External Database Access Options Report

Prepared for

Nuclear Regulatory Commission
Contract NRC-02-88-005

Prepared by

Edward Harloe

Center for Nuclear Waste Regulatory Analyses
San Antonio, Texas

June 1993

CONTENTS

TABLES

ACKNOWLEDGMENTS

1 INTRODUCTION

1.1 BACKGROUND

The CNWRA has participated directly in the analysis, design and implementation tasks of the Division of High-Level Waste Management (DHLWM) Advanced Computer Review System (ACRS) since March, 1992. The "DHLWM Advance Computer Review System Design Summary and Proposed On-Going Support Tasks" report (Johnson 1992), recommended several ongoing activities to compliment and support the ACRS that had been designed. The proposed activities were grouped into three primary development tasks and two support tasks. These tasks were discussed with DHLWM management and in a management meeting on October 18, 1992, the decision was made to pursue the two highest priority development tasks.

The CNWRA has initiated work on both tasks and this is the first report on the task to provide the DHLWM with "Access to External Technical Databases." The Procedural Agreement between the U.S. Nuclear Regulatory Commission (NRC) and Department of Energy (DOE) (U.S. NRC. 1992), as implemented, provides NRC read-only access to the DOE databases and the CNWRA is ascertaining the proper computer protocols for accessing selected technical databases, the data format, and controlling software for each database. For the selected set of databases, the CNWRA will establish a working connection between the DHLWM ACRS and the remote DOE database and will demonstrate reading, transmitting, and importing of a data file into a DHLWM analysis program. A similar such demonstration of capability will be conducted for selected other remote databases at agencies such as the U.S. Geological Survey (USGS). The CNWRA will provide an estimate of life cycle costs for maintaining and supporting access to the four selected external databases.

1.2 PURPOSE AND SCOPE

The purpose of this report is to identify external databases of interest to the DHLWM and to present options for implementing access to these databases. The following "External Databases" section (Section 1.3) lists databases identified as being of possible interest to the DHLWM and discusses methods for acquiring data from these databases for use by the NRC. The list is not comprehensive. Other useful databases will very likely be identified in the future.

Section 2 discusses the "Current Communications Infrastructure" of the sites involved. This section also discusses existing computer systems and networks. This background is necessary to help select the most appropriate access methods to develop.

The "Selected External Databases" section (Section 3) specifies which external databases have been selected as targets for the DHLWM to access and gives the criteria used to make the selection. These databases are (i) the Geographic Nodal Information Study and Evaluation System (GENISES) portion of the Yucca Mountain Project Office (YMPO) Technical Data Base (TDB), (ii) the YMPO Automated Technical Data Tracking (ATDT) system, (iii) a Participant's Database, such as the USGS, and (iv) Internet public data archives.

Section 4 discusses "Access Methods Options" for the selected databases and how various methods might be utilized to access additional databases in the future. The section also discusses the pros and cons of each type of access. Access methods discussed include modem dial-up, media transfer, and

several types of direct network access using Transmission Control Protocol/Internet Protocol (TCP/IP). The access methods in the final section will be discussed with the DHLWM and a consensus will be reached on which access method should be pursued for each selected database.

The next report is the activity will provide details about the hardware, software, and access procedures required to implement the selected access methods for staff using the ACRS in the DHLWM.

1.3   EXTERNAL DATABASES

1.3.1   Yucca Mountain Project Office

Databases at the YMPO 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.

1.3.1.1   Technical DataBase

The TDB contains the technical data stored and maintained by the Yucca Mountain Project staff that has been extracted from Technical Data Packages sent from participant sites. The TDB consists of two components, GENISES and Geologic and Engineering Materials: Bibliography of Chemical Species (GEMBOCHS).

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 YMPO, 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 about 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"[1] data and is currently estimated to be about 500 megabytes in size. The geographical support data is referred to as "Type 2"[2] data and is currently estimated to be about 2 gigabytes in size.

Due to the amount and variety of data contained in the GENISES, this system will probably be the focus of much of the effort to access technical data for use on the ACRS SparcStation 10 server in the DHLWM.

---

[1] "Type 1" data is the data extracted form Technical Data Packages sent to the YMPO. This data is subject to revision at any time.

