INTERNET ACCESS PLAN FOR EXTERNAL DATABASES

Prepared for

**Nuclear Regulatory Commission
Contract NRC-02-93-005**

Prepared by

**Rawley Johnson
Richard Murphy**

**Center for Nuclear Waste Regulatory Analyses
San Antonio, Texas**

April 1994

CONTENTS

CONTENTS (CONT'D)

TABLES

FIGURES

iv

ACRONYMS

ACRS	Advanced Computer Review System
ADAPS	Automated Data Processing System
ATDTS	Automated Technical Data Tracking System
AUTOS	Agency Upgrade to Office Systems
AWN	Agency Wide Network
CNWRA	Center for Nuclear Waste Regulatory Analyses
CRF	Central Records Facility
DOE	Department of Energy
DTN	Data Tracking Number
DWM	Division of Waste Management
EROS	Earth Resources Observation Systems
ESNET	Energy System Network
FTP	File Transfer Protocol
GEMBOCHS	Geologic and Engineering Materials: Bibliography of Chemical Species
GENISES	Geographic Nodal Information Study and Evaluation System
GIS	Geographic Information System
GLIS	Global Land Information System
GWSI	Ground Water-Site Inventory System
HLW	High-Level Waste
HTML	Hypertext Markup Language
IRM	Office of Information Resources Management
IRIS	Improved Records Information System
LAN	Local Area Network
LLNL	Lawrence Livermore National Laboratory
MNP	Microcom Networking Protocol
MOU	Memorandum of Understanding
NFS	Network File System
NRC	Nuclear Regulatory Commission
NTS	Nevada Test Site
NWIS	National Water Information System

ACRONYMS (CONT'D)

PC	Personal Computer
POC	Point of Contact
QW	Quality of Watersystem
RIB	Reference Information Base
RPC	Remote Procedure Call
RSH	Remote Sheel
SQL	System Query Language
SwRI	Southwest Research Institute
TCP/IP	Transmission Control Protocol/Internet Protocol
TDAS	Technical Database Access System
TDB	Technical Database
TDC	Technical Data Catalog
TDIF	Technical Document Information Forms
TDOCS	Technical Document Reference Database System
USGS	United States Geological Survey
WAN	Wide Area Network
WGB	Western Great Basin
WUSE	Water-Use Data System
WWW	World Wide Web
YMSCPO	Yucca Mountain Site Characterization Project Office

ACKNOWLEDGMENTS

1 BACKGROUND

The Center for Nuclear Waste Regulatory Analyses (CNWRA) has participated directly in the analysis, design, and implementation tasks of the Nuclear Regulatory Commission (NRC) Division of Waste Management (DWM) Advanced Computer Review System (ACRS) since March 1992. The DWM Advanced Computer Review System Design Summary and Proposed On-Going Support Tasks report (Johnson, 1992) recommended several activities to complement and support the ACRS. The two development tasks with the highest priority are the Technical Document Reference Database System (TDOCS) and the Technical Database Access System (TDAS). The CNWRA has initiated work on both tasks, and this document is the second report on TDAS. The first report, External Database Access Options Report (Harloe, 1993), identified four options for the DWM and CNWRA staff to use in accessing U.S. Department of Energy (DOE), participant, and public technical databases of interest to the high-level waste (HLW) program. This work is in compliance with the Procedural Agreement between the NRC and the DOE (Nuclear Regulatory Commission, 1993) which provides for NRC read-only access to the DOE databases.

Following delivery of the first report, discussions between the NRC and the DOE confirmed that the preferred access method of "remote file access" to DOE technical databases would not be available for DWM staff in the near term. At the same time, DOE indicated that Internet access would be made available to the DWM and CNWRA staffs for only the DOE Automated Technical Data Tracking System (ATDTS). Therefore, it was decided that, for this report to be of value, it should focus on Internet access, not only to the DOE ATDTS but also to other organizations' technical databases of interest.

1.1 PURPOSE AND SCOPE

The purpose of this report is threefold: (i) to describe existing DWM ACRS, CNWRA, and DOE computer hardware and software that will be utilized in the implementation of Internet access; (ii) to identify Internet procedures currently available for a proposed list of databases or types of data important in the HLW program; (iii) to provide a brief discussion of past CNWRA experience and other access methods for use in future database access.

At the present time, the DWM ACRS computer system, as well as the CNWRA systems, have the basic capability for reading data from tapes and other media, accessing external databases via Internet, and using dedicated line connections for remote file access. All of these access options are being used by DWM and CNWRA staffs with their own data and some amount of external data.

This report deals primarily with Internet access, the access method recently made available by the DOE for the ATDTS. No additional computer system hardware and software are needed to implement Internet access. However, there is a recognition that increased capacity in additional servers and disk storage will be necessary as access increases and the databases expand. This growth is factored into the annual capacity planning by the DWM and the CNWRA. For other methods that will be used in the future to access more significant DOE technical databases, appropriate database management systems and specific application software will be required. These items are discussed briefly in this report.

1.2   OVERVIEW OF THE PLAN

The goal for technical database users is to have a reasonably simple and transparent mechanism to quickly determine what data are available, where it is, and data transfer options to avoid delays in processing the data. This goal cannot be fully attained until the DOE technical databases are accessible and appropriate electronic transfer capabilities are in place.

The on-line access recently approved for the Yucca Mountain Site Characterization Project Office (YMSCPO) and ATDTS is ideal for browsing metadata and catalogs, where the intent of the user is to determine data availability, data quality, the existence of correlative data sets before ordering specific data. So far, most of the data used by DWM and CNWRA have been transferred on different media types and not by on-line electronic transfers. The reasons for this are very practical. Most of the data desired are not in well defined databases matching the needs of the user. The database systems typically do not allow the read-only type users to specify the format of exported data files. Additionally, the data volumes are large in many instances, and on-line transfer times would be too great.

