uncoordinated. Often customers who wish to use NARA records or information must visit a NARA repository or rely upon NARA staff to locate needed information.[4] What is lacking is a comprehensive, integrated set of finding aids to all NARA holdings.

Preservation:

Preservation was not part of this study. However, a digital master for electronic access should not be considered NARA's archival copy. Other studies made by NARA have pointed out the possibility of making microform copies of documents concurrently with scanning the documents or from the digital files.[5] Whether or not these microforms should be considered archival or preservation copies is not discussed in this report.

Section 2: Discussion of Users' Requirements and How They Could Be Met

The customer survey provided detailed information on what a given population wants from the National Archives. The following sections discuss the specific issues including the information, materials, and services that the users want to have available (information on getting started, finding aids, on-line ordering capability, and downloadable documents), user requirements regarding the presentation of the materials (interface, access points for information retrieval, and formats), and criteria for determining which documents should be digitized. Table 1 reviews the findings of the Nebraska study.

25.4 cm.); and a Macintosh Quadra 950 with 192 MB of RAM, 4 GB of hard disk space, SyQuest drive for removable storage, a Sharp JX-600 scanner (24 bit color, 600 dpi resolution, 11" x 17" (27.94 x 43.18 cm.) platen, reflection or transparent originals), a LeafScan 45 film scanner (24 bit color, 4000 line to 7000 line resolution depending on size of original, 35 mm to 4" x 5" (10.16 x 12.7 cm.) still photo negatives or transparencies), and a Kodak PCD 200 double-speed CD-ROM recorder purchased with authoring software to create ISO-compatible CD-ROMS's on either DOS, Windows, or Macintosh computers.

[4] Information from the NARA statement of work dated Spring 1994.

[5] National Archives and Records Administration, Digital-Imaging and Optical Digital Data Disk Storage Systems: Long-term Access Strategies for Federal Agencies, July 1994.

Table 1. Users' Requirements - Review of Findings from the User Study in Nebraska

Public wants electronic information, records, and services:
- access by browsing or searching with a "user-friendly" interface
- access by searching by subjects and events, personal names and titles, and place names and geographical areas
- copies of actual documents held by NARA by downloading the information
- on-line ordering of records or publications

Public wants information and records concerning:
- how to locate information or records and services including the finding aids
- information about records and related materials held by NARA
- Federal regulations
- education and culture
- genealogy
- science and technology
- Congressional records

Specific groups want information tailored to their needs, for example:
- teachers want everything available electronically so students can access records of interest
- environmentalists want historical maps and records pertaining to the evolution of water resources and weather
- Native Americans want information relating to treaties and land claims

Getting Started and How to Locate Information, Records, and Services of NARA:

The users and potential users need guidance on how to find the information that they need. General descriptive material about the National Archives and its mission is important. The introductory slide show presented at the beginning of the briefings in Nebraska would be an excellent overview. Once the audience is aware of the mission of the National Archives, a discussion of how to locate information, records, and services could be presented. This "getting started" and "how to locate material" information should be made available on video tape as well as on the Internet. It should appear on the NARA gopher because that is a popular starting place for users today. When the WWW server is available, this material should also be available there.

Finding Aids:

The overwhelming response of the participants in the summer 1994 study in Nebraska on the kinds of information needed was that users of the NARA resources want to receive copies of actual documents. However, they also stressed the need for indices and inventories of the complete holdings of NARA because they do not know what is available. Participants acknowledged the immense task of making such vast quantities of materials available. Further, they recognized the effort required to create inventories of NARA's holdings that could then be made available electronically.

To satisfy the need of the users and potential users to know what NARA contains, NARA must have a comprehensive catalogue of its holdings. Currently NARA has index and inventory materials in both electronic and paper formats. This diverse information should be made available in an integrated catalogue of NARA's holdings on the Internet.

What users need is to be able to search or browse on fields that contain terms pertaining to subjects and events, personal names or titles, and place names and geographical areas. They need to be able to mark items of interest or otherwise obtain a list of hits so that they may retrieve those items or further refine their list of hits. This integrated catalogue should contain descriptions of the core NARA record groups, series descriptions of photographs and maps, descriptive records of the holdings of the presidential libraries, and descriptions from other agency-wide systems. Although we have not worked with the MARC AMC format,[6] it is a format designed for archival materials and seems to allow fields that would provide the descriptive information needed by users. These fields include search terms as subjects and events, personal names or titles, and place names and geographical areas. An alternative to structured data records such as those of the MARC AMC would be ASCII text files based on authority lists with a full-text search capability.

On-line Ordering Capability:

The users and potential users of NARA resources would like to have a simple, quick way of ordering materials from the National Archives. In the meetings in Nebraska, participants requested that books of order forms be available in public libraries so that forms could be photocopied and mailed or faxed to NARA and that order forms be made available on the NARA gopher server. On the questionnaire, just under 70% of the participants requested that a mechanism for on-line ordering of records and publications be made available.

On-line ordering via E-mail would require that NARA develop a methodology for dealing with E-mail requests for materials and services. Although on-line ordering was a high priority of respondents to the questionnaire, attendees at meetings seemed to stress the need for easier access to the order forms themselves. Therefore, until a simple on-line subscription service is available, NARA should make order forms available in public libraries and on the gopher and WWW servers.

---

[6] Nancy Sahli, MARC for Archives and Manuscripts: The AMC Format, Chicago: The Society of American Archivists, 1985.

Downloadable Documents:

Not only do the users and potential users want information on getting started, finding aids, and on-line ordering capabilities, but also they want copies of documents held by NARA that they can download from the Internet. For some groups of NARA users, particularly the K - 12 teachers and students, sustained use of the NARA resources will be in the form of electronic access over the Internet. Further, these users will depend on the popular modes of access, such as gopher and WWW servers.

The majority of the users surveyed want to download written (textual) documents (91.0%) and photographs (61.1%). Maps were seen as an important resource by about one-third of the respondents to the questionnaire in each of the user categories. Audio, audio-visual, and electronic forms of data were lower priority items. Text and graphics are supported by current gopher browsers. Audio and video, as well as text and graphics, are all supported by current WWW browsers. The small groups of NARA customers requiring electronic data sets (government employees, miscellaneous professionals, and agriculturalists) might be better served by CD-ROMs or perhaps ftp access rather than gopher and WWW servers for data sets.

User Interface:

The users and potential users of the NARA electronic resources want an easy to use front end to the system that NARA provides. The graphical user interface should support text-based and visually based queries that permit users to express, in simple and perhaps in visual terms, queries concerning the availability, characteristics, and content of information that satisfies their requirements.

NARA can best provide service to their customers by using popular electronic access means, such as a gopher server and a World Wide Web (WWW) site. These interfaces would allow NARA to establish repositories of electronic information and holdings. In addition, NARA must accept and respond to E-mail from the public. This would allow the users to provide input to NARA regarding the materials that are available, as well as to make requests for additional or specific information. It would also provide a means for users to order copies of holdings that would then be made available on the public server or sent to them in a traditional format.

Data Access Points and Information Retrieval Techniques for Users:

The users surveyed in Nebraska want access by searching by subjects and events, personal names and titles, and place names and geographical areas. Users with specific informational needs want to get on the computer, key in or select their search terms, find what they are looking for, capture that information, and get off the system. Those who do not have a clear idea of their informational needs will need a browsing capability. In either case, the users want a simple user interface.

In addition, some groups of the survey participants would find it useful to search by time frame as well as by place name and geographic area. Searching by geographic coordinates was not included on the questionnaire. However, we can assume that some researchers would find this useful.

Given scale, projection, and a known point on the map, longitude and latitude could be used as coordinates for a search. Further, the user could add a time frame to his search criteria.

Formats for Dissemination of Information From and About NARA's Holdings:

The participants in the Nebraska survey felt that electronic access to NARA's information and holdings was very important. They also requested that CD-ROMs be produced for collections of materials appropriate for special user groups.

Criteria for Determining Which Documents Should be Digitized:

Selecting documents for digitization is a time consuming task. The following suggestions might facilitate the selection process. First, categories of holdings that should not be digitized are given. Then, some priorities for items to be digitized are given.

The following categories of holdings should not be digitized:
temporary records,
Agency restricted records, and
holdings that the Archivist determines will not be of general interest.

The participants of the Nebraska study said that they want access to everything. Realistically, this is neither possible nor necessary. Until the users know what is in the Archives, they would not be able to ascertain exactly what they need. However, the archivists are in a position to guide the selection of materials that should be made available. The following should be considered high priority items:
holdings that have traditionally received high use,
artifact or intrinsic value items,
photographs,
research sets based on geographic area, slice of time, or topic, and
holdings that the Archivist feels are of great importance.

Section 3: Meeting the Users' Requirements with Specifically Developed Systems

The participants in the Nebraska study made in the summer of 1994 wanted electronic access to provide for the following needs. The participants

- want to get information from NARA on how to get started on any system that NARA implements and they want guidelines on locating the information or records and services provided by NARA;
- want information about records and related materials held by NARA, hereafter referred to as finding aids;
- want to be able to order records or publications electronically; and
- want to download copies of actual documents held by NARA.

The goal of this section of the report is to present a strategy for meeting the users' requirements and estimated costs for the strategy presented.

Proposed Strategy for Providing Information on Getting Started and Locating Information or Records and Services:

NARA has a gopher site and has plans for a WWW server. These are the appropriate places to present information to the public for electronic access on how to get started and how to locate information or records and services through NARA. The participants in the Nebraska study stated that they need an easy to use interface.[7] Since the gopher and WWW servers are becoming ever more popular, they could be considered easy to use. However, care must be taken to present the basic "how to get started" information at the top level of the hierarchy. This would allow the new or novice user to select what is essentially a help file when he or she accesses the server.

In addition to presenting introductory information on the gopher and WWW servers, videotapes of the introductory slide show that was presented by NARA at public meetings held in Nebraska could be made available in public libraries. This suggestion was made by many of the participants in the Nebraska study.

Estimated Cost of Developing a Base System for Public Access:

A base system for public access should consist of two computers, a public access server and a development machine, in addition to the AIS/RAIS DBMS[8] machine used to support NARA's internal records. In order to maximize interchangeability of components and personnel, we recommend that these machines be identical. At the time of the writing of this document, Dual CPU Ultra Sparc SUN Servers are appropriate machines. The server that is used for public access

---

[7] The participants in the study did not specify what would make a user-friendly interface. There has been much published on this topic; a useful list of principles that would make an "interface more oriented to the needs and wants of users" is found in Siegfried Treu, User Interface Design: A Structured Approach, NY: Plenum Press, 1994, 196-197.

[8] AIS/RAIS refers to two unfunded NARA projects. We have used the models from these two projects as the basis for costing the traditional database component of this paper. The descriptive documents for these projects are (1) Standard Technology, Inc., "Archival Information System (AIS): Revised Cost Estimate, November 8, 1993," (hereafter referred to as AIS) and (2) National Archives and Records Administration, Office of Records Administration and Standard Technology, Inc., "Records Administration Information System (RAIS): Cost/Benefit Analysis," May 26, 1993, (hereafter referred to as RAIS). AIS/RAIS, as used in this report, refers to the functional capabilities of these proposed systems.

would be set up with a 10-megabit SMDS access to the Internet, an ISDN Access Unit,[9] and a Hierarchical Storage Management Unit (HSM) with 100 gigabytes of storage. The gopher and WWW server software that NARA has already purchased would be installed here.

We estimate that the SUN public access server would require 1.5 FTE[10] GS 11/12 system administrators in order to maintain the system on a full-time basis. These people would be responsible for maintaining the system software and for adding new files to the system as materials were readied by the staff for public access. The staging/development system would require a FTE GS 11/12 system developer. This individual would be responsible for maintaining the structure and integrity of the design of the content to be placed on the public server. The equipment and personnel needed for developing and running a robust electronic access server along with their costs are shown in Table 2.

Table 2. Costs for Providing Base Public Access to Electronic Information at NARA

COMPONENT	COST - $
Electronic Access Server - Dual CPU Ultra Sparc SUN Server with maintenance, 10-megabit SMDS Line and setup, ISDN Access Unit, HSM (Hierarchical Storage Mgmt including tape backup) with 100 GB storage at $2.2 K per GB	410.2 K
System Administrator for server - 1.5 FTE GS 11/12 with base salary of $50 K with a 2.5 multiplier	187.5 K
Staging/Development System - Dual CPU Ultra Sparc SUN Server with maintenance	76.2 K
System Developer - 1 FTE GS 11/12 with a base salary of $50 K with a 2.5 multiplier	125.0 K
TOTAL	**798.9 K**

[9] The ISDN access is to provide high speed dial up access for NARA's reference desk viewing stations. (These viewing stations will be discussed later in this report.) We believe that ISDN will provide adequate and cost-effective access for NARA's remote reference desks. If in the future, high use sites cannot meet customer demands, upgrading to faster service would be warranted.