[2] "Type 2" data is used to support the display of "Type 1" data by the GENISES software. This data is expected to be relatively static and includes data on fault locations, map borders, road locations, well locations, etc.

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 the: (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 this mass among these 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 the NRC and LLNL results in delivery of updates of the GEMBOCHS data and the EQ3/6 analysis software to the CNWRA in San antonio, Texas. These updates are currently done on magnetic tape. The 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.

1.3.1.2 Automated Technical Data Tracking System

The Automated Technical Data Tracking (ATDT) system could be very useful to the NRC as a means for determining both the data available in the TDB and related information, such as Principal Investigator, dates needed, and project related data and reports. The ATDT system contains data about the data packages sent to the YMPO. The data in the ATDT system 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 for the ATDT, for submission of information to the TDB and for transfer of data among participants or to outside parties.

1.3.1.3 Quarterly Technical Data Catalog

The Technical Data Catalog is the data product currently produced by the YMPO to document the data available in the TDB. This data catalog is produced by the ATDT system from the information on the TDIFs submitted and includes a description of all data acquired or developed by the DOE, acquisition methods, and location of the data. Entries in the Technical Data Catalog are indexed by a unique Data Tracking Number (DTN) that is assigned to each data package. This catalog is not currently available electronically.

1.3.1.4 Yucca Mountain Project Office Parameter Dictionary

As technical data packages arrive at the YMPO, parts of the technical data in these packages are placed into the TDB, typically into the GENISES data system. In order to control these technical data transfers and associate the technical data and programmatic information and documents, a parameter dictionary is being developed and is expected to be completed in the summer of 1993. 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 database.

1.3.1.5 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 YMPO so that participants may be informed about any future changes to the content or status of information in the database.

1.3.1.6 Central Records Facility

After the technical data from each technical data package is entered into the TDB at the YMPO, the data packages are forwarded to the Central Records Facility. At the Central Records Facility 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 Central Records Facility by the accession number assigned to each data package as it is entered into the IRIS.

1.3.2 Participant Data Archives and Local Records Facilities

The possibility of gathering data directly from one participant site will be determined in future activities of this task. It is known that organizational units of USGS and the National Laboratories participating in the Yucca Mountain Project are accessible through the Internet.

1.3.3 Public Data Archives on Internet

Other sources of data available in public archives accessible over the Internet may also be of value to the NRC. Options include direct access from DHLWM computers (if permitted by the NRC Office of Information Resource Management [IRM]) and access through the NRC "firewall" computer. [3] The "firewall" computer is a system set up for providing this type of access while providing increased security for the NRC network by effectively isolating other systems from the Internet.

---

[3]NRC "firewall" computer is a planned system that may be installed as a security-measure to isolate the NRC servers and workstations containing data from the Internet.

2 CURRENT COMMUNICATIONS INFRASTRUCTURE

The following sections give a brief overview of the current computer and communication infrastructure at the DHLWM, the CNWRA, and the YMPO. This information provides the necessary background for the discussion of access method options in the next section

2.1   NRC DIVISION OF HIGH-LEVEL WASTE MANAGEMENT

The DHLWM users have either IBM Compatible Personal Computers (PCs) running DOS and/or Windows, or Sun workstations running SunOS. A Silicon Graphics workstation will be added in the near future. The computers are attached to a token ring local area network (LAN). 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 multi-protocol routers to form the Agency Wide Network (AWN).

All connections to the Internet at the NRC must go through a "firewall" computer. 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 IRM by reconfiguring the wide area router used to connect the NRC with the CNWRA.

2.2   CENTER FOR NUCLEAR WASTE REGULATORY ANALYSES

Most of the CNWRA's technical staff have PCs running OS/2, Sun workstations, or Macintosh computers, as well as access to a Silicon Graphics computer. The CNWRA uses both token ring and Ethernet networks. Both network types are connected to the wide area router connecting the CNWRA with the NRC's AWN, but the router connection is configured to only allow communications from specific computers to pass over the link. CNWRA users also have access to the Southwest Research Institute network and, through it, to the Internet. VAX and Cray computers are accessed over this network for running codes such as the EQ3/6 code and others for specific technical disciplines.