It is envisioned that this report, particularly the Internet access listed in Table 3-1, will be updated and extended periodically in the future. This will require technical staff's time to catalog more completely the databases of interest. The Internet procedures can be maintained and executed on line and are described in this report which contains the following sections.

Section 2 discusses the current Internet communications infrastructure of the sites involved. This section also discusses existing computer systems and networks. This background helps potential users select the most appropriate access methods.

Section 3 provides a list of selected organizations and related technical databases for Internet access. At this time, most of the Internet access methods are yet to be determined for the organizations' technical databases. This section also includes descriptions of DOE databases and examples of CNWRA experience in acquiring external data.

Section 4 identifies actions that should be taken in the future as more technical databases and access options become available. The experience gained by DWM and CNWRA staff along with system and network improvements will promote rapid changes in the method for accessing DOE data.

2 INTERNET COMMUNICATIONS CAPABILITIES AND CONSTRAINTS

Both the NRC and CNWRA staff have a need to access databases maintained by the DOE at the YMSCPO and at other service points on the Internet. This section will describe Internet access capabilities and mechanisms for the DWM and CNWRA in the context of the existing Local Area Networks (LANs) and Wide Area Networks (WANs) for the two organizations as shown in Figure 2-1. In addition to the specific Internet access now supported by DOE, data may be provided from a number of organizations using different access methods as follows:

- Remote login using TELNET[1]. This is the most common way to browse data catalogs or make data orders.

- Remote file access via File Transfer Protocol (FTP)[2].

- Remote file access through a client/server mechanism, such as Network File System (NFS)[3] or Prospero[4].

- Remote file access through a navigation tool such as Gopher[5] or Mosaic[6].

- File ordering and retrieval via an E-mail scheme such as NETLIB[7], LISTSERV[8], or ARCHIE[9].

---

[1] *TELNET* describes both a standard protocol (MZL-STD-1782) as well as the client software package used to emulate a terminal and communicate with a remote host. On most computers the command is called *TELNET*, but some systems give other names to this function.

[2] *FTP* refers to the command used to retrieve and send files from one computer to another using the MIL-STD-1780 protocol from the TCP/IP family.

[3] The *Network File System (NFS)* was developed by SUN as a means of creating shared file systems over networks. The specification is given in request for comments 1094 (RFC 1094) available from INTERNET.Net.

[4] *Prospero* is a special network file system developed at the University of Washington. It is particularly useful for making ARCHIE Archives appear as local file systems.

[5] *Gopher* is an information organizing and retrieval tool from the University of Minnesota. It is a client/server system that it organized around menus of file dictionaries.

[6] *Mosaic* is the name of the client software developed at the National Center for Super Computer Applications (NCSA) for accessing the World Wide Web (WWW). Mosaic will follow hypertext links that span the Internet and can also be used as a Gopher client.

[7] *NETLIB* is a simple E-mail based file retrieval system developed at Bell Laboratories. It will send users files via return E-mail.

[8] *LISTSERV* means "List Server", which refers to E-mail distribution lists. Besides maintaining distribution lists, this software can archive files and distribute them via E-mail. There are several versions of this software available on the Internet.

[9] *ARCHIE* is the name of the client/server system for searching anonymous FTP sites for files. Certain ARCHIE sites also provide for E-mail queues of the ARCHIE database. This software was developed at McGill University in Canada.

Initially, mainly metadata (information about technical data) will be acquired over the Internet in the form of catalog browsing systems like the YMSCPO ATDTS. While remote login to YMSCPO computers via the Internet or dial-up service may allow implementations that record interactive sessions, the UNIX "script" command provides another means of doing the same thing, but data recorded in this manner require much cleanup. Thus, the client/server paradigm represents the preferred method for accessing both data and metadata and is described in more detail in Section 4. In certain cases, this application will require additional database and application software to be installed at the CNWRA and NRC.

The DOE operates the Energy System Network (ESNET), which interconnects many of the DOE research centers. This network provides a gateway to the Internet for many DOE centers and alternate routing for the larger centers with direct Internet connections (Figure 2-2). DOE Energy Management (EM) has a Memorandum of Understanding (MOU) with DOE Energy Reserves (ER) which will allow DOE/EM facilities to use ESNET for Internet access, but it is not yet known how this affects the YMSCPO.

2.1 NUCLEAR REGULATORY COMMISSION DIVISION OF WASTE MANAGEMENT SYSTEMS AND NETWORKS

The DWM technical users have either IBM 486 compatible Personal Computers (PCs) with DOS or Windows operating systems, or Sun Workstations with the SunOS operating system. A Silicon Graphics workstation, and SPARC10 and SPARC1000 servers on an Ethernet network support technical computing. The Ethernet network is connected to the DWM Token Ring LAN via a Wellfleet router/concentrator. The local Token Rings provide access to the Novell Netware-based file servers. These file servers provide the Agency Upgrade to Office Systems (AUTOS) suite of office automation software. The many local Token Rings around the NRC are connected together with multiprotocol routers to form the Agency Wide Network (AWN).