[10] FTE is used indicating "Full Time Equivalent."

Proposed Strategy for Generating Finding Aids:

There are two major groups for whom finding aids are important, the public users and the NARA users. Perhaps the most efficient way to build a comprehensive catalogue of NARA's holdings that would be of use to both groups would be through a three-phase strategy. The first phase would be to build a catalogue that comprises the descriptions of the 525 core NARA record groups,[11] the Cartographic and Architectural series descriptions, the Still Pictures series descriptions (estimated to number 1,200), and the descriptive records at the series or sub-collection level for the presidential library collections.[12] Phase II would involve adding to the entries from Phase I the descriptions from the other agency-wide systems, such as the microfilm locator and the descriptions in the Master Location Register. Phase III would involve adding descriptions at the document or folder level where appropriate to identify those items or series that are of particular interest or importance to large numbers of NARA's customers. For the public access users, the gopher server would be the most appropriate place for these finding aids to reside. For NARA users, these records would be part of an internal database system such as the AIS/RAIS proposals.

Providing for the Immediate Needs of the Public Users via a Gopher Server. Finding aids need to be made available on the public access server as rapidly as possible. The finding aids available on the gopher server should include the general getting started and how to find information materials, short descriptive files of the images that are available for downloading, and the descriptive files from the comprehensive catalogue that NARA creates for its own use.

Specifically, in the first phase of the development of the comprehensive catalogue, the descriptions of the 525 core NARA record groups, the Cartographic and Architectural series descriptions, the Still Pictures series descriptions, and the descriptive records for the presidential library collections should be made available on the gopher public access server. The finding aids for the images found in the Still Pictures and the Cartographic and Architectural collections are important since the Nebraska study participants stressed their interest in pictures and maps. Further, as each image file created by scanning original documents is made available to users on-line, an associated ASCII text file with finding aid information should also be made available both individually and collectively for keyword search. Fields or data in the text file should include the image ID, key

---

[11] Archival Publications and Accessions Control Staff of the National Archives, List of Record Groups of the National Archives and Records Administration, April 1994.

[12] Each of the presidential libraries has a list of holdings. The holdings are usually described at the folder level. The series or sub-collection level of description might be appropriate for a general catalogue. The types of records included in the presidential libraries are manuscripts (personal papers, Federal records, and presidential records), audiovisuals (still pictures, film, video tape, audio tape, and audio discs), oral histories (transcriptions and tapes), museum objects, and printed materials (books, serials, microform, other).

words for searching (subject fields, Federal Agency, geographic codes,[13] date, author / photographer / artist), etc. These files should become part of the gopher server.

Phase two would involve adding to the preceding descriptions from the other agency-wide systems such as the microfilm locator and the descriptions in the Master Location Register. Phase three would involve adding descriptions at the document or folder level where appropriate to identify those items or series that are of particular interest or importance to large numbers of customers of NARA.

For the public users, we recommend an Internet-accessible populated database on the gopher server with immediate access to a core group of descriptions. This core group should consist of the descriptions from phase one above. Because of the interest expressed by the user groups in photographs and maps, series level descriptions (with some item level control) of the images in the Still Picture Branch and the Cartographic and Architectural Branch should be made available as rapidly as possible. Phases two and three should be implemented gradually.

Providing for the Long-term Needs of the NARA Users via a DBMS. Depending on NARA's plans concerning a database for administrative purposes, alternative systems might make sense. For example, NARA might purchase COTS library software that accepts the Archival and Manuscripts Control descriptive format of the MARC format and is also Z39.50 compatible or a more modest on-line system such as those available from Cuadra Star or Innopac (as mentioned by the Archives staff). Alternatively, NARA might need its own archival information system such as the planned centralized archival system, AIS. If this system were built, the Description and Reference modules could provide the information that the users need for part of the NARA holdings. However, as proposed, this system does not include information from the Presidential Libraries or from the Office of Records Administration on records found in agencies. Also, it is not clear that the search strategies envisioned by the Reference and Description modules of AIS would be acceptable for the users of NARA holdings. The users want to search on subjects and events, personal names and titles, and place names and geographical areas. Perhaps adding features to the system and including some aspects of the RAIS system would allow the development of a comprehensive catalogue from which records could be made available on the public access server.

Ideally the comprehensive catalogue built for NARA's internal processes would contain the subset of data that would meet the users' needs. However, realistically, developing a comprehensive catalogue down to the folder level (and item level in some cases) that will satisfy the needs of the users is an admirable goal but does not seem possible in the short term. The users need information on the holdings available to them immediately. Therefore, it is probably best to

---

[13] Geographic coding using the identifiers specified in FIPS 55-2 allows unique referencing of geographic locations in the United States. By consistently using the 10 alpha-numerics, two for state, three for county, and five for place in materials of all types when a geographic location is identified, NARA would be able to readily link images, text, and other information formats with a geographic location and users would be able to search the files easily. (FIPS 55-2, Guideline: Codes for Named Populated Places, Primary County Divisions, and Other Locational Entities of the United States and Outlying Areas, NIST, 1994, that implements ANSI X3.47-1977.)

consider a two track approach. The records for phase one should be generated and made available on the public access server. These same data should be part of the records of the NARA internal database system. However, additional fields will be added to the records to meet the NARA administrative needs.

Estimated Costs for Developing a Unified Set of Finding Aids:

Providing for the Immediate Needs of the Public Users via a Gopher Server. The cost of populating a core database is difficult to determine. Although estimates have been published as to the cost of producing a MARC AMC record, they may or may not be appropriate for the NARA holdings. The NHPRC estimate of 2 to 2 1/2 hours to prepare an automated description of an archival collection could be used to determine the time and thus the cost of populating a gopher server with finding aid data. We have estimated 2 hours per set of materials. Further, we have estimated that a Federal employee works 1,760 hours per year. (Of the 52 weeks in a year, there are 2 weeks of sick leave, 2 weeks of holidays, and 4 weeks of vacation leaving 44 weeks at 40 hours per week or 1,760 hours of work per year.)

Table 3. Cost for Populating the Base Electronic Server with 10,000 Series Description Records

COMPONENT	COST - $
11.364 FTE Archivists - GS11/12 with base salary of $50 K with a 2.5 multiplier	1,420.5 K

To create 10,000 descriptive records for this activity would take 20,000 hours costing $1,420.5 K. This assumes that it takes 2 hours per record and that the work is done by a GS11/12 Archivist with a base salary of $50 K using a 2.5 multiplier to cover benefits and overhead. (See Table 3.)

Providing for the Long-term Needs of the NARA Users via a DBMS. Table 4 presents the cost of the Reference and Descriptive modules of the Archival Information System (AIS) as estimated in November of 1993 and the cost of the schedule component from the Records Administration Information System (RAIS) as estimated in May of 1993. Specifically, the costing for the line item "Schedule from Agencies" was extracted from RAIS Cost/Benefit TABLE 4-1 "APPLICATIONS DEVELOPMENT" for years 1994 and 1995 (years 0 and 1). Other line items are extracted from AIS Revised Cost Figure 5.1 for years 0 and 1. The $958.077K total of Table 4 represents the cost of developing the base components of a comprehensive, centralized database system for internal use by the Archives staff that would also provide information needed to meet the needs of the remote users of the National Archives. However, it would not be appropriate for the remote users to access the DBMS directly; the subset of records for public access should be replicated on the public server.

Table 4: Database Development Costs (Labor) for Base System with Functional Modules[14]

LABOR COSTS FOR SOFTWARE DEVELOPMENT FOR BASE SYSTEM WITH FUNCTIONAL MODULES	COST - $
Functional Analysis	41.373 K
Rapid Prototyping	37.562 K
Detailed Design	83.789 K
Foundation Software	104.558 K
Description	95.416 K
On-line Finding Aids	36.451 K
Authority Control Module	92.185 K
Reference	128.512 K
Schedule from Agencies	338.231 K
TOTAL	**958.077 K**

Table 5 presents the costing for the hardware and software for a customized DBMS. Line item "COTS Groupware and Development Tools" is extracted from RAIS Cost/Benefit "GROUPWARE" and "APPLICATIONS DEVELOPMENT TOOLS" for Years 1994 and 1995 (years 0 and 1). Costing for line item "Workstation Upgrade" is extracted from RAIS Cost/Benefit "WORKSTATION UPGRADE" for Years 1994 and 1995 (years 0 and 1). Costing for the remaining non-hardware line items ("DBMS," Topic Text Processing," "C Development Tools and Licenses," "Easel Workbench and Development Tools," "Software and Maintenance Support for DBMS," and "Software Maintenance Support for Easel Workbench") were extracted from AIS Revised Cost Figure 4.1 for Base Year and Option I (years 0 and 1).

The development of the system is based on previous studies made for NARA as noted above. As such, it is to be developed by a contractor. However, we believe that database entries can best be made by NARA personnel. Populating the database by using contracted personnel would not serve NARA's interests since the contractors would need extensive guidance and interaction with the archivists and the entries would have to be validated by the archivists. The following figures on populating the database assume that the entries are all to be made in a single year. Therefore, the personnel necessary is high, 568.182 FTE. It would be appropriate for the work to be spread over a number of years and use be made of volunteers to enter data that is already available. This

[14] Extracted from AIS Figure 5-1 years 0 and 1 and RAIS Figure 4.1 alternative 2 , years 1 and 2.

would lower the cost and be more efficient than the proposed scheme. However, for costing purposes, we are presenting a worst case scenario.

Table 5: Hardware and Software Costs for Implementing Customized DBMS[15]

HARDWARE, SOFTWARE DEVELOPMENT TOOLS, MAINTENANCE, AND RUNTIME LICENSES	**COST - $**
Hardware for Database System (Ultra Sparc with 50 GB storage at $2.2 K/GB and $1.2 K maintenance)	186.200 K
DBMS	307.571 K
Topic Text Processing	194.700 K
C/C++ Development Tools	9.015 K
Easel Workbench and Development Tools	190.218 K
Software Maintenance Support for DBMS	132.000 K
Software Maintenance for Easel Workbench	44.000 K
Workstation upgrade (50 workstations) plus misc hardware (scanner, FAX and print server) for schedule from agencies component	78.705 K
COTS groupware and development tools	118.250 K
TOTAL	**1,260.659 K**

As in the previous costing of database or descriptive entries, we have estimated 2 hours per record. As before, we have estimated that a Federal employee works 1,760 hours per year. To create 500,000 descriptive records for this activity would take 1,000,000 hours costing $71,022.75 K. This assumes that it takes 2 hours per record and that the work is done by GS11/12 Archivists equivalents with a base salary of $50 K with a 2.5 multiplier or 568.182 FTE to do the descriptions for 500,000 record series in a single year. Ideally this work should be done by archivists familiar with the National Archives. Realistically, to do the project in a timely manner, much of the work would have to be done by contractors. The estimates are based on Federal archivists doing the work but perhaps it could be done by using some combination of archivists and contractors.

[15] Ibid.

Table 6: Personnel Costs

PERSONNEL	COST - $
3 FTE for project management team to oversee work by contractors 1 FTE GS 13 (base $60 K and 2.5 multiplier) and 2 FTE GS 11/12 (base $50 K and 2.5 multiplier)	400.00 K
568.182 FTE GS 11/12 to populate the database with 500 K series records (base $50 K and 2.5 multiplier)	71,022.75 K
TOTAL	**71,422.75 K**

In addition to staff to do the data entry, a system administrator would be necessary to maintain the database. Further, it would probably take 3 FTE to form a management team for a project of this scope. (The 2.5 multiplier takes care of management and other costs in most cases. However, since this is an intensive task, additional personnel were budgeted.) Table 6 summarizes the personnel costs for populating a database with 500 K records in a single year.

Proposed Strategy for Providing On-Line Ordering Capability:

There are a variety of options for facilitating the ordering of NARA materials. The most basic suggestion made by participants in the Nebraska study was to make all the order forms available in a book in the public library. Forms could then be photocopied by users as needed. Customers would then mail the request to NARA with payment to cover the cost of the copies. Also, strongly recommended by the participants was that the forms be made available immediately on NARA's gopher server. Users could then download the forms and mail the order as in the previous option.

If NARA were in a position to scan documents on-demand with no cost to the customer, NARA could accept E-mail requests for copies of holdings, scan those holdings, and then inform the customer via E-mail of a site from which the material could be retrieved.

Where cost recovery and/or copyright fees are involved, there are two options. The first is for NARA to provide a subscription service whereby customers pre-pay and have their accounts debited as materials are obtained. This could cover the cost of mailing hard copies and copyright fees. The second option is for NARA to accept credit card orders via the WWW server once there is adequate security for such transactions.