2.3   YUCCA MOUNTAIN PROJECT OFFICE

The YMPO uses Sun workstations and PCs running DOS and Windows, as well as a Silicon Graphics computer. The PCs use TCP/IP software from Sun Microsystems to access file systems on the Sun file server using the Network File System (NFS) protocol. The GEMBOCHS database is maintained on a VAX computer. The YMPO uses Ethernet as the LAN and currently does not support TDB connections to outside networks.

3 SELECTED DATABASES

Three interrelated considerations lead to the selection of the four databases targeted for access by the DHLWM. These considerations are:

(i) The utility of the data in the database

(ii) The need to include the DOE, a participant's local data archive, and public data archives accessible over the Internet in the selected set of databases in order to establish access to one database of each of these three classes of databases

(iii) The fact that establishing access to some databases would provide an access mechanism that could be used to access additional databases of the same type in the future

The role that each of these considerations played in the selection of the target databases is explained in the following sections.

3.1 YUCCA MOUNTAIN PROJECT OFFICE GENISES

GENISES was selected because of the large variety of data it contains. As site characterization data is forwarded to the YMPO from participant sites, it will be entered into the GENISES, which will be used to control the distribution of this data (Newbury, 1992). YMPO plans to provide online access to the site characterization data centered around providing remote access to GENISES. Access to this database system will be complicated by the size of the database and the fact that the YMPO does not support connectivity to external networks at this time. The important factors in selecting the GENISES as a target for access are the utility of the data it contains and the fact that it is a major DOE data collection center.

3.2 YUCCA MOUNTAIN PROJECT OFFICE AUTOMATED TECHNICAL DATA TRACKING SYSTEM

The ATDT system is the only data system at the YMPO that currently supports any type of remote access. Participant sites may dial in to the ATDT system with a modem to enter or view TDIFs. The ATDT system provides information about data stored in the GENISES and information about data located in participant data archives. Gaining access to the ATDT system might also serve as a first step towards using the existing dial-in facilities for other purposes.

The ATDT system provides data about the site characterization data being acquired or developed. This includes data about the Principle Investigator, citation information, governing program plan, test and sample numbers, and data submission and transfer data. When the License Support System (LSS) becomes available it is expected that the ATDT system would provide the description data for the header of each Technical Data Package entered into the LSS (Statler, 1992).

3.3 PARTICIPANT DATABASE SYSTEM

Specific technical data may be available at participant data sites before it is available through the YMPO and may be useful to the NRC. A specific site has not yet been identified, but staff from both

the DHLWM and the CNWRA have expressed the desire to access data from the USGS and the National Laboratories.

3.4   INTERNET PUBLIC DATA ARCHIVES

This fourth database is not really a single database but has been selected because implementation of the access method to get a public data archive on the Internet is the same for many data archives. A list of some of the Internet data repositories of possible interest to the DHLWM will be compiled and provided as a reference when this access method is implemented.

4 ACCESS METHODS OPTIONS

Most of the selected databases could be accessed using more than one access method. Selecting access methods for external databases requires that many factors be weighed against one another. Factors that must be considered include 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. Table 4-1 summarizes access options that might be implemented for each selected remote database. The following sections discuss the access of each database in more detail.

Table 4-1. Database Systems and Access Methods

Databases	Methods			
	Electronic Media Transfer	**Dial-in Via Modem**	**Direct Network Access* Using TCP/IP**	**Remote File System on Dedicated Link**
GENISES	Supported. High capacity (2 Gbyte plus) 8 mm DAT tape cartridge.	Dial-in to ATDT at same site supported. No dial-in to GENISES supported.	No, not linked to any outside network at this time.	Planned. This is the YMPO planned approach for making GENISES available to remote sites.
ATDT	None. Distributed in hardcopy form.	Supported for participants.	No, not linked to any outside network at this time.	Not planned but link for GENISES might also support access to ATDT system.
Participant Database	By arrangement. CNWRA has received data on 9-track mag tape from Participants.	Unknown but likely.	Yes. Connected to Internet. Should require only account on machine with database.	Yes. Probably unlikely. Requires considerable work at both ends.
Internet Public Data Archives	None.	Very rare.	Yes. Usually the only access method available.	No.

* Direct Network Access may require users to first log on to an NRC "firewall" computer for reasons related to network security. This means that DHLWM users may see this Direct Network Access as a two-step process, Remote Site to and from "firewall" computer and "firewall" computer to and from DHLWM computers.