All connections to the Internet at the NRC must go through a "firewall" computer (Russell, A. Gangemi Sr., G.T. 1991). This procedure requires users to log into an account on this firewall system before being allowed to perform remote logins or file transfers involving computers on the Internet. One possible exception to this rule is that access to specific computers at the CNWRA may be granted by the NRC Office of Information Resources Management (IRM) by reconfiguring the wide area router used to connect NRC with the CNWRA. The primary NRC Internet access mechanism involves a firewall system running the Sun Gateway software. The NRC network administrator may enable any NRC user to pass through the firewall on designated Transmission Control Protocol/Internet Protocol ports (TCP/IP). The two key issues involved with Internet access then are:

- Performance of the firewall when many users are accessing Internet services. It could become a bottleneck because many FTP or TELNET sessions may slow it down.

- Nonstandard port usage must be properly managed. While most providers of TELNET, FTP, Gopher, Mosaic, and other access tools use the default TCP/IP ports, there are instances of such services running on nonstandard ports. Each such instance will require the firewall configuration to be updated.

The NRC has no technical obstacles to preventing access to any database available on the Internet.
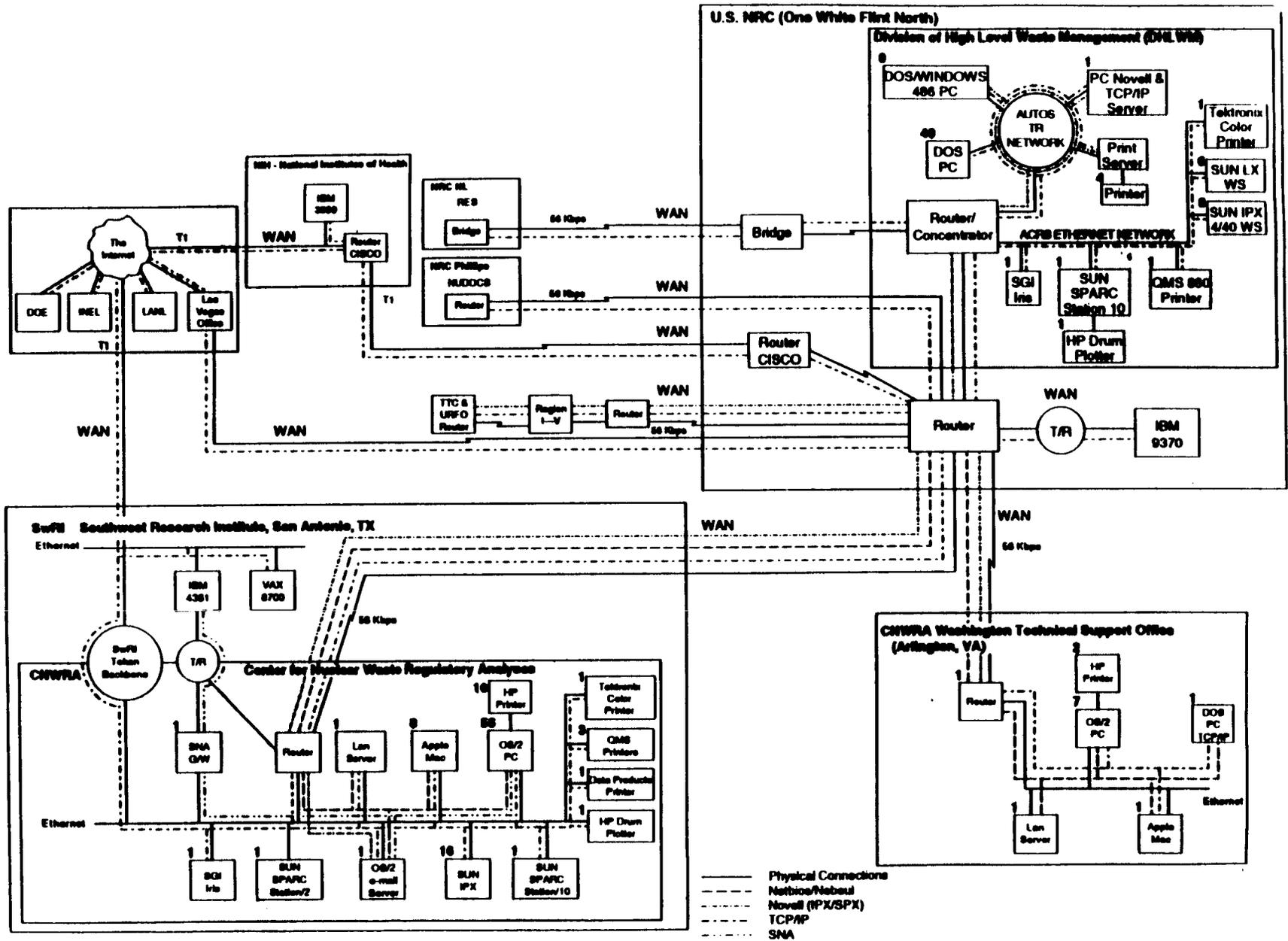
Figure 2-1. DWM and CNWRA system and network configuration