Estimated Costs for Providing On-Line Ordering Capabilities:

The cost for NARA to prepare a booklet of its order forms and provide these to public libraries is not estimated here. However, the cost would be recovered by the savings when NARA mails the forms as individuals request them. Further, if these forms were available for the public to download, the cost would be limited to scanning the forms and making them available on the servers and then photocopying them when they are received (although the public could be requested to submit multiple copies if NARA needs more than one copy of each request for materials). If NARA could process the orders from a single copy of the form, the expense of photocopying to emulate the current multi-page forms would be unnecessary. Further, if NARA could accept electronic orders and process them without paper copies, the savings would increase.

The option that is most appropriate for schools and other institutions would be for NARA to implement a subscription service. Although there would be an initial outlay for development and implementation of the system including design and software costs, the costs of maintaining the system could be recovered from the fees for service.

Table 7. Cost to Establish a Subscription Service for Electronic Access

COMPONENT	COST - $
Server software providing subscription capabilities	500 K
System development - training and other consulting fees	25 K

An example of such a system, actually as far as we can ascertain, the only such system, is the ExMENTIS™ system. ExMENTIS™ was developed by the ExLIBRIS Group for the Legal Information Center of Chicago-Kent Law Library[16] in cooperation with Professor Mickie Voges who provided the intellectual and legal design considerations. The software includes a group of programs that track document delivery and copyright fees and provides accounting for a subscription service based system.[17] If NARA had electronic copies of 100 K documents that would be of interest to 1000 subscribers such as schools, the initial cost of software and training on the software for such a system would be approximately $525 K. (See Table 7.) The annual

[16] M.A. Stapleton, "Quiet! Library Copyright Program at Work," Chicago Daily Law Bulletin (January 6, 1995), 141/4.

[17] ExMENTIS™ provides for the "downloading" of information or files to other ExMENTIS™ (including CURE™) systems. ExMENTIS™ enables the site to track the type of information used by subscribers so that the central site can continue to provide materials and services of interest to a particular group or groups of users.

maintenance of such a system would be approximately $100 K and includes ExMENTIS™ system upgrades and improvements, telephone support, and updates to the copyright fee tables. The costs could be recovered. These figures do not include the cost of putting documents on-line.

Optionally, customers could obtain CURE™,[18] a subset of ExMENTIS™, that would allow interaction with the NARA server for on-line ordering and look-up of copyright fees, etc. Such a system would not be required but would provide additional capabilities for schools or organizations. (Cost for CURE™ is $2,000 per site. The annual upgrade maintenance fee for each CURE™ site would be $300.)

Proposed Strategy for Creating Downloadable Documents for the Public:

The public would like to get information about NARA and its holdings, as well as electronic copies of documents held in NARA, at home, at work, and at public access stations such as schools, libraries, and government facilities.

Basically the system for providing downloadable documents to the public consists of the base electronic access server with additional HSM storage (that could provide for up to 25 million images although as costed, initial storage allows for 4.5 million images), a staging and development subsystem, a production scanning subsystem (consisting of 50 stations), a CD-ROM recording subsystem (consisting of 10 stations), and reference room subsystems (consisting of 100 viewing stations and 50 scanning stations).

The production scenario follows. NARA first determines the holdings to be made available on the public access server and determines the order in which they should be made available. Once these decisions have been made, the production scanning begins.

First the materials are prepared for scanning (staples removed, pages unfolded, damaged pages repaired or placed in mylar covers, etc., and given a unique ID) by the document preparation staff person. Once the materials are ready, they are scanned at 600 dpi[19] and the files saved as lossless

---

[18] CURE™ allows for copyright accounting and tracking at the local level down to the photocopier/ FAX machine function. The CURE™ system provides an interactive link to the ExMENTIS™ site. Requests can be made from the CURE™ site to the ExMENTIS™ site. A person at a CURE™ location can request information/research services, register for programs, etc., at the ExMENTIS™ location. Both sites can track the requests. This feature would take care of some of the communications and timeliness problems that exist between the Archives and remote users of the Archives.

[19] 600 dpi is used in this report as the suggested scan resolution. However, based on experience, NARA may determine that 300 dpi is sufficient for many classes of materials.

At the British Columbia Archives, "it was determined that a target image of 2400 x 3000 pixels in size, with 256 levels of grey or 16 million colours would be satisfactory in terms of quality, the ability to produce electronic reproductions, and yet still be reasonable in terms of file size. The resulting raw image files have physical storage sizes of 7 megabytes for black & white

JPEG compressed files[20] on magneto optical disk. It is expected that NARA-wide, the number of scans per day per single sheet, flatbed scanner would be 400 or 88,000 per year.[21] Two archivists would do quality control and create the descriptive ASCII file for each image. Images that are not of sufficient quality are deleted and the materials are rescanned. Once the quality of each image is deemed acceptable, the companion file with identifying and other descriptive information is created. Information would include the image ID, key words for searching (subject fields, Federal Agency, geographic codes, date, author / photographer / artist), etc.

It is recommended that each image file have an associated ASCII text file (i.e., the companion file) with a limited number of fields all of which could be searched. This item level indexing is the only document management approach that would meet the users needs. To have image files without descriptive text leaves the user with no way to search for desired materials. Browsing directories of image files by file identifier is not a viable solution. Likewise grouping images in a single file or even in a directory with a descriptive file would not be appropriate either since the download time for a large file or all the items in a directory would be excessive. Further, this approach would require that the users have very sophisticated graphics software if they are to browse a set of image files as their "search" method.

As the removable magneto optical disks are filled with JPEG files, they are taken to a CD-ROM recording station and the scanned files and accompanying ASCII text files are recorded on CD-ROM media. The CD-ROM would be used to populate the image hierarchy on the server. (These CD-ROMs are not to be confused with the CD-ROMs that might be produced for use by

---

photographs and 21 megabytes for colour. ... The targa image format is utilized for the capture and image correction processes. ... Once the image is scanned, it is viewed on a PC and corrected to adjust the tonal range and colour balance if necessary. ... Four new images are created from the corrected raw scan. ... a lossless JPEG compressed image ... a lossy compressed JPEG ... used for detailed examination and high quality printing ... a screen size image in GIF format, 800 x 900 pixels, with the tonal range reduced down to 32 tones for black & white images and 209 tones for colour images ... a small thumbnail size image in GIF format, 150 x 170 pixels, with the tones mapped out to a common color palette." Brant Bady, Imaging Analyst, BC Archives & Records Service.

[20] NARA must experiment with each type of scanned material to determine what compression should be used.

[21] In the case of maps and charts it is expected that only about 100 could be done per day giving 22,000 in a year. The number of scans per hour estimates vary greatly. The Chicago-Kent Law Library has a project with inner-city school students scanning journals and achieving 300 scans per student per hour. This figure was confirmed by the work done for the province of British Columbia retirement system called Superannuation. In contrast, NARA did a small test and found they only did 4 or 5 scans an hour. The only accurate way to project the work is to set up a controlled experiment using the equipment and personnel for a representative collection.

customers.) It is estimated that every five scanning stations would require a recording station[22] where the data is copied from the removable storage media to the master CD-ROM. (In order to preserve data integrity the image files are moved from the scanning stations to CD-ROM recording stations on hard media.) The CD-ROMs provide the input for the files that will be placed on the public access server. The CD-ROMs should then be kept as the digital master.

The next to last step in this process is for the staging and development person to convert the image files from JPEG to GIF for the screen size image and the small thumbnail size image and copy the companion ASCII text files onto the staging development machine and organize them into the hierarchical file structure designed for this purpose. The final step is for these new image sets to be placed on the public access server.

Estimated Cost of Developing and Populating the System of Downloadable Documents:

Table 8 shows the summarized costs for each of the subsystems necessary for creating and providing the public with access to digitized documents. The electronic access server, as shown in Table 2, is a server that is capable of interfacing with the public and responding to their requests for information. However, with only 100 GB of storage, it would not be able to meet the users demands for a large collection of images. Therefore, an additional 2,150 GB of storage must be added to the hierarchical storage management system for it to contain 4.5 million images (1/2 MB per image).

In order to scan the materials that NARA wants to make available, a variety of scanning equipment must be made available. We recommend 50 Pentium/MS Windows Production Scanning Stations. The configuration of each station should include 3 Pentiums connected by a local peer-to-peer LAN with a removable storage device (e.g., 200 MB MO drive) and appropriate media, and a scanner connected to one of the Pentiums. The majority of the stations would have base scanners. Base scanners are high resolution scanners with a scan area of up to 8.5 x 14 inches (21.59 x 35.56 cm.). To begin with, we would recommend that there be one special scanner for aerial photographs, one for still pictures, and one for maps and charts. Special scanners are necessary for those areas with special needs, i.e., documents on non-paper medium and documents of a size greater than 8.5 x 14 inches (21.59 x 35.56 cm.). Because the materials most requested by the study participants were those of historical or intrinsic value, no automatic sheet feeding scanners have been recommended.[23] Some projects have used sheet feeders with

---

[22] It would not be cost-effective to have a CD-ROM recording station for each scanning station. Therefore, in this report we have estimated that, on average, one CD-ROM recording station would be placed in close proximity to each group of five scanning stations. In cases where the scanning stations are dispersed, the magneto optical disk should be duplicated and one copy sent to a designated CD-ROM recording station location.

[23] When NARA reaches the stage where modern materials could be scanned using automatic feeders, there are good models available. For example, the Chicago-Kent Law Library and the National Library of Medicine's R&D Division have each scanned large collections of materials and made them available on-line for their subscribers. The high volume scanning

individual documents encased in mylar protectors but that approach did not seem appropriate for the initial stage of this project. Further, scanning the microfilm was not high on the list of the study participants so that option was not considered in this initial discussion. Each scanning station set-up would require space for document preparation and document holding and the personnel. The staff for each station would include the following: a GS-4/5 to do document preparation, a GS-4/5 to do scanning, a GS-9 archivist to do quality control and a descriptive ASCII text file for each image, and a GS-7 to assist the archivist with quality control and description creation. (See Table 9 for the cost associated with the scanning stations.)

Table 8. Cost of Hardware, Software, and Personnel to Upgrade the Base Electronic Public Access System (Table 2) to Provide Scanning and Access Capabilities for Approximately 4.5 Million Scans Requiring 2250 G of Storage

	Unit Cost - $	**# of Units**	**Total Cost - $**
Additional storage for Electronic Access Server (This cost allows storage for 4.5 million scans or 2250 GB).	4,730.00 K	1	4,730.0 K
Optional Distributed Servers	597.70 K	3	1,793.1 K
Scanning Stations	323.50 K	50	16,175.0 K
Recording Stations (1 for every 5 scanning stations)	87.50 K	10	875.0 K
Printers (dye sublimation and laser for each site)	10.00 K	50	500.0 K
Reference Room Viewing Stations	13.75 K	100	1,375.0 K
Reference Desk Scanning Stations	65.00 K	50	3,250.0 K
TOTAL			**28,698.1 K**

The cost of the CD-ROM Recording Stations (Table 9) is based on one recording station for each five production scanning stations. In order to keep the hardware interchangeable the recording stations should be Pentium/MS Windows based CD-ROM Recording Stations with a removable storage device of the type used by the scanning stations. Appropriate software, media (i.e., CD-ROM blanks), and a GS-7 staff person would be required for each station.

---

stations could be clustered at a few sites or widely dispersed. In fact, it would be expected that they would be moved around as projects are completed. Provision is also made in this plan for a scanning station at each site. These are not production systems but rather are used to meet customers' needs at the reference desks. (See discussion of reference desk scanning stations in this report.)

Table 9. Cost of Production Scanning Stations[24] and CD-ROM Recording Stations

	Unit Cost - $	# of Units	Total Cost - $
Production Scanning Station - base station with 400 scans / day (88,000 / year)		50	16,175 K
Hardware, software, media	86.0 K		
Personnel: Doc Prep (GS4/5 base $20 K) Scanning (GS4/5 base $20 K) QC/Description (GS9 base $30 K) QC/Description Asst (GS7 $25 K)	 50.0 K 50.0 K 75.0 K 62.5 K		
Recording Station for CD-ROMs		10	875 K
Hardware and Software	15.0 K		
Media - 1000 disks	10.0 K		
Personnel: Computer Specialist (GS7 $25 K - 2.5 multiplier)	62.5 K		

In addition to the cost of scanning and describing the materials and placing them on the public access server, NARA must make access to the electronic server available in reference rooms at the NARA sites. It is estimated that 100 viewing stations would be necessary. These would be distributed among 50 NARA sites. A few of the sites might require up to 10 viewing stations while many others would just require a single public access station in their reference room. Again, for consistency, it is recommended that these machines be Pentium/MS Windows Workstations. The public access server should provide ISDN access or the functional equivalent for high speed dial up to allow the reference desk access that does not use the bandwidth that has been allocated for the public at large. Each of the 50 reference room sites would also require a dye sublimation printer and a laser printer for providing copies of images to the users. The cost of the media must be recovered from the customers. The cost for these components is shown in Table 10.

