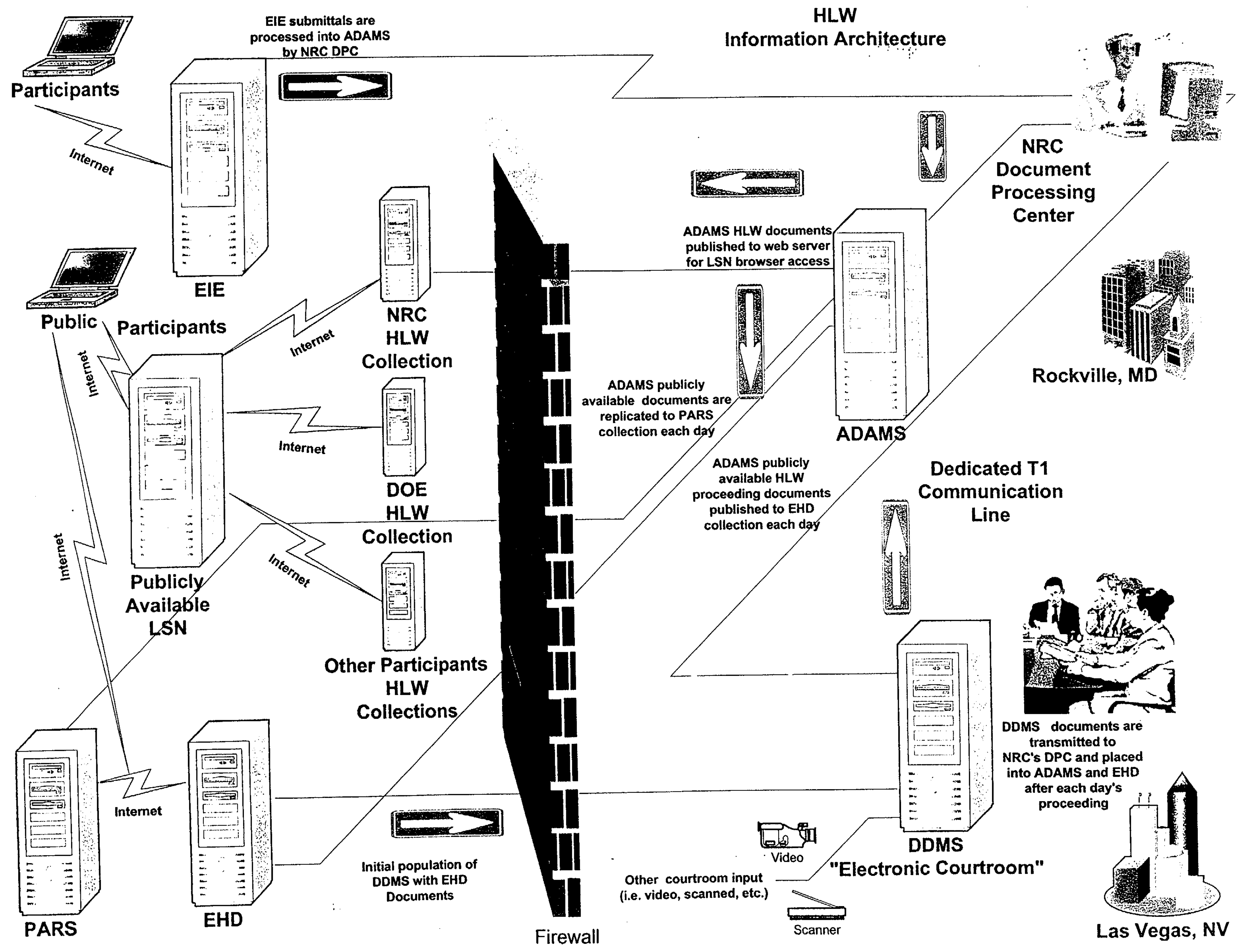
ATTACHMENT 3

MEETING HANDOUTS


HLW
Information Architecture
Participants
EIE submittals are processed into ADAMS by NRC DPC
Internet
EIE
Public
Participants
Internet
NRC HLW Collection
Internet
DOE HLW Collection
Internet
Other Participants HLW Collections
Internet
Publicly Available LSN
Internet
PARS
EHD
Initial population of DDMS with EHD Documents
Firewall
ADAMS HLW documents published to web server for LSN browser access
ADAMS publicly available documents are replicated to PARS collection each day
ADAMS publicly available HLW proceeding documents published to EHD collection each day
ADAMS
NRC Document Processing Center
Rockville, MD
Dedicated T1 Communication Line
DDMS documents are transmitted to NRC's DPC and placed into ADAMS and EHD after each day's proceeding
Video
Other courtroom input (i.e. video, scanned, etc.)
Scanner
DDMS
"Electronic Courtroom"
Las Vegas, NV