Backbone Sites

ANL - Argonne National Lab
ARM - Atmospheric Radiation Measurement (Project)
BNL - Brookhaven National Lab.
CEBAF - Continuous Electron Beam Accelerator Facility
CIT - California Institute of Technology
FIX - Federal Interagency eXchange
FNAL - Fermi National Accelerator Laboratory
FSU - Florida State Univ. (Supercomputer Computations Research Institute)
GAC - General Atomics Corp.
ISU - Iowa State University (Ames Laboratory)
ITER - International Thermonuclear Experimental Reactor (Project)
LANL - Los Alamos National Lab.
LBL - Lawrence Berkeley National Lab.
LLNL - Lawrence Livermore National Lab.
MIT - Massachusetts Institute of Technology
NYU - New York University
OER - Office of Energy Research, Washington, D.C.
ORNL - Oak Ridge National Lab.
PNL - Pacific Northwest Lab.
PPPL - Princeton Plasma Physics Lab.
SLAC - Stanford Linear Accelerator Lab.
SNLA - Sandia National Laboratory, Albuquerque
SSC - Superconducting Super Collider
UCLA - Univ. of California, Los Angeles
UTA - Univ. of Texas, Austin (Fusion Research Center)
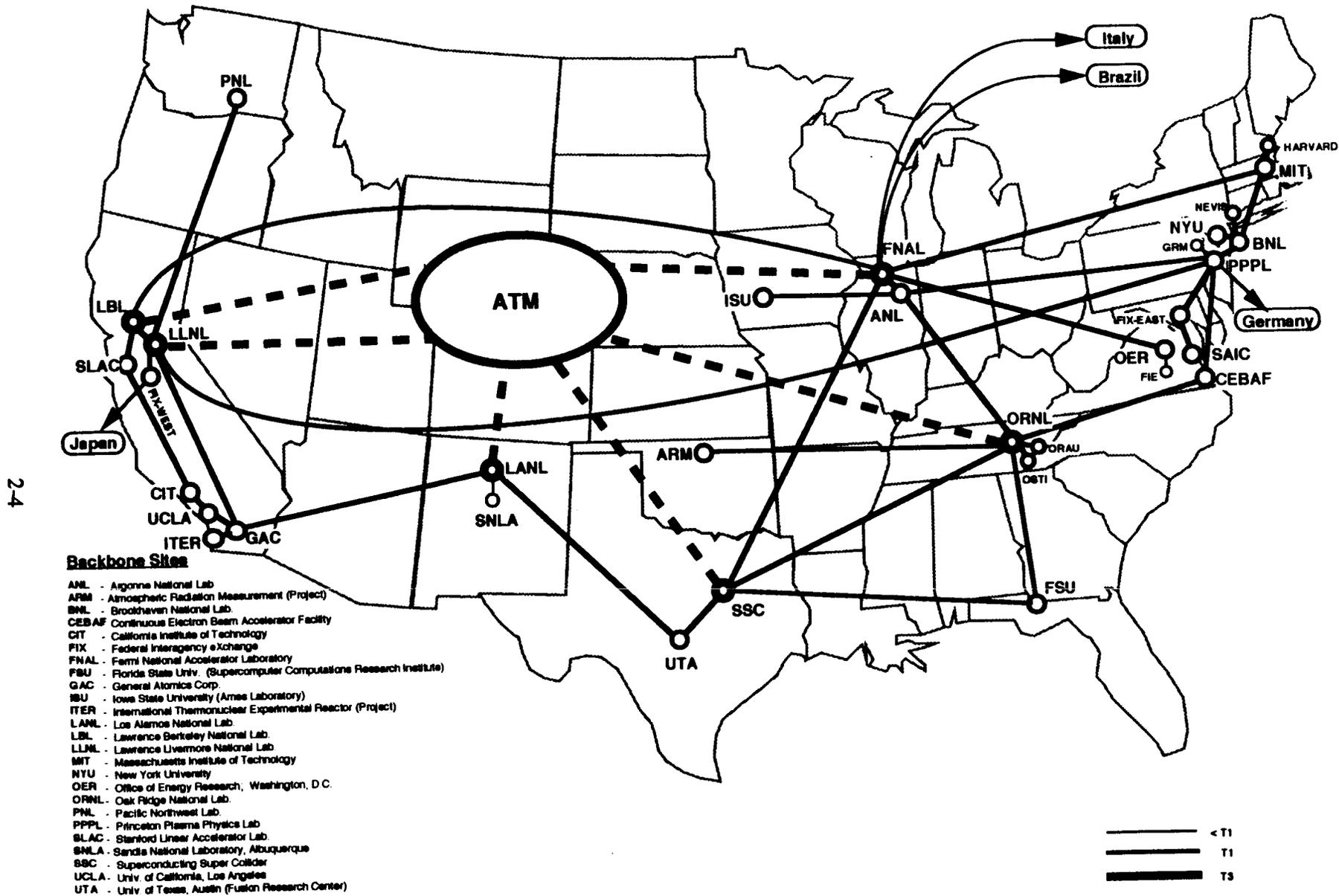
< T1
T1
T3

Figure 2-2. ESNET Backbone 1993 Map

2-4

2.2   CENTER FOR NUCLEAR WASTE REGULATORY ANALYSES SYSTEMS AND NETWORKS

Most of the CNWRA technical staff have Sun SPARC10/IPX, Macintosh Quadra 950, or IBM PS/2 95 workstations with SunOS, System 7, or OS/2 operating systems, respectively. Access to Sun SPARC10 servers for the Geographical Information System (GIS) and FORTRAN applications and a Silicon Graphics ONYX Server is also available on an Ethernet network. All of the CNWRA workstations and PCs are connected to the wide area router connecting the CNWRA with the NRC WAN. The router connection is configured to only allow communications from specific computers to pass on the link. CNWRA users also have access to the Southwest Research Institute (SwRI) network and, through it, to the Internet. VAX and CRAY computers are accessed over this network for running codes.

The CNWRA office in San Antonio has Internet access without restrictions. There are multiple independent circuits to the Internet via the SwRI internal network. Because SwRI has a direct T1 line to the backbone node at Rice University, network loading should never be a factor in remote database access. Also, there are three independent ways to reach the Internet from SwRI, so circuits to Internet should always be available.

The CNWRA internal network is being upgraded to support client/server applications. Central database and file servers provide files and data to client applications at user's workstations using TCP/IP protocols, such as; Remote Procedure Call (RPC) or NFS. File transfer and remote login between systems is accomplished using the FTP, TELNET and Remote Sheel (RSH) protocols. All CNWRA clients have TCP/IP software for their respective operating systems. These are the same protocols and access methods used to communicate with the many providers of services and databases on the Internet.