Besides the viewing stations, each of the NARA reference room sites would require a scanning station consisting of a Pentium and a removable storage device (Table 10). A base scanner connected to the Pentium would be adequate for the majority of the stations although special

[24] The cost for base scanning stations has been estimated high to cover the cost of the three special scanning stations. The special purpose scanners are for aerial photographs, still pictures, maps, and charts.

scanners should be provided for those areas with special needs.[25] The staff required per station would be a GS-4/5 to do document preparation and scanning and a GS-9 archivist to do quality control and to create the ASCII description of each image.

Table 10. Cost of Reference Room Viewing and Scanning Stations

	Unit Cost - $	# of Units	Total Cost - $
Reference Room Viewing Station		**100**	**1,375 K**
ISDN Access	1.50 K / yr		
Hardware, software, media	6.00 K		
Personnel: System Admin. (GS 11/12 .05 FTE $50 K)	6.25 K		
Printers (1 dye sublimation and 1 laser printer for each of the 50 sites)	**10.00 K**	**50**	**500 K**
Reference Desk Scanning Station		**50**	**3,250 K**
Hardware, software, media	15.00 K		
Personnel: Doc Prep & Scan (GS4/5 base $20 K)	50.00 K		

In theory, NARA's agency-wide LAN/WAN could have a gateway to the public access server. However, realistically, sustained use of the LAN/WAN for passing image files would slow down its operational functionality. Therefore, it is suggested that NARA staff use the public viewing stations for anything more than occasional access to the public server.

As the NARA electronic access server grows in size, it might be desirable to have a distributed system. To distribute the system, additional servers could be added. The cost for these additional servers is shown in Table 11.

The implementation plan assumes that NARA will integrate the scanning of documents into their present facilities. Costs have been estimated to cover the cost of equipment and personnel but not space. It is our recommendation that a non-contractor or contractor in combination with NARA

[25] Special scanners for the cartographic and architectural branch must accommodate a broad range of input sizes. The core of the collection averages 2 to 2 1/2 feet by 3 feet (60.96 x 91.44 cm.) but ranges from 8 1/2 inches by 11 inches to 2 feet by 20 feet (21.59 x 27.94 to 60.96 x 609.6 cm.) for railroad holdings. Aerial photos are standard at 9 inches by 9 inches (22.86 x 22.86 cm.) on a negative roll. Other special needs must be ascertained.

archivists approach be taken because it keeps control of the materials in the hands of the archivists and provides a transition from the current work environment of paper documents to the future environment that will have fewer paper documents and more electronic files.

Table 11. Cost of Additional Servers to Create a Distributed System

	Unit Cost - $	# of Units	Total Cost - $
Optional Distributed Servers		**3**	**1,793.1 K**
10-megabit SMDS Line	71.0 K / yr		
ISDN Access Unit	30.0 K		
Installation and Setup of 10-megabit SMDS and Router	13.0 K		
Dual CPU Ultra Sparc SUN Server	75.0 K		
Maintenance on SPARC	1.2 K / yr		
HSM (Hierarchical Storage Mgmt including tape backup) with 100 GB storage at $2.2 K per GB	220.0 K		
System Administrator - 1.5 FTE GS 11/12 with base salary of $50 K with a 2.5 multiplier	187.5 K / yr		

Standards for Providing Electronic Documents for the Users:

ASCII text files should be used to distribute information about NARA and its holdings over the Internet. ASCII files could be marked up with SGML tags if NARA so chooses. The decision to tag text should be based on the purpose for which the text will be used. In the case of documents, SGML is particularly helpful when the text will be used for a variety of purposes and thus reformatted. Further, SGML tags can be used to facilitate searching. For example, to locate all the files with Greek words, one could search on the tag for Greek symbols in order to retrieve those documents. Currently, HTML is used for WWW documents. It cannot be parsed by any standard SGML parser and does not provide the search capabilities of SGML.[26] However, this does not make it less useful for the limited capabilities it is providing. Further, the Portable Document Format, PDF, is currently being tested towards becoming a standard. It does not make sense to convert retrospective files to this format until such time as it has proved to be a format that

[26] V. Balasubramanian and Helen Ashman, "HTML: Poison or Panacea?" ACM SIGLINK Newsletter (December 1994), p. 27.

is widely used. Once NARA accepts electronic documents in PDF as the archival record, the issue of retrospective conversion to PDF should be revisited.

Image files should be used to distribute historical textual documents, photographs, line drawings, and maps over the network. Currently GIF is the most commonly used image format on the Internet. Therefore, until another format is widely used, GIF is probably the standard of choice for the image files for the users. However, this might change in the very near future. The British Columbia Archives use GIF for thumbnail and screen size images and JPEG for "archival" images. Software for Internet access has begun to embed JPEG file support, e.g., the Netscape WWW browser. It is recommended that NARA use JPEG as the format for the digital master.[27]

The electronic images for users would be created from the digital masters. They could be offered in a variety of file types and resolutions, e.g., GIF, TIFF, 100 dpi, 300 dpi,[28] 600 dpi if necessary.

The participants in the study requested textual and graphic materials. Users of the Internet are likely to have software on their computers that allows them to view text and images. Currently audio and video can also be transferred over the Internet. However, today a large percentage of the users of the Internet do not have viewers that allow them to use these materials. Until there is wide use of audio and video, the standards for such materials are in flux. Today the audio format that goes across platforms is IMA ADPCM AUDIO CODEC. (Otherwise the most popular formats are AIFF for MAC, AU for SUN, and WAV for PC.) The options for electronic transfer of motion pictures are MPEG and Quicktime. At this time, MPEG has no audio capability but the Quicktime format has both audio and video and thus is appropriate for sound motion pictures. As for mixed or multimedia, MHEG and HyTime are both possibilities for the future. Until there is a larger market for the diverse media formats to be accessed over the Internet, it is difficult to predict which of the standards will have reached maturity.[29]

---

[27] Only NARA officials know the direction that the Archives is taking regarding electronic records and the concept of replacing some paper archival copies with electronic archival copies. This blueprint does NOT suggest replacing paper documents with electronic versions. The idea of having a JPEG digital master is that there is an electronic copy of each image in a format that conforms to an ISO standard. Further, it is suggested that this JPEG digital master be made at a high resolution (600 dpi, for example, if that is found to be of a quality that NARA would find adequate for all uses of the image). In any case, image files created by scanning original documents should be stored in a standard format with compression such as JPEG, CCITT IV FAX, or possibly JBIG.

[28] Chicago-Kent Law Library has found that scanning text (including small footnotes), maps, and patent drawings at 300 dpi is adequate for their users. The British Columbia Superannuation project has found that scanning at 200 dpi serves their purposes. They are not interested in artifacts on the pages other than the text, however.

[29] NARA must keep in touch with institutions with similar goals and with what is happening on the Internet. There is no reason for NARA to be a trend-setter but NARA should be ready to emulate what others are doing when appropriate. Specifically, NARA must watch and see which of the standards are commonly used.

Summary of the Approach for Meeting the Needs of the Nebraska Study Participants:

In order to meet the needs indicated by a group of potential users of NARA's electronic access server, NARA must provide electronic access to information on getting started and on locating the information or records and services provided by NARA, to NARA's finding aids, to order forms, and to copies of actual documents held by NARA.

The surveyed individuals in Nebraska are computer literate and expect to use all the electronic services that NARA provides. NARA has already established a NARA gopher and if this server contained pointers to carefully designed screens that provided information on getting started, locating materials in NARA, and a set of order forms for obtaining copies of NARA holdings, a good percentage of the needs expressed by the study participants would be met. When the WWW server is available it should also incorporate this information.

As for meeting the desire of the participants for a complete set of finding aids, the proposed solution is only a compromise. What the users want and need is document level records that can be searched by subjects and events, personal names and titles, and place names and geographical areas. Researchers no longer accept the concept of a human intermediary who filters their informational needs. They have learned that information relating to their research interests might be found in diverse places, and they want to be able to poke around and find new sources. By recommending that NARA create 10,000 descriptive records and place them on a gopher server to meet the immediate needs of the public, the spirit of the public users' needs is being addressed. Further, by suggesting that NARA create 500,000 descriptive records and develop a database system for these records, the needs of NARA are being addressed. Obviously, neither of these approaches meets the explicit needs of the users but it is not clear that given what they want, an item level inventory of NARA's holdings, they would be able to use it. An item level inventory for non-digitized records is beyond comprehension in scope.

As concerns the desire for copies of actual documents over the network, the participants in Nebraska identified specific sets of materials that were appealing to them. The most popular clusters of records specified as potentially useful were information related to Congress and Federal regulations and laws, including the Federal Register; agency records concerning education/culture, public lands, parks, and environment, and science and technology; and records of individuals such as genealogy and military service records.

Information providers, local and family historians, government employees, business and law professionals, and veterans are working with the present NARA systems for information retrieval and, at least to a certain degree, are getting their informational needs met. K-12 educators, agriculturalists, environmentalists, and perhaps post secondary educators are potential new users of the NARA resources. Therefore, priority should be given to providing materials for these groups. These groups, like the others, specified written or textual documents as of major importance. The few environmentalists, however, specified maps as their first choice and text as their second. The other three groups chose photographs as their second choice.

These four groups also want the same materials as the participants as a whole. The K-12 group of educators had the broadest interests and is eager to introduce original sources into the classroom starting in early grades. The K-12 educators are most interested in agency records concerning education/culture and science and technology. Education/culture was also the most important area for post secondary educators. Federal regulations and laws were high on the list of all four groups.

An appropriate way to begin to meet the users' needs would be to choose a particular historical period and make indices and records available on-line. The set of materials should include photographs and maps as part of the materials made available. The historical period could be based on a theme, such as westward expansion, or specific years, such as 1860 to 1880. To be most effective, all NARA units containing materials on the period chosen should contribute materials of various media types. Further, the materials made available should include holdings related to Congress and Federal regulations and laws; to Agency records concerning education/culture, public lands, parks, and environment, and science and technology; and records of individuals such as genealogy and military service records.

Table 12 presents a possible distribution of scanning effort that would provide the users with over 4.5 million scanned images that would support a focused effort to give users a large set of materials for limited topics. It also uses the configuration of scanning stations recommended above, i.e., 47 base scanners and 3 special needs scanners.

Table 12. Proposed Scanned Images by Type for Year 1

Type of Holding	Total in NARA	**Projected Number to be Scanned in Year 1**	Projected Size of Scanned Holdings
Aerial Photographs	9,000,000	**88,000**	44 Gigabytes
Still Pictures	12,000,000	**88,000**	44 Gigabytes
Maps, Charts, and Architectural and Engineering Plans	>4,000,000	**22,000**	22 Gigabytes
Motion Picture, Sound, and Video Recordings	>310,000	**0**	0
Microforms	<300,000	**0**	0
Computer Data Sets	<6,000	**0**	0
Pieces of Paper	5 - 9 billion pages	**4,400,000 pages**	2.2 Terabytes

Again, although NARA's customers and potential customers indicated that they wanted all of NARA's holdings available electronically, realistically such a vast resource would overwhelm the users. It is appropriate that the archivists determine the high impact and frequently requested documents and supporting evidence and make those holdings available to the public electronically.

No matter what is made available, it will not meet everyone's needs. However, it is important to develop a set of guidelines and start making material available. Sets of materials being made available should be clearly defined so that periodic review of use and usability can be made. Based on results of such reviews, priorities for future sets of materials to be made available can be modified without compromising the integrity of the conceptual electronic access plan. In order to meet the needs of the largest number of new users, it is suggested that NARA begin by scanning the types of materials indicated in Table 12. Making audio recordings available is not addressed but since it does not require excessive bandwidth, it would not be a problem. The recommended format at this time would be IMA ADPCM AUDIO CODEC. As for video and motion pictures, the infrastructure that is currently available makes transmission of these media impractical.

Discussion of the Extent to Which Digitization of Documents Would Meet the Public's Informational Needs:

It was very clear from the user study in Nebraska that a higher proportion than expected of the sample population has access to electronic networks and would like to receive documents from NARA in electronic form.

There are also groups for which electronic access does not appear feasible in the near future. There was considerable concern expressed regarding equality of access. This implies the need to provide traditional access to the information and holdings as well as electronic access. If the materials that groups and individuals without electronic access need were available electronically, the time necessary for an archivist to access it electronically and print it out and mail it to the user would be less than the current time needed for pulling collections from the shelf, finding the appropriate documents, and then photocopying them.