NRC/DOE Meeting on Electronic Submissions

June 25-26, 2002

Agenda

- **Introduction - NRC**
- High-Level Waste (HLW) Information Architecture - NRC
- Use of NRC's Electronic Information Exchange (EIE) - NRC
- Electronic Hearing Docket (EHD) Document Format - NRC
- Electronic Courtroom - NRC
- Technical Issues with a Potential Electronic License Application (LA) Submission and Other Large Documents - NRC
- Status of DOE's Effort for Putting Its Documentary Collection Out on the Licensing Support Network (LSN) - DOE
- Status of NRC's Effort for Putting Its Documentary Collection Out on the Licensing Support Network (LSN) - NRC
- Status of Search and Retrieval Testing - DOE
- Status of NRC LSN Test Server - NRC
- Meeting Summary – NRC

June 25-26, 2002 Page 2

NRC/DOE Meeting on Electronic Submissions

Janet Schlueter
High Level Waste Branch
Division of Waste Management
(301) 415-7264
JRS1@NRC.GOV

June 25-26, 2002 Page 3

Agenda

- Introduction - NRC
- **High-Level Waste (HLW) Informat on Architecture - NRC**
- Use of NRC's Electronic Information Exchange (EIE) - NRC
- Electronic Hearing Docket (EHD) Document Format - NRC
- Electronic Courtroom - NRC
- Technical Issues with a Potential Electronic License Application (LA) Submission and Other Large Documents - NRC
- Status of DOE's Effort for Putting Its Documentary Collection Out on the Licensing Support Network (LSN) - DOE
- Status of NRC's Effort for Putting Its Documentary Collection Out on the Licensing Support Network (LSN) - NRC
- Status of Search and Retrieval Testing - DOE
- Status of NRC LSN Test Server - NRC
- Meeting Summary – NRC

June 25-26, 2002 Page 4

HLW Information Architecture

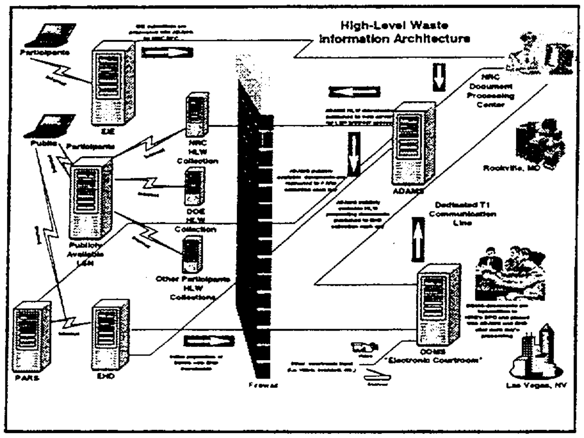
Wil Madison
Office of the Chief Information Officer
U.S. Nuclear Regulatory Commission
(301) 415-7221
WLM@NRC.GOV

June 25-26, 2002 Page 5

Purpose

- Provides a conceptual overview of the various information collections that will be used to support the high-level waste (HLW) proceedings
- Other presenters will focus on specific components, processes, and procedures for:
  - Electronic Information Exchange (EIE)
  - Electronic Hearing Docket (EHD)
  - Digital Data Management System (DDMS)
  - Licensing Support Network (LSN)

June 25-26, 2002 Page 6



Glossary of Terms

- ADAMS - Agencywide Documents Access and Management System
  Electronic information system that maintains NRC's unclassified official program and administrative records in a centralized electronic document repository.
- DDMS - Digital Data Management System
  Hearing room with digital information retrieval, utilization, and display capabilities to conduct a major portion of the HLW repository licensing proceeding. It permits the creation and use of an integrated, comprehensive digital record for the proceeding.
- DPC - Document Processing Center
  Focal point for entry of hard copy and electronic documents and related materials that originate outside the Agency into ADAMS.
- EHD - Electronic Hearing Docket
  Official hearing docket of the NRC for the proceeding on the application of DOE to store high level radioactive waste at a designated storage facility.