2.3   DEPARTMENT OF ENERGY YUCCA MOUNTAIN SITE CHARACTERIZATION PROJECT OFFICE SYSTEMS AND NETWORKS

The YMSCPO uses Sun workstations and PCs running DOS and Windows, as well as a Silicon Graphics computer and Sun server to support the Geographic Nodal Information Study and Evaluation System (GENISES) database. The PCs use TCP/IP software from Sun Microsystems to access file systems on the Sun file server using the NFS protocol. The Geologic and Engineering Materials: Bibliography of Chemical Species (GEMBOCHS) database is maintained on a VAX computer. The YMSCPO uses Ethernet as the LAN and currently does not support technical database connections to outside networks.

Recently, DOE requested the following information and provided the seven step procedure, as quoted below, from the NRC to initiate Internet access to the YMSCPO ATDTS.

"The NRC should provide the DOE with the name of a single point-of-contact (POC) with signature authority to authorize access for all NRC staff. The NRC should appoint a technical POC for day-to-day operational contact (such as notification of system unavailabilities), resolving user questions, interactions with YMSCPO technical staff, etc.). In addition, we need an estimate from the NRC of the number of concurrent NRC users that will be accessing the ATDTS database.

The following procedures related to the NRC access to the ATDTS are envisioned:

i) The NRC POC will sign all requests for NRC personnel requesting YMSCPO computer accounts. The NRC POC will also be responsible for notifying the YMSCPO in request of termination of any accounts.

ii) Each NRC person requesting a YMSCPO computer account will be required to read an unclassified computer security briefing and sign an acknowledgement of responsibilities.

iii) Each YMSCPO computer account will be given on a "per-user" basis, each NRC individual requesting access must have their own YMSCPO computer account that is not accessed or used by anyone else.

iv) The NRC user may be required to obtain some type of application-specific training. Depending on the number of NRC staff requiring training, it would probably be most cost-effective for the NRC to send a few people to the YMSCPO for training, and these individuals could train any additional NRC staff.

v) The computer accounts that will be granted to the NRC staff will be captive and restricted accounts that will have read-only access to the ATDTS data.

vi) The YMSCPO will support a vt100 terminal emulation, via dial-in access. It is strongly recommended that Microcom Networking Protocol (MNP) error-correcting modems be utilized. Dial-in access is supported for 2400 to 9600 baud rates. Internet access is planned within 90 days of the NRC response.

vii) The Technical Data Catalog is output in a text file in landscape mode 132 characters across. This is the standard report format. If customized reports are required, those requirements should be made known."

3 TECHNICAL DATABASE INTERNET ACCESS PLAN AND EXPERIENCES

This section contains the identification of organizations, databases, and access methods for external databases of interest to the DWM and CNWRA. Section 3.1 describes the major DOE databases of interest including the ATDTS which is the only one the DOE supports on Internet at the current time.

Section 3.2 provides a list of organizations, databases, and Internet access methods/status (Table 3-1) for the, DWM and CNWRA staff use at the current time. It is recommended that this information be kept on line electronically because it is likely to change as the remote database systems evolve and new systems become available. A Gopher or World Wide Web (WWW) server provides a good way to accomplish this. Table 3-1 will be provided electronically in HyperText Markup Language (HTML) for this purpose. When it is used with a client program like Mosaic, it not only documents the access procedures but it can also perform the procedures.

Section 3.3 contains examples of the CNWRA staff's experience in accessing some of the databases. As stated in the overview of this plan, much technical work remains to be done to catalog technical data and establish access procedures that assure that the appropriate data can be made available to support the HLW program.

3.1 DEPARTMENT OF ENERGY DATABASE DESCRIPTIONS

Databases at the YMSCPO in Las Vegas, Nevada, will be the highest priority for electronic access. All project participants submit data to this location. Therefore, providing electronic access to systems in this office could potentially supply the NRC with access to much of the data required for analysis and review.

The Technical Database (TDB) contains the technical data stored and maintained by the YMSCPO staff that has been extracted from Technical Data Packages sent from participant sites. The TDB consists of two components, GENISES and GEMBOCHS.

3.1.1 Geographic Nodal Information Study and Evaluation System

The GENISES portion of the TDB contains large amounts of data of potential use to the NRC. Participant data, maps, boundaries, fault data, seismic data, and other types of data are included in this system. As Project participants send data packages to the YMSCPO, the technical data contained in these data packages are entered into the GENISES. In order to support the viewing and analysis of this technical data, a large set of geographical information including boundaries, fault lines, well locations, roads, streams, etc., is used to provide a graphical depiction of the Yucca Mountain site. The technical data itself is referred to as "Type 1" data and is currently estimated to be about 500 megabytes in size. The geographical support data is referred to as "Type 2" data and is currently estimated to be about 2 gigabytes in size.

Due to the amount and variety of data contained in GENISES, this system will probably be the focus of much of the effort to access technical data for use on the ACRS SPARC10 server in the DWM when access is supported by the DOE.