Recommendations for Achieving Widest Possible Dissemination:

It is important for NARA to use services and interfaces that are commonly available to current customers and potential customers, such as gopher and WWW servers. Remote customers have or will have access to the Internet and will have computers capable of viewing images. To reach the largest number of users, a block of 800 phone numbers could also be used. NARA must provide materials in popular formats using familiar interfaces. Further, NARA must provide CD-ROMs of sets of materials for researchers and educators. These CD-ROMs should contain the documents and related finding aids. It is not necessary for NARA to produce expensive interfaces with paths through the materials. Such items are expensive to produce and are more appropriately left to the value-added vendors.

Commercial CD-ROMs could be produced for collections of materials appropriate for special user groups. These could be made from the digital masters thus allowing for a high quality image to be provided. Alternatively, thumbnail images or small copies of the photographs available electronically could be placed on commercial CD-ROMs. This would allow browsing of the images without the necessity of downloading large chunks of data. Users could then download specific images, order traditional copies, or even use the thumbnail, if that resolved their informational needs. The commercial CD-ROMs should be made available at NARA sites and could be sold or given to libraries, special user groups, and individuals.

Recommendations Regarding NARA's Planned and Existing Information Systems:

Based on the extreme interest in electronic access to information and government mandates to make information available to the public, it is clear that NARA must make a major effort to implement electronic access to its information and holdings. The information that we have received from NARA does not lead us to believe that NARA currently has the resources or strategic plans to create systems that would meet the users' informational needs. However, the gopher, WWW, and other electronic access systems currently available are important first steps and could provide the basis for providing the public with information from and about NARA's holdings electronically.

The existing hardware, software, and netware that are available or budgeted are appropriate to support NARA's infrastructure. Additionally, they could support prototypical experiments that would provide empirical evidence as to the best approaches to take towards meeting the customers' needs. Empirical data could also be collected regarding the various procedures that are necessary to prepare information and documents for on-line dissemination. This data would allow a more accurate budget regarding the time and cost of developing the electronic access project than we have been able to estimate.

NARA has also had plans for developing a database for records management. The records management issue is one that we cannot deal with here. AIS/RAIS, as planned, would probably have met NARA's needs. However, such a system would not necessarily meet the needs of the customers. The customers want to know the entirety of NARA's collections. To respond to this need in a general way, NARA might provide information on the collection level, i.e., NARA has a certain number of aerial photographs taken in the 1930's with whatever description was provided by the Agency submitting the photographs to the Archives. However, when the user then wanted to find out if there was a picture of his hometown available on-line, it would not be satisfactory for him to have to download a large set of images and open each file to see if it were the correct image. Traditionally NARA has received descriptions of holdings at a collection level, i.e., often at the box or group of boxes level. While some holdings have been further processed and finding aids have been generated, this is not true of the holdings NARA-wide. Although NARA had projects planned that would have standardized the overall record keeping of the Archives, even those plans would not have provided information down to an item level that would be useful to remote users. It is not enough for users to know that there is a reference to their topic contained in "lot 49" because "lot 49" might consist of 100 boxes. Even if this were modified to "lot 49" being 40 image files, the user today would not be adequately equipped to scan those forty files to find the section that pertained to his topic.

To respond to the specific requests of the customers, NARA would need an item level inventory of the holdings. If such an inventory were available, customer requests for copies of documents would escalate and would cover the whole gamut of holdings. One of the major impediments to providing public access to copies of documents held in the Archives will be the need for search and browse capabilities at the item level, especially when the documents are presented to the user as images. To our knowledge, providing users with item level search capabilities has not been considered by NARA. NARA collects artifacts that are grouped in collections. The artifacts in a collection might all be from the same Federal Agency's office from a given time period. There is a unifying factor here that is defined on the descriptive form that is submitted to NARA with the material. However, that material was not generated with the idea that it would become a unit that would then be of interest to the public as an entity. Within collections there will be folders that might be limited to an easily defined event, person, place, etc. However, unless the materials are inventoried to the item level, it will be difficult to convince the public that they have received all of the relevant information for their topic.

Textual materials that are presented to the public as searchable ASCII text would not have to be indexed at the item level. Units of materials that are clearly related topically so that they can be given content subject headings (rather than format headings, e.g., memos, letters, reports, etc.), could be retrieved and then a further full-text search could be made to determine relevance to the users' informational need. However, the bulk of the old materials from NARA that is made available to the public will be most desirable in image format. In this way the user will have visual clues present in the original rather than just the text as converted from the original. Of course, ideally the user could have both an image file and a text file. Realistically, however, doing quality control on scanned files that have been read through optical character recognition software is tedious because the rate of errors is substantial even on printed materials.[30]

NARA has disparate systems that could provide the nucleus of a NARA-wide information system that would meet the stated needs of the customers. However, NARA lacks an institution-wide plan for integrating the pieces into a system that satisfies their mission of providing the American people with access to the information and materials that have been entrusted to them. No outsider

---

[30] Beth Oddy's work on the Adult Education Archives at Syracuse University could serve as a model for doing OCR on a portion of each document. The text thus generated could provide the index terms.

Once scanned as bit-mapped images, NARA's holdings would be accessible and networkable, and would retain the visual integrity of the original document. However, character encoded documents would allow full-text search at the expense of the non-textual elements. Ideally, NARA would provide images and character encoded versions of textual documents. (This would mean having full text plus an image database.) However, while some optical character recognition (OCR) software is excellent, it is not perfect. At .999 accuracy, 25,000 pages would have 87,000 errors while at .98 accuracy there would be about 1,740,000 errors. At $20 per hour it would cost between $2,605 and $52,200 to do the corrections on the 25,000 pages depending on whether there were .999 accuracy or .98. (Terry Menta, Strategic Planning for Electronic Image Management.) At this point, OCR is not recommended as a major component of NARA's electronic access project.

can provide NARA with a blueprint that could be followed blindly to NARA's goal. NARA must re-evaluate the existing systems in light of the customers' perceived informational needs and the suggestions offered in this report. With insights gleaned from the outsiders' viewpoints, NARA staff could cooperatively establish an information system to meet the needs of the customers.

Section 4: Proposed Computer Architecture for Serving the Public

The proposed architecture for electronic access to the National Archives is modular. The modules include servers with large storage devices that the users access over a high speed network, subsystems necessary for the production of files for public access (production scanning stations and CD-ROM recording stations), additional servers to create a distributed system, and NARA's operational environment subsystem.

Components Necessary for the Base Public Access Subsystem:

Electronic Access Server:
- 10-megabit SMDS Line
- Router
- ISDN Access Unit
- Dual CPU Ultra Sparc SUN Server
- Appropriate Software
- HSM Storage System
- System Administrator (1.5 FTE GS 11/12)

Staging and Development System:
- Dual CPU Ultra Sparc SUN Server with CD-ROM Reader
- Appropriate Software
- System Developer (1.0 FTE GS 11/12)

Additional Servers to Create a Distributed System (for each site, the requirements are):
- 10-megabit SMDS Line
- Router
- ISDN Access Unit
- Dual CPU Ultra Sparc SUN Server
- Appropriate Software
- HSM Storage System
- System Administrator (1.5 FTE GS 11/12)

Subsystems Necessary for the Production of Files for Public Access:

Production Scanning Stations - 50 each with the following:
- Document preparation and document holding areas
- 3 Pentium/MS Windows Workstations connected via a local peer-to-peer LAN
- Removable storage device (e.g., 200 MB MO drive) with media
- Scanner connected to one of the Pentiums
  - base scanner for the majority of the stations with special scanners for those areas with

special needs, i.e., documents on non-paper medium, documents of a size greater than 8.5 x 14 inches
Appropriate software
Staff per station:
Clerical (1 FTE GS 4/5) to do document preparation
Clerical (1 FTE GS 4/5) to do scanning
Archivist (1 FTE GS 9) to do quality control and ASCII description files for images
Archivist (1 FTE GS7) to assist GS 9 with quality control and ASCII descriptions

CD-ROM Recording Stations - 10 each with the following:
Pentium/MS Windows Pentium
Removable storage device of type used by the scanning stations
CD-ROM recorder and media, i.e., CD-ROM blanks
Appropriate software
Computer Specialist (1 FTE GS 7)

NARA's Operational Environment Subsystems:

Reference Room Viewing Stations (100 stations at about 50 sites, 1-10 stations per site):
Pentium/MS Windows Workstation with 16 MB RAM and 1 GB Disk
ISDN access to the Public Access Server
(ISDN or functional equivalent for high speed dial up to allow the reference desk access that does not use the bandwidth allocated for the public)
Appropriate Software
System Administrator (0.5 FTE GS 11/12)
Reference Desk Scanning Stations (about 50 sites):
Pentium/MS Windows Workstation
Removable storage device with removable storage media
Scanner - base scanner for the majority of the stations with special scanners for those areas with special needs (see footnote 22)
Appropriate software
Clerical (1 FTE GS 4/5) to do document preparation and scanning
Archivist (1 FTE GS 9) to do quality control and ASCII descriptions
Reference Room Printers - 50 sites:
Dye Sublimation Printer
Laser Printer
Records Management Subsystem (NARA internal system):
Dual Ultra Sparc SUN Server with 50 GB storage
Appropriate Software
Database Management System Software
Novell Workstations with "Groupwise" (NARA-wide LAN/WAN workstations)

Discussion/Justification of the Proposed Computer Components and Architecture:

The modules are designed to accomplish specific tasks. Together the modules provide the services necessary for electronic access. The number of production scanning, CD-ROM recording, and reference desk viewing and scanning stations provided above is given for costing purposes. In order to facilitate the implementation of the project, the modules should be cloned the desired number of times. All basic scanning stations should be identical, all CD-ROM recording stations should be identical, etc. Introducing a large variety of basic scanners, for example, would introduce unnecessary complexity especially concerning the interchangability of personnel and procedures. Scanned image quality is dependent on resolution, the method of transporting the object being scanned on the scanner to create the image, reflectance and drop-out colors, background tracking and thresholding, image enhancement and noise removal filters, the photographic capability (dithering, grayscale, color), etc. Usability of the scanner depends on the paper types handled (weights, sizes, quality), ergonomics (operator positioning, paper handling, operator controls, ease of operation), integration effectiveness (protocol used, availability of interface boards, ease of integration), etc. All of these issues must be taken into consideration when the scanners are selected.

The modular approach has taken into consideration the usual bottlenecks in imaging systems. The imaging cycle includes data preparation, scanning, description, quality control, indexing, storage, and retrieval. Bottlenecks are usually at the description and quality control task station because these aspects are manual operations that involve viewing each image. Further, description involves recognizing and entering the index retrieval terms from the authority lists. To mitigate the bottleneck, two individuals are placed at each quality control and description function.

The HSM is the on-line storage for the public access server. The HSM hardware includes hard disks and DLT tape for backup of the disks. The removable magneto optical disks are written at the scan stations as temporary storage of the scanned images. The purpose of using rewritable media is to allow the deletion of unacceptable images as determined by quality control. The CD-ROMs are the archival digital master, as well as the source of the files that are provided on the HSM. The procedure for moving files to the HSM is as follows. The files are read from the CD-ROM by the staging/development machine. They are then converted to the format appropriate for electronic distribution and placed in the file system hierarchy of the public access server.

Upgrading the LAN/WAN at Archives II (College Park) and possibly at Archives I (downtown Washington, DC) to 100-megabit service will probably be necessary after the public access server has evolved to the point where it is a useful research tool for NARA archivists in meeting the demands of their daily work. The presidential libraries have traditionally made their own decisions regarding technology. Given the small number of personnel at the other remote sites, it is unlikely that service greater than the 10-megabit LAN/WAN speed will be required.

A database of some sort will be the foundation of NARA's internal infrastructure. The series of companion ASCII text files created to identify the images made available to the public could be input into whatever database option is used.

500 KB is used in this report as an average size for each item image and its accompanying files. The accompanying files are a thumbnail image and a text file. Scanned paper item images will be smaller and photos larger. However, this seems to be a reasonable average for the items and the accompanying files on the electronic public access server.

Section 5: Estimated Cost of Implementing NARA's Electronic Access System as Proposed

Table 13 brings together the previous discussions to provide a chart of the cost of the components necessary for a public access system. The on-line subscription and records management subsystems have not been included here. Although they support electronic access, they are not essential to getting gopher and WWW servers up and populated with what the users requested. These servers can meet the basic needs of the users by providing information about and from NARA, finding aids, order forms, and access to NARA holdings.