June 25-26, 2002 Page 8

Glossary of Terms (cont.)

- EIE - Electronic Information Exchange
  Allows NRC to exchange material related to official agency business with its customers and other Federal agencies across the Internet. The EIE system uses public key infrastructure and digital signaturing technology to authenticate documents and validate the person submitting the information.
- HLW – High-Level Waste
  High-level radioactive wastes are the highly radioactive materials produced as a byproduct of the reactions that occur inside nuclear reactors.
- LSN - Licensing Support Network
  Provides a single place where the partie, governmental participants, and potential parties to the licensing hearing can search for documents from any/all of participant collections in a uniform way.
- PARS - Publicly Available Records System
  Public version of ADAMS, an information system that provides access to all documents made public by the NRC since November 1, 1999. Several hundred new documents are added daily. Permits full text searching and the ability to view document images, download files, and print locally.

June 25-26, 2002 Page 9

Agenda

- Introduction - NRC
- High-Level Waste (HLW) Informatic ı Architecture - NRC
- **Use of NRC's Electronic Information Exchange (EIE) - NRC**
- Electronic Hearing Docket (EHD) Document Format - NRC
- Electronic Courtroom - NRC
- Technical Issues with a Potential Electronic License Application (LA) Submission and Other Large Documents - NRC
- Status of DOE's Effort for Putting Its Documentary Collection Out on the Licensing Support Network (LSN) - DOE
- Status of NRC's Effort for Putting Its Documentary Collection Out on the Licensing Support Network (LSN) - NRC
- Status of Search and Retrieval Testing - DOE
- Status of NRC LSN Test Server - NRC
- Meeting Summary – NRC

June 25-26, 2002 Page 10

Electronic Info. Exchange (EIE) Process and Pilots

John Skoczlas
Office of the Chief Information Officer
(301) 415-7186
JAS1@NRC.GOV

June 25-26, 2002 Page 11

Presentation Outline

- EIE System
  - Definitions
  - Regulatory Issue Summary
  - Formats
  - Parameters
- Future Rulemaking
  - Formats
  - Parameters
  - Pilot Projects
- Other Information
  - Digital certificates
  - Digital signatures
  - System demonstration
  - DOE participation in EIE

June 25-26, 2002 Page 12

Definitions

- Electronic Information Exchange (EIE)
  - Allows NRC to exchange material related to official agency business with its customers and other Federal agencies across the Internet. The EIE system uses public key infrastructure (PKI) and digital signaturing technology to authenticate documents and validate the person submitting the information.

Definitions (cont.)

- Portable Document Format (PDF)
  - Adobe® PDF is a universal file format that preserves all the fonts, formatting, graphics, and color of any source document, regardless of the application and platform used to create it
- Tagged Image File Format (TIFF)
  - A tag-based image format. TIFF is designed to promote universal interchanges of digital images.
- Public Key Infrastructure (PKI)
  - The architecture, organization, techniques, practices, and procedures that collectively support the implementation and operation of a certificate-based public key cryptographic system.

Definitions (cont.)

- Digital Signature
  - A transformation of a message using an asymmetric cryptosystem such that a person having the initial message and the signer's public key can accurately determine whether the transformation was created using the private key that corresponds to the signer's public key and whether the message has been altered since the transformation was made
- Digital Certificate
  - An electronic verification that allows for PKI use. The certificate has security requirements that can not be compromised by other than fraud.

Regulatory Issue Summary

- EIE implemented by Regulatory Issue Summary (RIS) 2001-05, "Guidance on Submitting Documents to the NRC by Electronic Information Exchange or on CD-ROM," dated January 28, 2001
- The RIS:
  - Allows for voluntary electronic transmission for Part 50 submittals
  - Defines the current EIE system and alternate use of CD-ROM
  - Gives guidance to licensees submitting regulatory documents electronically
  - Does not apply to submittals in NRC adjudicatory proceedings

EIE System (cont.)