3.1.1.1 Yucca Mountain Site Characterization Project Office Parameter Dictionary

As technical data packages arrive at the YMSCPO, technical data are placed into the TDB, typically into the GENISES data system. The programmatic information and document references are stored in the ATDTS and published in the Technical Data Catalog (TDC). The TDC lists data by its Data Tracking Number (DTN) and gives a brief English description of the data. A single entry in the catalog may represent the entire set of data collected under a particular study and submitted by some principal investigator. Details about the data content are not present in the catalog. For example, data on well levels, casing information, and well depths collected in a study may all be represented by a single item, with a single DTN, in the catalog. In order to deal with this grouping of different types of data under a single DTN, the YMSCPO is developing Parameter Dictionary. The Parameter Dictionary will consist of a detailed listing of "normalized" parameters such as, geologic, hydrologic, meteorologic, reaction kinetics, environmental, waste package, etc., which will be used to synthesize the technical data in a "package" into a simple shopping list for the users of the databases.

3.1.1.2 Geologic and Engineering Materials: Bibliography of Chemical Species

GEMBOCHS is a database and software library containing data and codes for three types of modeling related to prediction of the geochemical environment of the waste package and geologic system. GEMBOCHS supports modeling: (i) thermodynamic behavior of chemical species as a function of temperature and pressure; (ii) equilibrium distribution of elemental mass among coexisting species as a function of temperature, pressure, and bulk composition; and (iii) redistribution of mass among species as a function of imposed chemical and/or thermal disequilibrium.

The GEMBOCHS data and codes reside on a VAX computer system at the Lawrence Livermore National Laboratory (LLNL). An existing agreement between NRC and LLNL results in delivery of updates to the GEMBOCHS data and the EQ3/6 analysis software to the CNWRA. These updates have been made in a variety of ways including electronic on line transfers and mailing magnetic tapes and disks, and are available to the DWM. LLNL is accessible over the Internet and LLNL staff use Internet E-mail to support modifications and requests by participants for additions to the GEMBOCHS system.

3.1.1.3 Automated Technical Data Tracking System

The ATDTS is a means of determining both the data available in the TDB and related information, such as principal investigator, dates entered and other project related data and reports. The ATDTS contains information about the data packages sent to the YMSCPO. The data in the ATDTS are collected on Technical Document Information Forms (TDIFs). A TDIF is completed by principal investigators and submitted with data packages. The TDIF is a multipurpose form that provides information to the ATDTS for submission of information to the TDB and for transfer of data among participants or to outside parties. The ATDTS is also used to track data sent from the Project Technical Database to any requesting site.

3.1.1.4 Quarterly Technical Data Catalog

The TDC is the data product currently produced by the YMSCPO to document the data available in the TDB. This data catalog is produced by the ATDTS from the information on the TDIFs and includes a description of all data acquired or developed by the DOE, acquisition methods, and

location of the data. Entries in the TDC are indexed by a unique DTN that is assigned to each data package. This catalog will be accessible electronically as part of the ATDTS using Internet.

3.1.1.5 Reference Information Base

The Reference Information Base (RIB) contains representations, such as averaged and normalized forms, of some of the data available in the TDB and provides standard values of certain parameters for engineering design and performance assessment calculations. The RIB is currently maintained as a controlled document and distributed by the YMSCPO so that participants may be informed about any future changes in the content or status (e.g., qualification of previously nonqualified data) of information. The major interest for the licensing review process will be in the actual technical data rather than representations, averages, and normalized forms of the data available from the RIB.

3.1.1.6 Central Records Facility

After the technical data from each technical data package is entered into the TDB at the YMSCPO, the data packages are forwarded to the Central Records Facility (CRF). At the CRF, a summary of the contents of each data package is entered and indexed into the Improved Records Information System (IRIS), and the data packages are microfilmed and stored. Access to the actual data package may be requested from the CRF by the accession number assigned to each data package as it is entered into the IRIS.

As data are forwarded from participant sites to the YMSCPO TDB, the ATDTS is updated to indicate that the data now resides in the TDB system. When participant data are forwarded to the TDB, it is also sent to the CRF at the same time. The participant assigns a DTN to each package of technical data forwarded to the YMSCPO. This DTN is sufficient for retrieving data from the TDB but a second index, called the Accession Number, is assigned to each data package by the CRF as it is entered into the IRIS. The accession number assigned to each data package is also recorded in the ATDTS.

3.2 LIST OF ORGANIZATIONS, DATABASES, AND ACCESS METHODS FOR INTERNET

In addition to the Internet procedure established by DOE for NRC to access the ATDTS quoted in Section 2.3, there is a general procedure for requesting data offline for the other databases described in Section 3.1 and listed in Table 3-1. The procedure to be used for requesting data from the TDB or the CRF requires filling out a request form to retrieve data by its DTN. A similar procedure is required for requesting data from the CRF except that both the DTN and the CRF Accession Number are needed.

Table 3-1 provides a list of organizations, databases, and Internet access methods/status for the DWM and CNWRA staff use at the current time.

Table 3-1. List of organizations, databases, and Internet access methods

Organizations	Database	Internet Access Methods/Status
DOE/YMSCPO	GENISES	*TBD. Offline tape and disk transfers only. Assuming Internet access, client/server access is possible, perhaps using NFS.
DOE/YMSCPO	GEMBOCHS	*TBD. Offline tape and disk transfers only from DOE. FTP access from LLNL has been done by CNWRA.
DOE/YMSCPO	ATDTS	Internet Access approved by the DOE.
DOE/YMSCPO	Quarterly TDC	Internet Access approved by the DOE as part of the ATDTS.
DOE/YMSCPO	RIB	*TBD. Report published periodically by the DOE.
DOE/YMSCPO	Information Storage/Retrieval Access Management System (Infostreams/(IRIS)	*TBD. General request procedure available.
DOE/YMSCPO	High Level Waste Technical Data Packages	*TBD. General request procedure available.
USGS	Meteorological Database (PC database), Well Level Records, Southern Great Basin Seismic Network Design, WATSTORE Data, Digital Evaluation Data and Digital Line Graph Data, Digital Geologic Map Database, National Earthquake Information Service Data, Well Log Data, Geophysical (gravity, magnetic, seismic reflection, and refraction data), Satellite Image Data and Aerial Photographs/Video Images	These USGS databases access methods for Internet will be requested by letter and as they are provided will be included in the electronic version of Table 3-1 on Internet for NRC and CNWRA users.
USGS	Earth Science Data Directory	Mosaic via http://info.er.usgs.gov/