4.1 GENISES ACCESS METHODS

The advantages and disadvantages of each access method relative to the GENISES are summarized in Table 4-2. The Electronic Media Transfer option is supported by the YMPO. In this

option a remote site would install the required Ingres, ARC/INFO, and ARC/VIEW software on a remote computer and then load GENISES data onto that computer from 8 mm tapes requested from the YMPO. The standard procedure of submitting a work request form to the YMPO would be used to request updated data as desired.

Table 4-2. GENISES Access Methods

GENISES Access Methods	Electronic Media Transfer	Dial-in Via Modem	Direct* Network Access Using TCP/IP	Remote File System on Dedicated Link
PRO	Probably quickest way to get TDB data onto DHLWM computers.	Dial-in line exists at YMPO. Would need account and Ingres knowledge to access data.	Could allow rapid transfer of very recent data.	Approach planned by YMPO for remote access. Allows access to all data without YMPO linking to outside network.
CON	Requires agreement for tape production and some delay in preparing, shipping, and then extracting data from tapes. Requires work to load data from tapes and set up for GENISES access each time an update tape is received.	Not currently supported by YMPO. Very slow file transfers.	Requires YMPO to connect to an outside network. This is not planned. User must produce data files and manually transfer them back to the user's own system.	Requires dedicated link between YMPO and remote access site. Link will only support GENISES.

The dial-in option could be implemented with existing hardware and software if the YMPO was willing to allow use of the ATDT dial-in lines for remote log-in to the GENISES file server. Since the GENISES software will not run over this type of link, the remote user would need to issue Ingres database commands to extract the desired data and then use a file transfer protocol to transfer these files to the remote computer. The file transfers would be slow, probably having a throughput rate below 9600 bits per second.

The GENISES is a combination of the Ingres database system, the ARC/INFO Geographical Information system, and the ARC/VIEW graphical viewing software for the ARC/INFO system. The YMPO is currently not supporting outside networks. The plan for making the GENISES available to participant data sites involves a dedicated network link to the YMPO from each remote access site. A PC or Sun workstation at the remote site would run the Ingres, ARC/INFO, and ARC/VIEW software while accessing data and files from the file server at the YMPO over the dedicated link using NFS software. Large data files that rarely change would also be distributed to the remote sites so that these

large files could be accessed at the remote site instead of over the link from the YMPO. The network links are called "dedicated" because they only support this remote access to selected YMPO file systems. Such links would not truly connect networks at each end. This means that the YMPO network would be isolated from the networks at each remote GENISES access site. The standard TCP/IP File Transfer Program (FTP) and remote log in program (TELNET) would not be supported by these links. This option is promising because it is supported by the YMPO and has the potential to make GENISES data available on DHLWM computers as soon as it is online at the YMPO.

The Direct Network Access using TCP/IP protocols option is not really an option for GENISES access unless the YMPO links their network to the Internet. There are no current plans to implement such a link.

The most promising options for accessing data from the GENISES database are Electronic Media Transfer and setting up a Remote File System on a dedicated link. Both of these options require that at least one computer at the remote site be loaded with the required Ingres, ARC/INFO, and ARC/VIEW software. The remote site would then be able to use data requested from the YMPO on tape. The addition of a dedicated link to the YMPO would mean that the remote GENISES user could access actual current GENISES data and would always be using the latest data.

4.2 AUTOMATED TECHNICAL DATA TRACKING SYSTEM ACCESS METHODS

The advantages and disadvantages of each access method relative to the ATDT system are summarized in Table 4-3.

Table 4-3. ATDT Access Methods

ATDT Access Methods	Electronic Media Transfer	Dial-in Via Modem	Direct Network Access* Using TCP/IP	Remote File System on Dedicated Link
PRO	The Technical Data Catalog lists ATDT contents. Currently distributed in hardcopy form. Electronic form would be easier to search.	Currently supported for entry of TDIFs into ATDT system. Would allow remote users to check on availability of expected data in the GENISES.	Could provide access to the same software as direct dial-in but be faster, cheaper, and easier to use.	None. Not a good option for this database.
CON	Not currently supported by YMPO. Would still be quarterly updates.	System was designed for TDIF entry. Might be of limited use for finding new data.	Not supported. YMPO has no plans to link to Internet.	Expensive way to access low volume data files. Would require remote copies of ATDT access software.