- Files limited to 15 megabytes (Mb) per transmission
- Allows option to submit larger documents (over 15 Mb) in CD-ROM
- No paper copy need accompany EIE or CD-ROM submittals
- Uses digital certificates and signatures
- Expanded to include some Part 40 and 70 submittals

EIE Accepted Formats

- Adobe® PDF
- Corel® WordPerfect
- Microsoft® Word
- ASCII
- Multi-page TIFF
- Corel® Presentations
- Corel® Quattro Pro
- Microsoft® Excel
- Microsoft® PowerPoint

EIE Preferred Formats

- PDF Image+Text
- PDF Normal
- PDF Image
- Multi-page TIFF

June 25-26, 2002 Page 19

EIE Parameters

- No more than 15 Mb per submittal
- No:
  - Classified material
  - Safeguards information
  - Privacy Act information
  - Other non-public documents
- Only acceptable browsers (see NRC web site for versions) are:
  - Netscape
  - Internet Explorer

June 25-26, 2002 Page 20

E-Rule and Guidance (Next EIE Version)

- Consistent with requirement to reduce government paperwork (Government Paperwork Elimination Act of 1998):
  - NRC is considering revising its regulations to permit voluntary electronic submissions
  - NRC may issue guidance addressing details of electronic submission
  - NRC may revise its regulations to allow electronic submission of documents in all NRC adjudications (10 CFR 2.108 requires paper filings)

June 25-26, 2002 Page 21

Future Rulemaking

- May permit voluntary electronic submissions by all licensees, vendors, and applicants by EIE, E-mail, CD-ROM, or diskettes
- May include guidance addressing details of electronic submissions
- May allow electronic submission of documents in all NRC adjudications

June 25-26, 2002 Page 22

Future EIE Formats

- PDF Image+Text
- PDF Normal
- PDF Image
- Multi-page TIFF

June 25-26, 2002 Page 23

Future EIE Parameters

- May increase submittal size
- May exclude:
  - Restricted Data
  - National Security Information
  - Privacy Act information
  - Proprietary information
  - Other non-public information
- One paper copy to be submitted with a CD-ROM

June 25-26, 2002 Page 24

EIE Pilot Projects

- Adjudicatory Pilot
  - Will test the submittal, distribution, and service of electronic hearing documents
  - Provides for the submittal and transmission of documents to all or selected participants
  - Provides service verification record
  - Access only to those who are approved participants of a particular hearing

June 25-26, 2002 Page 25

EIE Pilot Projects (cont.)

- Criminal History File Pilot
  - Will test submittal and transmission of criminal history file information
  - Will provide for certificate level security and point-to-point encryption
  - Will transmit scanned fingerprint images as well as text material
  - Will test files larger than 25 Mb

June 25-26, 2002 Page 26

Digital Certificates

- Available online
- Submitters must be authorized and name provided on an Authorized Certificate List
- Digital certificate requests are reviewed and approved by the NRC
- Digital certificate is provided by the NRC

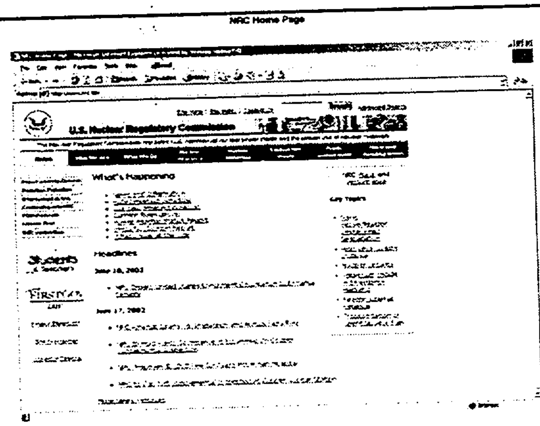
June 25-26, 2002 Page 27

Digital Signatures

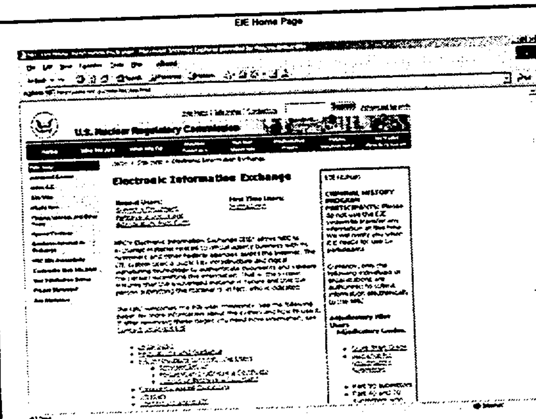
- Required for documents submitted under oath and affirmation
- Not required for other submittals
- Each submittal must have a "secure transmittal authorization" by an individual who is on the Approved Certificate List