Table 13. Cost of System Components Including Personnel to Create 4.5 Million Images and to Provide Access within NARA Facilities and via the Internet to NARA Customers

COMPONENT	COST EACH ($)	QTY	YEAR 1 COST - $	$ COST PER YEAR IN FUTURE YEARS (plus replacement of hardware)
Electronic Access Server		1	597.7 K	259.7 K
10-megabit SMDS Line	71.0 K / yr			
ISDN Access Unit	30.0 K			
Installation and Setup of 10-megabit SMDS and Router	13.0 K			
Dual CPU Ultra Sparc SUN Server	75.0 K			
Maintenance on SPARC	1.2 K / yr			
HSM (Hierarchical Storage Mgmt including tape backup) with 100 GB storage at $2.2 K per GB	220.0 K			
System Administrator - 1.5 FTE GS 11/12 with base salary of $50 K with a 2.5 multiplier	187.5 K / yr			

COMPONENT	COST EACH ($)	QTY	YEAR 1 COST - $	$ COST PER YEAR IN FUTURE YEARS (plus replacement of hardware)
Additional HSM Storage		1	4,730.0 K	4,950 K / 4.5 million images when added to base system
HSM (Hierarchical Storage Mgmt including tape backup) with 2,150 GB storage (1/2 MB per image) at $2.2 K per GB	4,730 K / 4.5 million images when added to base system			
Optional Distributed Servers		3	1,793.1 K	779.1K
10-megabit SMDS Line	71.0 K / yr			
ISDN Access Unit	30.0 K			
Installation and Setup of 10-megabit SMDS and Router	13.0 K			
Dual CPU Ultra Sparc SUN Server	75.0 K			
Maintenance on SPARC	1.2 K / yr			
HSM (Hierarchical Storage Mgmt including tape backup) with 100 GB storage at $2.2 K per GB	220.0 K			
System Administrator - 1.5 FTE GS 11/12 with base salary of $50 K with a 2.5 multiplier	187.5 K / yr			
Electronic Access Production Components				
Staging/Development System		1	201.2 K	126.2 K
Hardware/Software	75.0 K			
Maintenance on Sparc	1.2 K / yr			
System Developer - 1 FTE GS 11/12 with a base salary of $50 K with a 2.5 multiplier	125.0 K			
Production Scanning Station - base station with 400 scans / day		50	16,175.0 K	237.5 K
Hardware, software, media	86.0 K			
Personnel: Doc Prep (GS4/5 base $20 K) Scanning (GS4/5 base $20 K) QC/Description (GS9 base $30 K) QC/Description Asst (GS7 $25 K)	 50.0 K 50.0 K 75.0 K 62.5 K			

COMPONENT	COST EACH ($)	QTY	YEAR 1 COST - $	$ COST PER YEAR IN FUTURE YEARS (plus replacement of hardware)
Recording Station for CD-ROMs		10	875.0 K	725.0 K
Hardware and Software	15.0 K			
Media - 1000 disks	10.0 K			
Personnel: Computer Specialist (GS7 $25 K - 2.5 multiplier)	62.5 K			
Reference Room Viewing Station		100	1,375.0 K	775.0 K
ISDN Access	1.50 K / yr			
Equipment	6.00 K			
Personnel: System Admin. (GS 11/12 .05 FTE $50 K)	6.25 K			
Reference Room Printers		50	500.0 K	
Equipment per site	10.00 K			
Reference Desk Scanning Station		50	3,250.0 K	2,500.0 K
Equipment	15.00 K			
Personnel: Doc Prep & Scan (GS4/5 base $20 K)	50.00 K			
TOTAL			**29,497.0 K**	**10,352.5 K**

Table 14 summarizes the costs by subsystem. Again, although a number of units is given, this number should be modified to reflect the needs of the National Archives at the time implementation of a public access system is started.

Table 14. Summary Cost Based on Approximately 4.5 Million Scans per Year Requiring 2250 G of Storage

	Unit Cost - $	# of Units	Total Cost - $
Electronic Access Server and Additional HSM Storage (This unit cost includes storage for 4.5 million scans. If fewer scanning units are used, the on-line storage component could be decreased.)	5,327.70 K	1	5,327.7 K
Optional Distributed Servers	597.70 K	3	1,793.1 K
Staging/Development System	201.20 K	1	201.2 K
Production Scanning Stations	323.50 K	50	16,175.0 K
Recording Stations (1 for every 5 scanning stations)	87.50 K	10	875.0 K
Reference Room Viewing Stations	13.75 K	100	1,375.0 K
Reference Room Printers	10.00 K	50	500.0 K
Reference Desk Scanning Stations	65.00 K	50	3,250.0 K
TOTAL			**29,497.0 K**

Section 6: Alternative Scenarios for Implementing the Computer Architecture

NARA might use a variety of approaches for providing electronic access to its holdings.

Using a Service Bureau or Contractor to Scan and Describe Holdings:

Service bureaus might be hired to scan and develop ASCII descriptions of the images. The strength of such an approach is the experience in the technical aspects of scanning that the service bureau brings to the project. The weakness is the lack of flexibility since once a contract is signed, as long as the work is performed to the standard specified, NARA might not be able to modify the

work as it goes along. Using service bureaus to scan and describe series that are highly predictable and well documented would probably make sense if NARA has the funds to do massive digitization. The cost for service bureau digitization would not require the outlay for digitization equipment but unless there is funding to digitize massive amounts, it is probably not cost-effective since Archivists must still be involved in quality control and help with the descriptions.

If a contractor or service bureau were to be used, NARA must provide space and a clear set of expectations concerning the way the scanning and description is to be done. The greatest limitation in such a scheme is that in order to maintain integrity of the records and responsibility for the work flow, NARA would have to carefully document the sets of materials provided to the contractor for scanning and relinquish access to them until the contractor returned them. Depending on the value of the materials, the frequency of need for the Archivists to access these materials, and the amount of time that the materials were to be unavailable to the Archivists, turning materials over to a contractor, even if they are kept at the NARA site, might not be acceptable.

From the contractors' point of view, doing work at NARA's sites would probably mean that they are limited to working in some restricted set of days and hours and that their equipment is only available to do NARA work.

Grants and Partnerships:

NARA might also choose to develop requests for grants in cooperation with other institutions. In these cases NARA would provide masses of holdings. The cooperating party might be interested in either the potential of experimenting with the automation of the process or in the subject matter of the holdings being processed. The cost of seeking grants could be limited to the cost of writing the proposal and meeting with potential partners. However, it could also include matching funds with the amount of the grant. To effectively seek grants and partnerships, NARA must have a very clear idea of what sets of materials it will make available, what the characteristics of the sets are that might appeal to other researchers or even commercial enterprises, and what they can contribute to each potential endeavor. NARA must then be aggressive in presenting the opportunities to the communities of potential partners. Obviously, NARA will ultimately have to make adjustments in order to work with a partner, but unless NARA has a clear plan of action any partnerships based only on what someone else is willing to do for NARA will be of limited value to NARA's long term goal of providing the customers with needed resources.

Cooperation with On-Going Projects:

NARA might be in a position to provide access to its materials to gain entry into on-going digitization projects. Alternatively, NARA might seek to replicate someone else's approach and submit data to the on-going project that would help broaden the base of the work. Again NARA must first determine its plan of action and then seek projects that might benefit from parallel work. Two such possibilities that could be investigated are the Alexandria Project at USCB with its emphasis on getting maps on the Internet and the Image Evaluation project at the Library of Agriculture.

Imitating Successful Projects:

In early February 1995 the British Columbia Archives public access server was announced on the Internet. In order to provide remote access to its holdings, the British Columbia Archives established a pilot gopher server in April 1993. In 1994 they added functionality with UWI Masque forms-based software. Presently, with their operational system, they provide their users with access to their electronic holdings using gopher, WWW, and Masque servers. Materials are scanned, files checked for quality, and a companion ASCII text file using a Masque form is constructed. Links among the files are created. This electronic forms approach allows the structuring of descriptive data for management, and the inclusion of links for search and retrieval without the level of effort required for traditional database development.

Another project that offers parallels to an approach that NARA might consider is the on-line ordering and delivery capabilities system in use at Chicago-Kent Law Library. Depending on the plan of action that NARA develops, the Chicago-Kent Law Library could provide a model for on-line ordering with traditional delivery of paper-based copies of holdings or a model for on-line ordering with electronic delivery of the copies of holdings. ExMENTIS™ provides full delivery/accounting capabilities. This would allow NARA to recover costs or to cover copyright fees through billing or pre-paid billing mechanisms.

In-House Validation Experiment:

As mentioned in the first section of this document, NARA has current and budgeted capabilities to facilitate preparation for electronic access. NARA has scanners, a digital camera, a CD-ROM writer (footnote 3), and a variety of PCs, as well as a gopher server, currently available. Further, NARA plans to install a SUN development workstation, a WWW server, and a NARA-wide Novell LAN/WAN. These resources could be the nucleus of a system designed to begin meeting the needs of users for public access. In fact, at least part of these resources are serving the public; the gopher has been up for over 6 months.

However, if the available resources are to be used to meet the users' needs, goals and milestones must be established. The needs of the users must be taken into account by implementing services that begin to address their stated needs. For example, once the plan of action is established, the gopher should be modified to contain explanatory files on how the users/customers can get started using the resources currently available, guidelines on locating the information or records and services provided by NARA, and forms for ordering materials from NARA. The gopher could also contain a file with a diagram showing the plan of action and what has actually been accomplished and what the next steps are.

Further, a production scanning station should be set up to process a representative sub-set of documents and data should be collected to provide empirical evidence for time and effectiveness of the scanning procedure discussed previously. Each of the scanned images in the sub-set should have a companion ASCII text file of a descriptive nature. The images and descriptive files should then be made available on the gopher server.

Section 7: Next Steps

Prototypes:

NARA has plans for two prototypes that will help guide the future of electronic access. The first of these is a digitization project in conjunction with the University of Nebraska Press. Entitled "The Gallery of the Open Frontier" this project aims at making available to the public a digital image library of photos, paintings, and drawings that pertains to the history of the American West[31]. After the first 2000 images from the National Archives are digitized, the Press will work with scholars and specialists to determine which additional images would be of significance.

The second planned prototype is a virtual kiosk.[32] NARA plans to create an archive of post-Civil War records relating to the settlement of the West using a variety of multimedia materials. The emphasis of the project is to explore if an SGML/HTML based system, with links between the descriptive levels, would serve for NARA as the basis for a system of description.

In addition, we recommend that NARA develop a prototype electronic public access server based on the British Columbia Archives model.[33] This model is not unlike the system proposed for providing downloadable documents in this report. The operational forms-based system,[34] referred to as BCARS, contains "roughly 100,000 textual descriptions with 13 cataloging fields in the image database. Over 5000 images[35] have been converted and are on-line. ... that represents 35

---

[31] Brief proposal description by Michael Jensen of the University of Nebraska Press, "The Gallery of the Open Frontier."

[32] Information provided by NARA.

[33] The British Columbia Archives model consists of a conversion subsystem (image scanning and correction, and image conversion and processing), textual database indexing and retrieval subsystems (database conversion, linking images and text, and indexing and retrieval), and presentation subsystems (Masque, gopher, and WWW clients).

The resulting collection of ASCII text and image files is managed by an application that also provides index and search engines. Development of the system is done with Masque forms but access to the system can be with either gopher or Masque clients.

[34] The current forms-based system is based on a pilot that ran from April 1993 to September 1993. The pilot had a very limited number of images initially but a large number of textual cataloguing entries from an existing cataloguing database. Subsequent to the pilot, the software was completely rewritten to provide functionalities necessary based on user and staff feedback. The implementation of the system continued concurrently with public access via gopher during 1994 and in February 1995 WWW access was announced.

[35] The creation of the 5000 electronic image files was done by existing personnel and required approximately 580 hours.

gigabytes of raw image data that has been captured and processed through the system."[36] The software tools used for this implementation were developed by a company created by the University of Victoria in partnership with the Provincial Government Systems Corporation.[37] This approach showed that an archive facility's digitization project could quickly and effectively meet the demands of its customers.

NARA should use resources currently available to develop a similar prototype. Some additional resources, such as, software would be needed.[38] The purpose of this effort would be to make available to remote users on the internet a sub-set of NARA's holding within a few months. For the public, this would demonstrate NARA's commitment to meeting their customers' needs. For NARA, this would provide an opportunity for NARA to evaluate this approach and compare it with the SGML/HTML[39] approach used in the Virtual Kiosk.

Conclusion - Where does NARA go from here?

First of all, NARA must set up a prototype public access system that addresses the needs of the customers. The major need expressed by the customers and potential customers is for information on getting started and locating information or records and services available through NARA. In a sense this prototype has been started with the gopher server. NARA must continue putting materials on their gopher server and establish a World Wide Web (WWW) server. The gopher and WWW servers should contain explanatory files on how the users/customers can get started using the resources currently available, guidelines on locating the information or records and services provided by NARA, and forms for ordering materials from NARA. Also, NARA must identify small sets of archival materials and make them available electronically. These sets of materials should include photographs, maps, and textual documents. Once these materials are

---

[36] Brant Bady, Imaging Analyst, BC Archives & Records Service.