The Electronic Media Transfer option could take two forms for the database. The raw database files could be transferred, or the quarterly Technical Data Catalog produced from the ATDT system, which is currently distributed in hardcopy form, could be distributed on electronic media.

The dial-in option is currently supported and would be a first step towards electronic data exchange between the NRC and the YMPO. The current system is designed to support participant data sites in entering TDIFs into the ATDT system. It might not support look-up and browsing functions very well.

The Direct Network Access using TCP/IP option is not really an option unless the YMPO links to the Internet.

The most promising options for accessing the ATDT system are dial-in via modem and electronic media transfer, probably of the quarterly Technical Data Catalog.

4.3   PARTICIPANT DATABASE

The advantages and disadvantages of each access method relative to accessing participant database systems is summarized in Table 4-4.

Table 4-4. Participant Database Access Methods

Participant Database Access Methods	Electronic Media Transfer	Dial-in Via Mmodem	Direct Network Access* Using TCP/IP	Remote File System on Dedicated Link
PRO	Simple and quick if support can be found to transfer data onto tape for shipment.	Allows remote user to select data. Facilities probably in place. This must be verified for selected database.	Allows remote user to select data. Facilities probably in place. Faster and easier to use than dial-in.	Would allow remote users to directly access all files in remote file system.
CON	Requires support from lab that already must prepare, document, and ship its data to YMPO.	If modem access is allowed, TCP/IP would probably be allowed also. Modem transfer is very slow.	Will require user to extract data into files to be transferred back to remote site. Requires users go through NRC "firewall" system.	Expensive way to access data probably accessible over Internet. Would require remote copies of any access software normally run on remote system.

The Electronic Media Transfer option has been used by the CNWRA to receive data from participants in the past. This method would be best suited for large data sets that do not change rapidly.

The dial-in option and the TCP/IP option would provide similar functionality but the TCP/IP option would allow faster file transfers. Both options would entail remote users logging on to a computer at the participant's office and either entering database queries or running database access software. Data

could be captured as it scrolled across the screen, or the data could be extracted into files and then transferred back to the remote user's computer.

The Remote File System option would allow remote users to directly access a set of files on a participant computer. This option requires a dedicated network link and system configuration modification and support on each end of the link.

The most promising options for accessing data from a selected participant database are using TCP/IP access, Electronic Media Transfer, or dial-in access.

4.4 INTERNET PUBLIC DATA ARCHIVES ACCESS METHODS

The only access method available for retrieving data from public Internet archives is using the TCP/IP protocols over the Internet. This is shown in Table 4-5.

Table 4-5. Internet Public Archives Access Methods

Internet Public Data Archives Access Methods	Electronic Media Transfer	Dial-in Via Modem	Direct Network Access* Ising TCP/IP	Remote File System on Dedicated Link
PRO	None.	None.	The only access option available but these archives were built with this access method in mind. Online documentation is provided to help remote users locate and retrieve the data in each archive.	None.
CON	Not supported.	Not supported.	Public archives can be unavailable or become so busy that network connections are refused.	Not supported.

4.5 IMPLEMENTATION OF ACCESS METHODS

The implementation scheme for each of the remote database access options listed here will be presented in detail in the next report in this activity. This section outlines the basic form envisioned for each type of access.

Electronic Media Transfer will probably take the form of 8 mm DAT cartridges, 9-track magnetic tapes (reels), and IBM/PC formatted floppy disks. GENISES data would arrive on 8 mm

cartridges for loading directly into the DHLWM SparcStation 10 server. The GENISES software (Ingres, ARC/INFO, and ARC/VIEW) would also be installed on this Sun computer. The Unix "TAR" format would be the preferred tape format for all tape media. IBM/PC disks could be loaded on either PC or Sun workstations. It is expected that this external data would reside on the NRC SparcStation 10 and that software would be available to allow DHLWM PCs and workstations to access the data on the Sun.

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 from the site being dialed into. This job 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.