June 25-26, 2002 Page 28

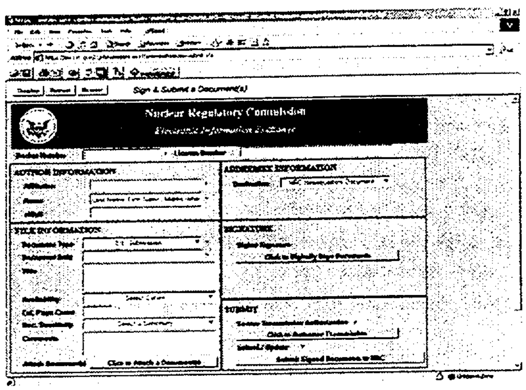
NRC Home Page



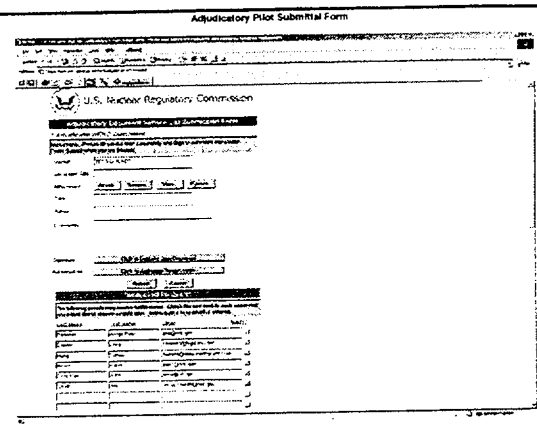
EIE Home Page



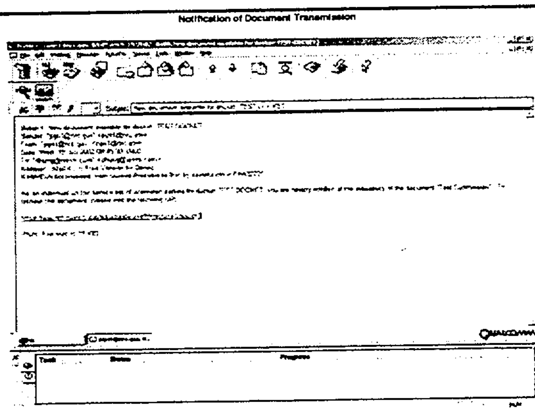
General Submission Page



Adjudicatory Pilot Submittal Form



Notification of Document Transmission



DOE Participation in EIE

- DOE may participate in the current EIE process
- Current process has number of constraints
- Submissions limited to 15Mb
- Larger documents could be submitted via CD-ROM

June 25-26, 2002 Page 34

Agenda

- Introduction - NRC
- High-Level Waste (HLW) Information Architecture - NRC
- Use of NRC's Electronic Information Exchange (EIE) - NRC
- **Electronic Hearing Docket (EHD) Document Format - NRC**
- Electronic Courtroom - NRC
- Technical Issues with a Potential Electronic License Application (LA) Submission and Other Large Documents - NRC
- Status of DOE's Effort for Putting Its Documentary Collection Out on the Licensing Support Network (LSN) - DOE
- Status of NRC's Effort for Putting Its Documentary Collection Out on the Licensing Support Network (LSN) - NRC
- Status of Search and Retrieval Testing - DOE
- Status of NRC LSN Test Server - NRC
- Meeting Summary – NRC

June 25-26, 2002 Page 35

EHD for a Potential HLW Proceeding

Emile L. Julian
Office of the Secretary
(301) 415-1966
ELJ@NRC.GOV

June 25-26, 2002 Page 36

Presentation Outline

- 10 CFR Part 2, Subpart J, docket requirements
- Electronic docket
- Service of documents
- Pleadings and other docket submissions
- Background materials

June 25-26, 2002 Page 37

10 CFR Part 2, Subpart J, Docket Requirements

- Electronic Docket Establishment
  - Secretary will establish an electronic docket for the HLW proceeding commencing with the docketing of the DOE application (10 CFR 2.1013(a))
  - Secretary must determine that the application is accessible through the electronic docket (10 CFR 2.1012(a) and 2.1013(a))
  - Electronic docket will also serve as the Pre-application docket for filings in disputes on document availability during the pre-license application phase (10 CFR 2.1010(d))