[37] The software is now a commercial product called Masque. The information in the application is as portable as possible. The descriptive information is maintained in ASCII text files. The images are all in common non-proprietary formats at sizes and quality levels selected by the Archives. All of these are stored as discrete files on the server; any application that can read these formats could access and use the data. The data is currently being accessed by gopher and Masque clients. WWW access is also available. The operating system of the server is UNIX. (The source of this information was Brant Bady, Imaging Analyst, British Columbia Archives and Records Service.)

[38] Masque server software (allows 3 instances on the same machine - 1 development, 1 beta, 1 production), client software, training, and first year maintenance would cost about $30 K. Additionally, an optional indexing program could be customized to NARA's requirements for about $25 K.

[39] SGML is an International Standard but HTML is still evolving. The gopher and HTTP protocols are stable.

available, NARA should review the methodology used for making the materials available, the time it took, and so on thus providing an opportunity for NARA to establish its long term plan for public access based on empirical evidence from its own experience. The previous section on prototypes addresses possibilities for implementing prototypes that would serve as testbeds for future development.

Second, NARA must deal with the lack of a comprehensive catalogue in a pragmatic way that meets the needs of the public. The participants in the study in Nebraska expressed interest in knowing what the National Archives has. The public would like a comprehensive catalogue of the holdings and assumes this would be at the item level. More in keeping with the methods used by archivists for describing holdings, we recommend that NARA make available descriptions of the 525 core NARA record groups, the Cartographic and Architectural series descriptions, the Still Pictures series descriptions, the descriptive records for the Presidential Library collections, and the descriptions of the other agency-wide systems (e.g., microfilm locator and the Master Location Register). NARA needs to establish a coordinated system into which all these diverse descriptive materials are placed. Ideally this system would evolve from the prototype based on the British Columbia Archives model. The comprehensive catalogue of the prototype could be used to bring together all the data presently in electronic form, scanning and OCR could be used to bring in paper finding aids, and data entry could bring in archivists' notes. Such a pragmatic approach will preserve the material and make it available immediately. As the next step archivists should be encouraged to systematically go through the on-line finding aids using authority lists to limit variants of spelling and choice of terms. Further, fields should be added to the records to facilitate searching by subjects and events, personal names and titles, place names and geographical codes, and dates. This data might then form the basis of a central database system.

Third, NARA must provide a system for users to place orders for information and materials electronically. To begin with, order forms should be made available on the gopher server. As NARA develops its electronic access system, an on-line ordering capability should be developed.

Fourth, NARA must identify sets of holdings that are of interest to large groups of customers, prioritize these sets of materials, and actively seek cooperative partnerships with on-going projects or seek funding for new projects. However this activity is funded, NARA must begin putting images of holdings on-line with accompanying item level ASCII text files containing descriptions. The prototypes mentioned are a good place to begin this activity but the need is for a large scale operational system.

As materials are made available on the Internet, NARA must collect data regarding the use of the materials. Further, as the comprehensive catalogue is made available and users become aware of additional materials, requests for these materials must be used to modify the priorities that may have been established. The goal of this electronic public access project is to make available the materials that the customers want.

APPENDIX A: Electronic Access System Diagrams

The diagrams that follow are intended to provide a visual explanation of some of the components discussed in the Electronic Access Blueprint.

Diagram 1: Electronic Access Components Overview

The major components itemized in this report are the electronic access server, the staging and development subsystem, the production scanning and CD-ROM recording stations, the on-line ordering subsystem, the reference rooms' viewing and scanning stations, and NARA's database subsystem. The overview diagram shows the public access system and NARA's subsystems for the preparation of materials for public access and NARA's operational environment.

Diagram 2: Production Scanning Stations

The production scanning stations are shown as consisting of three Penium/MS Windows class workstations on a local peer-to-peer LAN. One of the workstations has a scanner as well as a device for removable media. The other two workstations are used for quality control and for creating the ASCII text descriptive file for each image.

Diagram 3: CD-ROM Recording Stations

The CD-ROM recording stations are Pentium/MS Windows class workstations with a CD-ROM recorder and a device to read the removable media created by the scanning stations. Each CD-ROM recording station could probably handle the input of five production scanning stations.

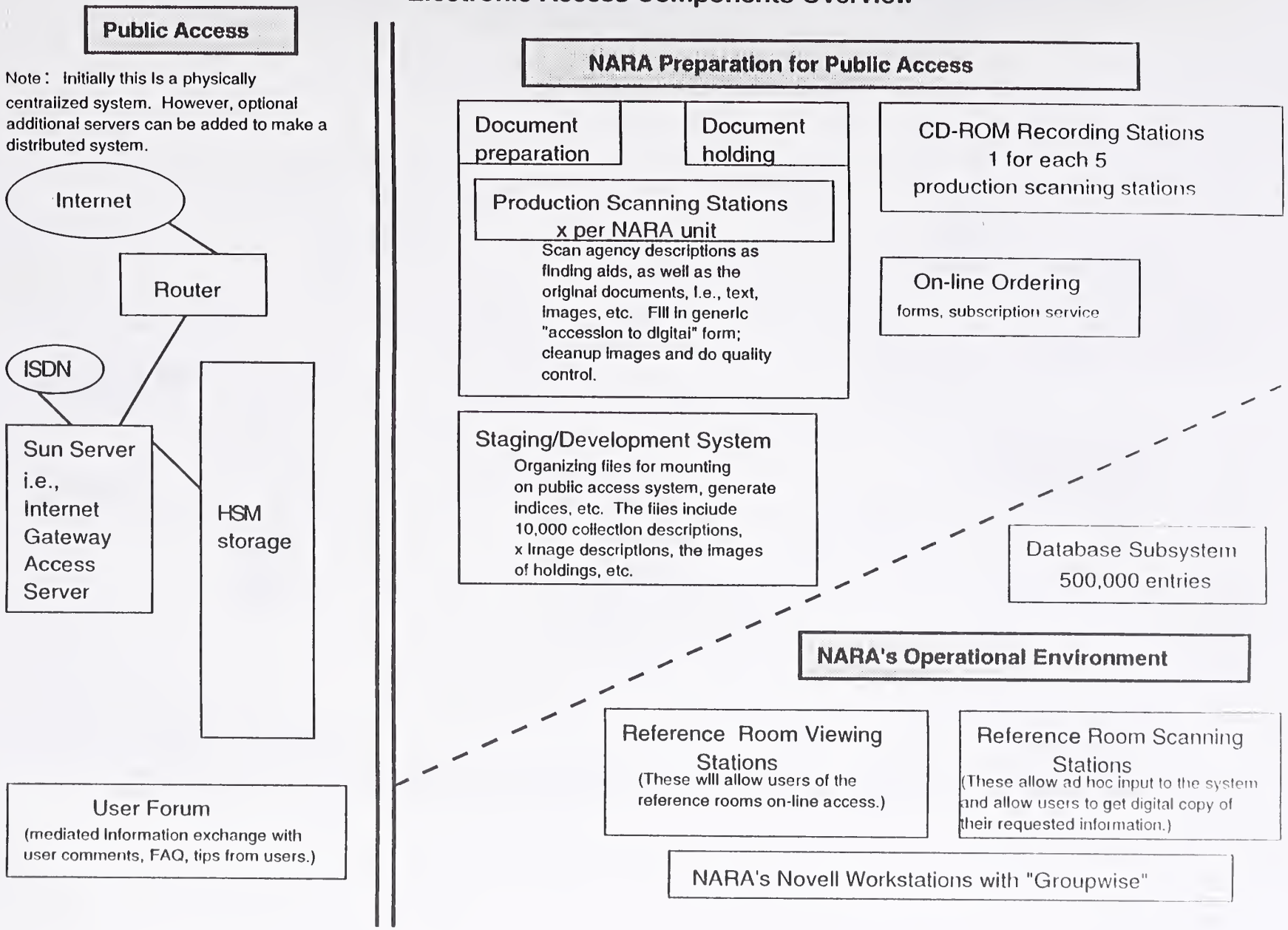
Diagram 4: Reference Room Viewing Stations

Each reference room viewing workstation is equipped with a 20-inch display and has ISDN access to NARA's public access server.

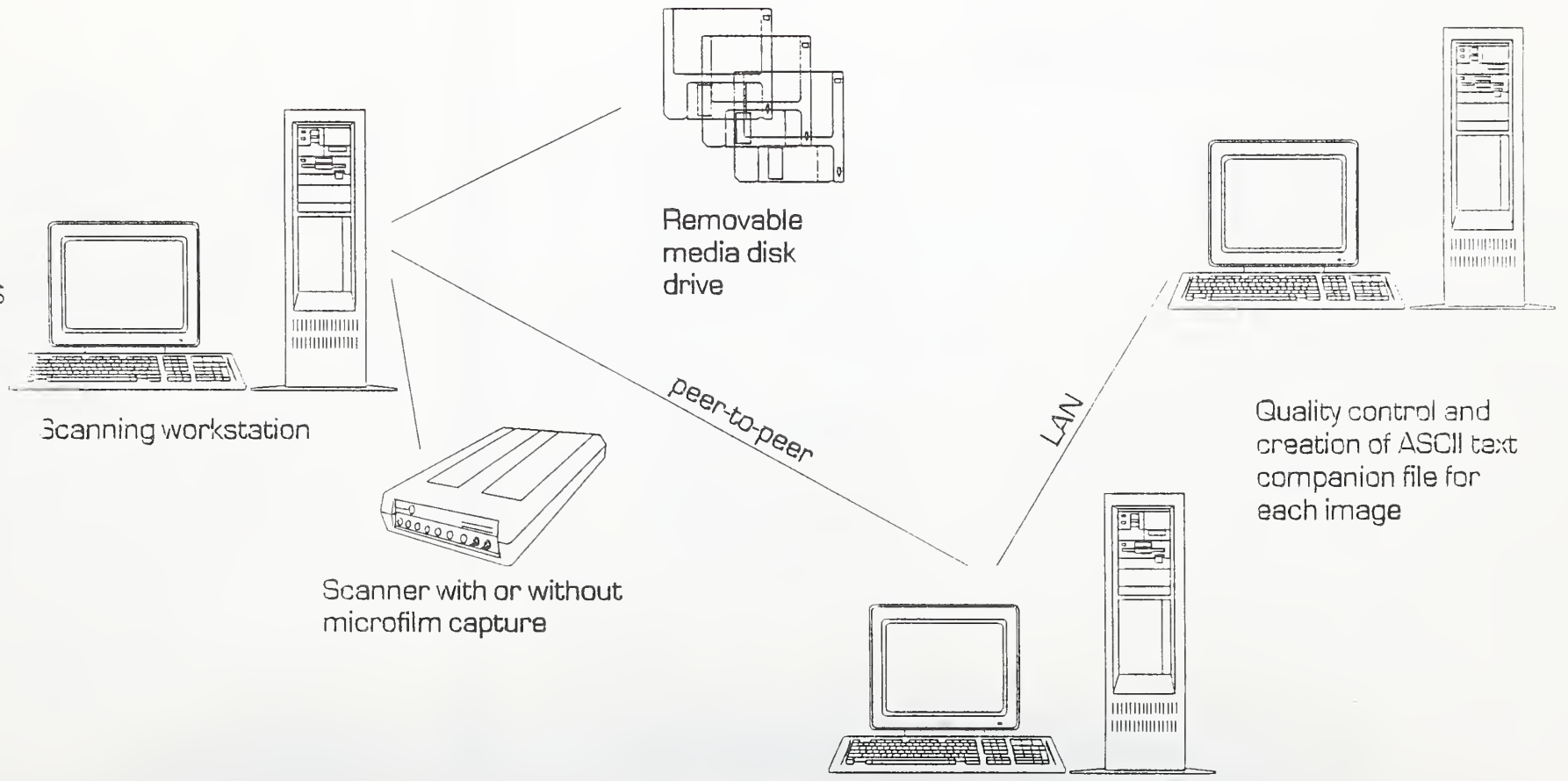
Diagram 5: Reference Desk Scanning Stations

Each reference desk scanning workstation is equipped with a scanner of the type needed for the variety of holdings at the NARA site where it is located. Customers could obtain digital copies of holdings requested.