Organizations	Database	Internet Access Methods/Status

USGS	GLIS. Contains descriptions of datasets and some data samples. Will hold indices to land processes imagery from EOS.	1) TELNET glis.cr.usgs.gov
		2) X-window interface via TELNET
		  Xglis.cr.usgs.gov
		  Note: Prior to using TELNET, the command "xhost +" must be used to allow the GLS x-window client to connect to the local x-window server
		3) Dial-up 605-594-6888 8N1
		4) Mosaic via http://info.er.usgs.gov/fact-sheet/glis/GLIS.html

DOE/National Laboratory and Nevada Test Site (NTS)	Stratigraphic Cross-Section Data, Sorption Data (SSDMS II), Density/Porosity Data — Yucca Mountain Tuffs, Sorption data, Meteorological Database, Well Level Database	These DOE National Laboratory databases access methods for Internet will be requested by letter and as they are provided will be included in the electronic version of Table 3-1 on Internet for NRC and CNWRA users.
National Weather Service	Meteorological Data and maps (Rainfall, runoff, watershed data)	*TBD
Earth Resources Observation System (EROS)	EROS Data Center searchable index	1) Mosaic via wais://sun1.cr.usgs.gov:210/EROS/Data/Center 2) WAIS (address TBD)

*TBD - To be determined

3.3 CENTER FOR NUCLEAR WASTE REGULATORY ANALYSES EXPERIENCE IN ACCESSING TECHNICAL DATABASES

Examples of CNWRA experience in using data and databases are provided below to indicate the current approaches being used, the difficulty in acquiring good data, and the need for more on line data access to expedite delivery and use of the data. As stated earlier, the primary data needed are in the DOE TDB.

3.3.1 United States Geological Survey Data for the Death Valley Region

Data on wells, springs, and seeps in the Death Valley Region were obtained from the USGSs National Water Information System (NWIS), a distributed database which resides on District office minicomputers. This database consists of a shared site file, the Automated Data Processing System (ADAPS), the Ground-Water Site Inventory System (GWSI), the Quality of WaterSystem (QW), and the

Water-Use Data System (WUSE). In order to obtain data for the entire Death Valley Region in Nevada, separate data retrieval requests had to be made to the Carson City office, which supplied data from all areas outside NTS and the Nellis AFB bombing range. In both cases a formal written request for specific records was submitted to the cognizant data manager. After the retrievals were made, the files containing the data were placed by the data manager in a publicly accessible directory on the Data General Avion workstation at Carson City where they were transferred to the CNWRA by anonymous FTP.

This is not an automatic procedure. However, after details were worked out regarding the cost of performing the database retrievals and the method of payment, the procedure was more straightforward.

3.3.2 Department of Energy and United States Geological Survey Yucca Mountain Region Data for Regional Hydrology Research and Other Projects

Geochemical and hydrologic information from wells and springs in the Yucca Mountain Region are being compiled. This information is being used by the Regional Hydrology Research Project and others. This data is being collected and placed within a spatial framework comprised of topography, political boundaries, lithologic distribution, tectonic features, etc.

Data sources include topographic, tectonic, and political boundaries from the CNWRA Tectonics Research Project, geologic information from the DOE GIS and geochemical information from USGS Open File reports (these data were not available in electronic form and had to be scanned or entered by hand).

Additional data are being sought, including fluid isotopic measurements, fluid chemical compositions that can be traced to specific producing horizons within the rock strata, and long-term (years) variation in water compositions.

3.3.3 Digitized Data from United States Geological Survey and Published Literature for the Western Great Basin

The Volcanic Systems of the Basin and Range Research Project requires construction of a GIS for tectonics and volcanism of the Western Great Basin (WGB), which includes the Yucca Mountain region. The primary source of data for the GIS has been published literature, which has been digitized by the CNWRA for use with the ARC/INFO GIS software. Only a small amount of WGB data exists in electronic formats, and these data are generally at too large a scale to be useful in volcanism research. The electronic data consist of digitized base maps and regional geophysical surveys, primarily from the USGS.

3.3.4 Formal Requests for Department of Energy Data Through the Nuclear Regulatory Commission

Because there are large amounts of volcanism data available in the published literature, computerized databases are required in order to efficiently review DOE prelicensing and licensing documents. These databases also support technical assistance and research at the CNWRA. Currently, DOE data are not available electronically on line. During the recent review of the DOE topical report on Extreme Erosion, original analyses and sample locations were requested from the DOE by the NRC.

Although some paper copies of these data were provided after a few months, the complete data set was received after about six months. Similarly, some locations and stratigraphic designations for geochemical analyses of the Lathrop Wells volcano were requested by the NRC ten months ago and these data have yet to be provided. In both instances, the lack of pertinent data has hindered an effective review of DOE documents. Direct CNWRA Internet access to the appropriate DOE technical databases would help alleviate this problem.

3.3.5 Direct Internet Access to the Yucca Mountain Site Characterization Project Office Technical Databases

Direct Internet access to the YMSCPO TDBs, such as GENISES and GEMBOCHS, and to USGS databases containing topographic, geologic map, geophysical, and aerial imagery data, would be extremely useful to CNWRA staff. Ideally, one would TELNET to the appropriate data catalogs and FTP the data to their systems. Even the ability to FTP data from remote hosts would be a vast improvement, such as in systems made available to most universities and government scientific agencies. The current system of a hard-copy, out-of-date data catalog and written requests through the NRC does not result in the timely receipt of necessary data.