June 25-26, 2002 Page 38

10 CFR Part 2, Subpart J, Docket Requirements (cont.)

- Document submission requirements
  - Filings are to be submitted to the docket using a password security code for electronic submission (10 CFR 2.1013(c)(1))
  - Filings are to be submitted in searchable full-text, or, if not suitable for submission in searchable full-text, by header and image as appropriate (10 CFR 2.1013(a)(2))
  - Filings are to be served on the participants in the proceeding electronically (10 CFR 2.1013(c)(1))
  - Proof of service of filings will be satisfied when the sender of the electronic document receives an electronic acknowledgement from the recipient (10 CFR 2.1013(c)(4))

June 25-26, 2002 Page 39

10 CFR Part 2, Subpart J, Docket Req. (cont.)

- Content of the Electronic Hearing Docket (EHD)
  - Filings that are typically found in an NRC adjudication (10 CFR 2.1013(a)(2). Also see description of docket contents in Background materials. The description is also found on the NRC/EHD web site at http://ehd.nrc.gov/EHDpublic/browser.asp)
  - Orders of the Commission or Presiding Officers (10 CFR 2.1013(c)(5))
  - Transcripts of proceedings and hearing exhibits (10 CFR 2.1013(b))

June 25-26, 2002 Page 40

Electronic Docket

- Located on the NRC web site (http://www.nrc.gov/reading-rm/ehd.html)
- Location allows direct access to the docket
- Contains folders
  - Folders will contain filings, orders, transcripts, exhibits, and protective order documents
- Search capability
  - Full-text
  - Document attributes
  - Scroll through folder content

June 25-26, 2002 Page 41

Electronic Service of Documents

- Electronic Information Exchange (EIE) will facilitate service of pleadings, orders, and other documents
- EIE provides document authentication
- EIE will allow secure submission of protective order documents

June 25-26, 2002 Page 42

Pleadings and Other Docket Submissions

- NRC plans to adopt Adobe Acrobat® Portable Document Format (PDF) as a file format requirement for pleadings and other submissions to the docket
- NRC will provide more detail on filing submissions and file format requirements at the conclusion of the EIE pilot and large document review

Background

- Description of Electronic Hearing Docket (EHD)
- Accessing the High-Level Waste Electronic Hearing Docket (HLW EHD)

Description of Electronic Hearing Docket (EHD)

- Purpose of docket

  To provide parties to adjudicatory proceedings before the NRC, designated Atomic Safety and Licensing Boards, or presiding officers with readily available access to hearing documents. The central location of hearing documents in the EHD provides several ways to find documents on a proceeding. One can find documents in the appropriate hearing folder or by using the search tools provided on the EHD web site. Although the EHD has been established as the NRC's electronic docket for a potential adjudication on the DOE selected site for the storage of high-level radioactive waste, eventually it will be used as the hearing docket for all adjudications. Presently, hearing documents can be found in ADAMS.

- Adjudicatory docket definition

  This definition comports with the authority of the Secretary under 10 CFR 2.702 to maintain an adjudicatory docket. Under the Secretary's authority, all documents related to a given proceeding, whether or not relied upon by the Commission, an Atomic Safety and Licensing Board, or a presiding officer as part of the decision-making process, are deemed adjudicatory documents. The adjudicatory docket is divided into two parts: the decisional docket (at times referred to as the hearing docket) and the related correspondence docket.

Description of EHD (cont.)

- Decisional docket contents

  Documents that are actually considered b_ either the Commission exercising original jurisdiction or sitting as an appellate body or by Atomic Safety and Licensing Boards or presiding officers in the course of hearings are filed on the decisional docket.

  Documents that are likely to be filed before a Licensing Board or a presiding officer include Commission referral of intervention petitions and hearing requests, late filed intervention requests, party contentions and responses thereto, motions and associated pleadings, evidence presented at hearings (documentary exhibits and oral testimony), limited appearance statements, proceeding transcripts, In Camera filings, Board notifications, letters from parties to the Licensing Board, findings drafted by the parties, settlement agreements and various Board orders and memoranda such as Board establishments, scheduling orders, pre-hearing conference and post conference orders, hearing decisions and termination or dismissal orders. In addition to the above, for proceedings conducted pursuant to 10 CFR Part 2, Subpart L, a hearing file as defined by 10 CFR 2.1231(b) is submitted by the NRC Staff to the adjudicatory docket.

Description of EHD (cont.)