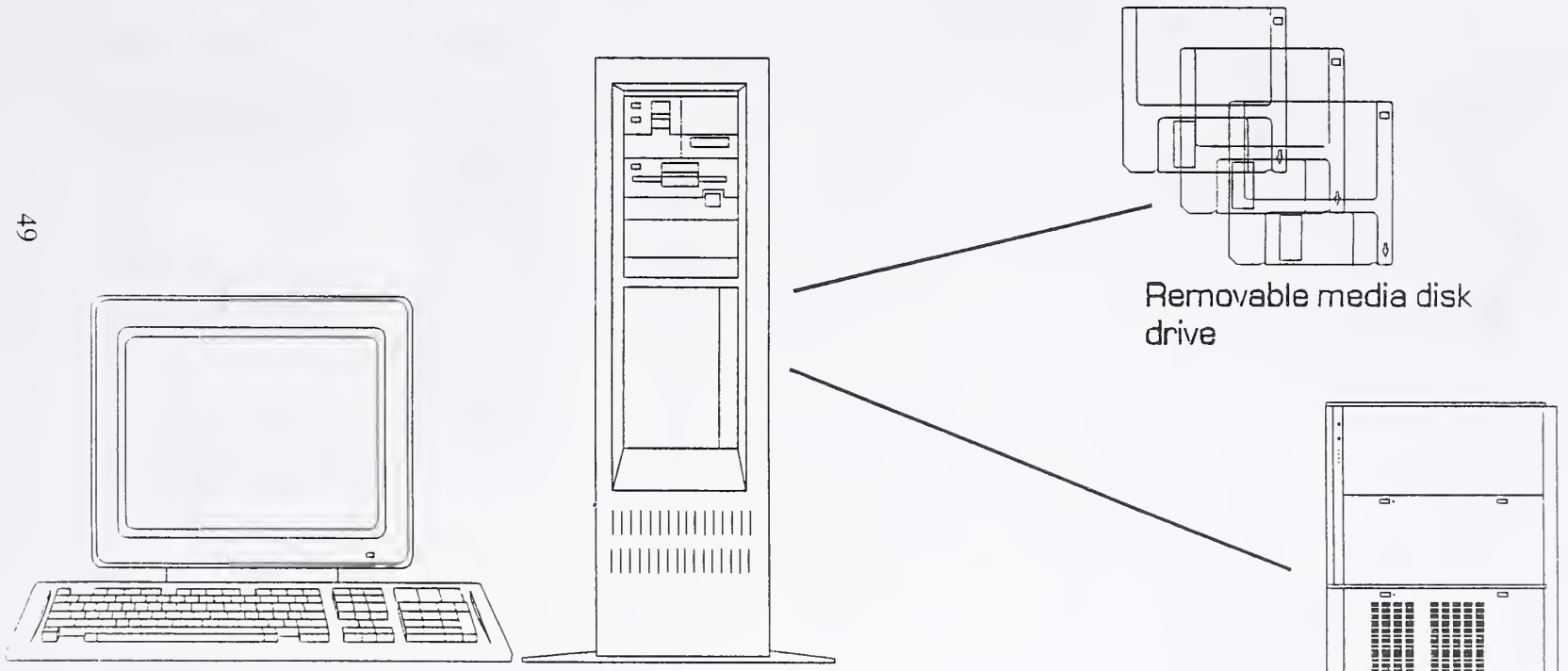
Electronic Access Components Overview



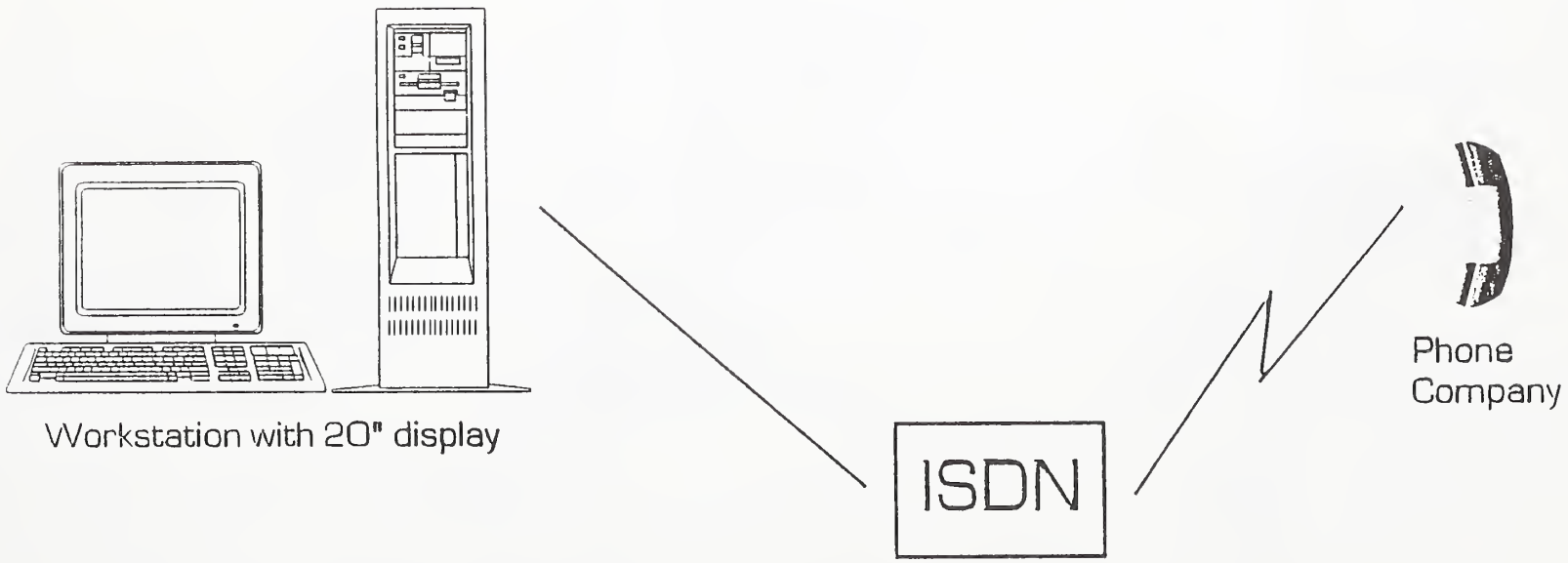
Pentium/MS Windows Scanning Stations



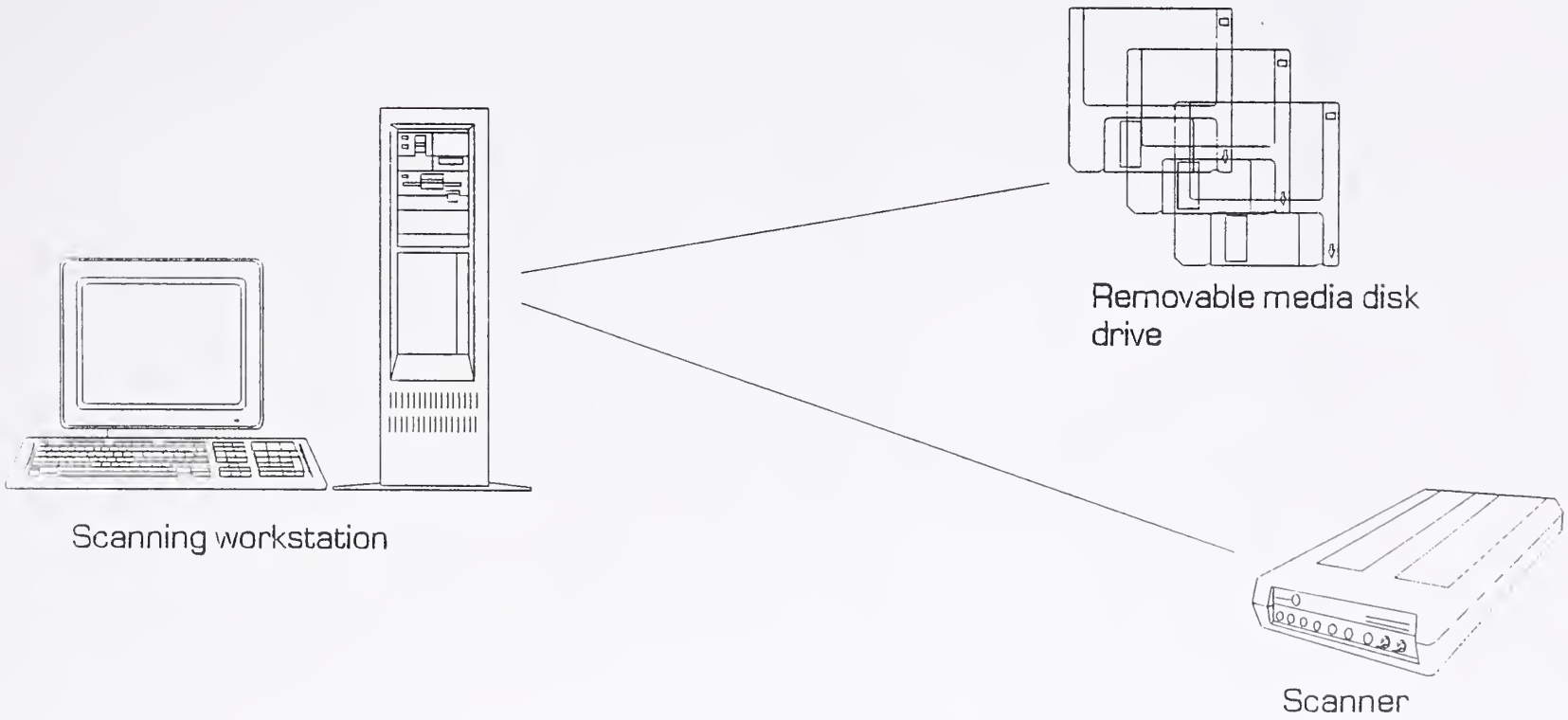
Pentium/MS Windows CD-ROM Recording Stations



Reference Room Viewing Stations



Reference Desk Scanning Stations



ANNOUNCEMENT OF NEW PUBLICATIONS ON COMPUTER SYSTEMS TECHNOLOGY

Superintendent of Documents
Government Printing Office
Washington, DC 20402

Dear Sir:

Please add my name to the announcement list of new publications to be issued in the series: National Institute of Standards and Technology Special Publication 500–.

Name _____

Company _____

Address _____

City _____ State _____ Zip Code _____

(Notification key N-503)

☆ U. S. Government Printing Office: 1995 - 386-627 (37912)

NIST Technical Publications

Periodical

Journal of Research of the National Institute of Standards and Technology—Reports NIST research and development in those disciplines of the physical and engineering sciences in which the Institute is active. These include physics, chemistry, engineering, mathematics, and computer sciences. Papers cover a broad range of subjects, with major emphasis on measurement methodology and the basic technology underlying standardization. Also included from time to time are survey articles on topics closely related to the Institute's technical and scientific programs. Issued six times a year.

Nonperiodicals

Monographs—Major contributions to the technical literature on various subjects related to the Institute's scientific and technical activities.

Handbooks—Recommended codes of engineering and industrial practice (including safety codes) developed in cooperation with interested industries, professional organizations, and regulatory bodies.

Special Publications—Include proceedings of conferences sponsored by NIST, NIST annual reports, and other special publications appropriate to this grouping such as wall charts, pocket cards, and bibliographies.

National Standard Reference Data Series—Provides quantitative data on the physical and chemical properties of materials, compiled from the world's literature and critically evaluated. Developed under a worldwide program coordinated by NIST under the authority of the National Standard Data Act (Public Law 90-396). NOTE: The Journal of Physical and Chemical Reference Data (JPCRD) is published bimonthly for NIST by the American Chemical Society (ACS) and the American Institute of Physics (AIP). Subscriptions, reprints, and supplements are available from ACS, 1155 Sixteenth St., NW, Washington, DC 20056.

Building Science Series—Disseminates technical information developed at the Institute on building materials, components, systems, and whole structures. The series presents research results, test methods, and performance criteria related to the structural and environmental functions and the durability and safety characteristics of building elements and systems.

Technical Notes—Studies or reports which are complete in themselves but restrictive in their treatment of a subject. Analogous to monographs but not so comprehensive in scope or definitive in treatment of the subject area. Often serve as a vehicle for final reports of work performed at NIST under the sponsorship of other government agencies.

Voluntary Product Standards—Developed under procedures published by the Department of Commerce in Part 10, Title 15, of the Code of Federal Regulations. The standards establish nationally recognized requirements for products, and provide all concerned interests with a basis for common understanding of the characteristics of the products. NIST administers this program in support of the efforts of private-sector standardizing organizations.

Order the **following** *NIST publications—FIPS and NISTIRs—from the National Technical Information Service, Springfield, VA 22161.*

Federal Information Processing Standards Publications (FIPS PUB)—Publications in this series collectively constitute the Federal Information Processing Standards Register. The Register serves as the official source of information in the Federal Government regarding standards issued by NIST pursuant to the Federal Property and Administrative Services Act of 1949 as amended, Public Law 89-306 (79 Stat. 1127), and as implemented by Executive Order 11717 (38 FR 12315, dated May 11, 1973) and Part 6 of Title 15 CFR (Code of Federal Regulations).

NIST Interagency Reports (NISTIR)—A special series of interim or final reports on work performed by NIST for outside sponsors (both government and nongovernment). In general, initial distribution is handled by the sponsor; public distribution is by the National Technical Information Service, Springfield, VA 22161, in paper copy or microfiche form.

U.S. Department of Commerce
National Institute of Standards and Technology
Gaithersburg, MD 20899

Official Business
Penalty for Private Use $300