4 OTHER ACCESS METHODS AND DIVISION OF WASTE MANAGEMENT SYSTEM REQUIREMENTS

Most of the selected databases could be accessed using more than one method. Selecting access methods for external databases requires that many factors be considered, including project data needs (type, volume, time periods), security issues, cooperation from remote sites, existing network capabilities, ease of use, frequency of data changes in remote databases, and the ability to procure hardware, software, and communication links for the future.

The implementation scheme for remote database access options other than Internet will be overviewed here to identify potential future system requirements and implementation plans.

4.1 ELECTRONIC MEDIA TRANSFERS

Electronic media transfers have been made and have taken the form of 8 mm DATA cartridges, 9-track magnetic tapes (reels), and IBM/PC formatted floppy disks. GENISES data have been loaded directly into the DWM SPARC10 server from tapes. The GENISES software (ARC/INFO and ARC/VIEW) is installed on Sun servers at both the DWM and CNWRA. Transfers of GEMBOCHS data from VAX VMS systems will probably require that a VAX VMS system be provided to read the electronic media. VMS systems typically use different tape formats than Unix systems. The UNIX "TAR" format is the preferred format for all tape media. IBM/PC disks can be loaded on either PC or Sun workstations. It is expected that this external data would reside on the NRC SPARC10 and that software would be available to allow DWM PCs and workstations to access the data on the Sun.

The additional software that may be required to be completely compatible is the Ingress Database software ($25,000). However, it is possible to use other System Query Language (SQL) compatible databases such as Oracle ($40,000) with some data conversion effort.

4.2 DIAL-IN VIA MODEM

Dial-in access is already practiced by some DWM staff members. Dial-in via modem will probably be implemented by allowing users to run terminal emulation software and access a shared modem pool. Dialing into remote databases can sometimes require that users remember several different things that they must type to successfully retrieve data. This can be made easier by providing hardcopy "how to" and on-line documentation, and by customizing the telecommunications software to simplify the job. All of these options will be considered during the next phase of this activity.

4.3 REMOTE FILE SYSTEM USING A DEDICATED NETWORK

Setting up a Remote File System using a dedicated network line is included here because it is the future access method that the YMSCPO has decided to support for remote GENISES access by participants. This option allows the YMSCPO to tightly control and monitor the data retrieved from their databases and allows them to remain unconnected to outside networks. To implement this access method, a data link must be established between the YMSCPO and the SPARC10 server in the DWM. The required speed on the line has not been determined. This approach is still in the evaluation stages at the YMSCPO. Provisions for paying the monthly leased line fees would need to be arranged.

Access to the YMSCPO GENISES will probably be the most complex undertaking. The establishment of remote access into the YMSCPO GENISES server over a dedicated phone link will require coordination between the YMSCPO, the NRC IRM group, the DWM, two to three different telephone companies, and one or more vendors of network hardware. This option is equivalent to extending the YMSCPO network over phone lines to a single remote server, in this case the GENISES SPARC10 server in the DWM. It is important to understand that the dedicated link is not being used to join two networks. The dedicated phone line is capable of providing full network connectivity and, in fact, this is why such dedicated links are typically used. In this case the YMSCPO plans to limit network traffic over the link to remote, read-only access to specific directories on the GENISES server. The software used would be Sun's NFS.

Once the link is in place, the GENISES software would be loaded onto the SPARC10 server. The exact configuration of the software required to run GENISES will need to be determined so that appropriate software may be acquired and installed. Some of the large static data sets used by GENISES would be requested on 8 mm tape and would also be loaded onto the DWM SPARC10 server. This would allow the DWM SPARC10 server to act as a remote GENISES node and access the actual GENISES data files on the Sun providing file services for the GENISES at the YMSCPO.

The ORACLE and Ingres database software mentioned earlier will also support this operation.

5 CONCLUSION

The DWM and CNWRA have the basic capability for each of the access options discussed in this plan, except for those few items mentioned as future system options in Section 4. Access to Internet sites will require that DWM accounts be established on the NRC firewall computer. Some mechanisms must be put in place to allow users to log into this computer so that they can then perform remote logins to, and file transfers to and from, Internet sites. Sun users have this capability through the FTP and TELNET programs that are standard installations on their Sun workstations. PC users will log in to the firewall using TCP/IP network drivers to allow FTP and TELNET access. Remote file access using dedicated links will require compatible software on both DWM and DOE systems in the future, with database management and application specific software being the primary additional items.

In conclusion, Table 3-1 will continue to be updated on line for Internet access methods by the DWM and CNWRA. Appropriate support is planned by DOE for access to their complete technical database. This will permit the requirements for technical database users to be met.

6 REFERENCES

Johnson, R.D. 1992. *Department Of Waste Management Advanced Computer Review System Design Summary and Proposed On-Going Support Tasks*. San Antonio, TX: Center for Nuclear Waste Regulatory Analyses.

Harloe, Ed. 1993. *External Database Access Options Report*. San Antonio, TX: Center for Nuclear Waste Regulatory Analyses.

Nuclear Regulatory Commission. 1993. Revised Procedural and Project-Specific Agreements Between NRC and DOE for the High-Level Waste Repository. Washington, DC: Nuclear Regulatory Commission.

Russell, G.T., and A. Gangemi, Sr. 1991. *Computer Security Basics*. Sebastopol, CA: O'Reilly and Associates, Inc.