Direct Network Access using TCP/IP will probably require DHLWM users to first log into an NRC "firewall" computer. The users would then run the TELNET program to log into remote systems. The types of terminal emulation provided by TELNET on the "firewall" computer will need to be investigated. One a user is logged onto a remote computer, data transfer will typically take place in one of two ways. The first method involves the user, often using a database access screen, searching for and displaying data of interest. Most TELNET implementations allow the data being displayed to also be stored as a file. This is a simple mechanism that works well for small amounts of data. Larger transfers would more likely be performed by using software on the remote computer to generate a file containing the data to be transferred. This file could then be transferred back to the NRC "firewall" computer with the FTP program. This method is similar to the dial-in method in that remote terminal emulation is being used. As with the dial-in option, methods of making the job easier for users must be considered.

Setting up a Remote File System using a dedicated network line is included here because it is the access method that the YMPO has decided to support for remote GENISES access by participants. This option allows the YMPO 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 YMPO and the SparcStation 10 server in the DHLWM. The requires speed on the line has not been determined. This entire approach is still in the planning stages at the YMPO. Provisions for paying the monthly leased line fees would need to be arranged. Once the link is in place, the GENISES software would be loaded onto the SparcStation 10 server. Some of the large static data sets used by the GENISES would be requested on 8 mm tape and also loaded onto the Sun. This would allow the Sun 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 YMPO.

5 CONCLUSION

The survey of external database systems and external database access options presented here is intended to serve as a basis for discussion between the CNWRA and the DHLWM. The goal of these discussions is to determine which access method should be attempted for each of the four (4) selected database systems. The hardware and software required to implement the selected access methods will be discussed in the next report for this activity.

The requirement for access to Internet sites will require that 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 come standard on their Sun workstations. PC and Macintosh users will need software to allow them to login to the firewall system. This could take the form of either TCP/IP network drivers to allow FTP and TELNET or the form of commercial telecommunications software such as CrossTalk, Smart Term, or Procomm Plus.

Media transfer on 8 mm DAT tape from Unix systems can be handled on the SparcStation 10 in the DHLWM. Transfers on 9-track tapes would require additional hardware or the use of hardware outside of the DHLWM. Transfers of data from VAX VMS systems will probably require that a VAX VMS system be located to read the electronic media. VMS systems typically use different tape formats than Unix systems.

Dial-in access is already practiced by some DHLWM staff members. If the amount of dial-in access is expected to increase it may be necessary to examine the dial-out capabilities within the DHLWM. For heavy use it may be necessary to upgrade the number of dial-out lines available, the speed of the modems used for dial-out, or the software used during dial-out sessions.

Access to the YMPO GENISES will probably be the most complex undertaking in this activity. The establishment of remote access into the YMPO GENISES server file system over a dedicated phone link will require coordination between the YMPO, NRC's IRM group, the DHLWM, 2 to 3 different telephone companies, and one or more vendors of network hardware. This option basically amounts to extending the YMPO network over phone lines to a single remote workstation, in this case the GENISES computer in the DHLWM. It is important to understand that the dedicated link is not being used to join two network here. The dedicated phone line is capable of providing full network connectivity and, in fact, this is what such dedicated links are typically used for. In this case the YMPO plans to limit network traffic over the link to remote, read only, access to specific directories on the GENISES file server. The software used would be Suns's Network File System (NFS).

Regardless of the establishment of a remote file system over a dedicated phone line or the use of electronic media to transfer GENISES data from the YMPO to the DHLWM, a Sun computer in the DHLWM will need to be running the GENISES support software in order to use the data once it is accessible. It is known that the basic software packages include ARC/INFO, ARC/VIEW, and the Ingress database system. The exact configuration of the software required to run GENISES will need to be determined so that the appropriate software may be acquired and installed.

6 REFERENCES

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

Newbury, C.M. 1992. Management of Scientific and Engineering Data Collected During Site Characterization of a Potential High Level Waste Repository. *Proceedings, Third International High-Level Waste Conference.* Las Vegas, NV. Washington, DC: U.S. Department of Energy.

Statler, J. 1992. The Flow to Licensing: Technical Data Tracking and the Licensing Support System (LSS). *Proceedings, Third International High-Level Waste Conference.* Las Vegas, NV. La Grange, IL: American Nuclear Society.

U.S. Nuclear Regulatory Commission. 1993. *Revised Procedural And Project-Specific Agreements Between NRC And DOE For The High-Level Waste Repository.* Washington, DC: U.S. Nuclear Regulatory Commission.