Documents that are likely to be filed before the Commission are intervention and hearing petitions, appeals and petitions for review, motions and associated pleadings, and appellate briefs. Orders that may be issued by the Commission as part of the hearing process include scheduling orders, orders extending the time for filings or for Commission rulings, and Commission Legal Issuances (CLIs) which are Commission decisions resolving appellate matters or motions originating before the Commission. In addition to orders, the hearing docket includes final agency action memoranda issued by the Secretary in the event the Commission declines to take review of a decision of the Licensing Board.

- Related correspondence docket

  Maintained to accommodate those filings that would not be considered by the Licensing Board or the Commission as part of the hearing process. These documents include documents generated as part of the discovery process such as interrogatories and answers to interrogatories, depositions, requests for admissions and admissions, requests for production of documents, prefiled testimony, and correspondence between the parties.

  It shou_d be noted that Federal Courts of Appeal usually consider only the Decisional doc_et including the final decision of the Commission in reviewing NRC adjudications.

Description of EHD (cont.)

- High-level waste proceeding

  The EHD will contain documents that are generated in a pre-application hearing as well as a hearing on the DOE license application. Prior to the filing of the application by the DOE, disputes could arise among potential parties on document access and document availability. In accordance with 10 CFR 2.1010, a pre-license application presiding officer will be designated to preside over such disputes. The Secretary, pursuant to 10 CFR 2.1010(d), shall maintain a docket to receive pre-application proceeding pleadings. The authority of the Secretary to maintain a pre-application docket is derived from the Secretary's authority under 10 CFR 2.1013 to maintain a docket for a proceeding which may follow the docketing of the DOE licenses application. The EHD is structured to accommodate both proceedings.

Description of EHD (cont.)

- Filing of documents

  The EHD is a visual representation of the official docket maintained by the Secretary of the Commission. Parties to any proceeding under 10 CFR Part 2, Subpart J, will be expected to utilize the NRC's Electronic Information Exchange (EIE) process for the submission of pleadings. Documents may also be filed in other forms directly with the Secretary if EIE filing is not possible because of the size of the electronic file or the nature of the filing, such as video or audio media.

Accessing the High-Level Waste Electronic Hearing Docket

Start at NRC Home Page

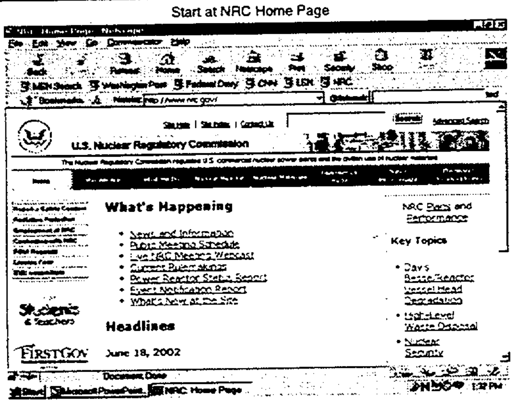


Figure 1

Click on the Electronic Reading Room Tab and select HLW Electronic Hearing Docket

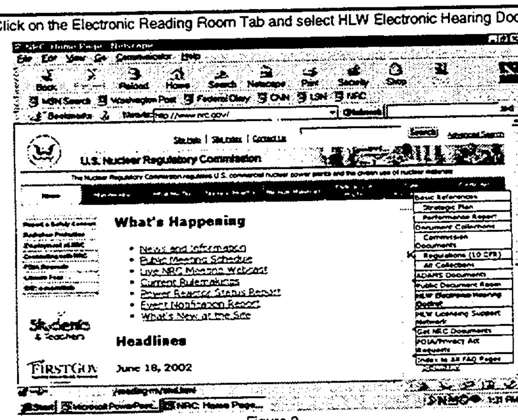


Figure 2

Click on Login for Public Users

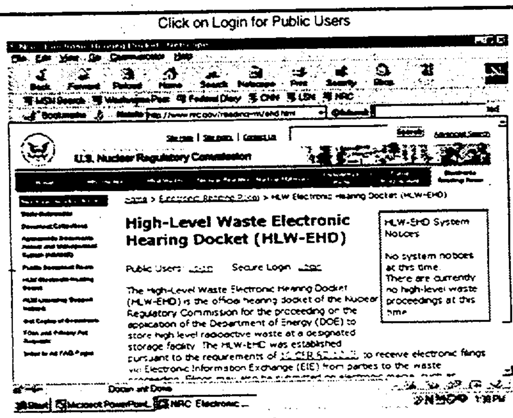


Figure 3

Click on Logon

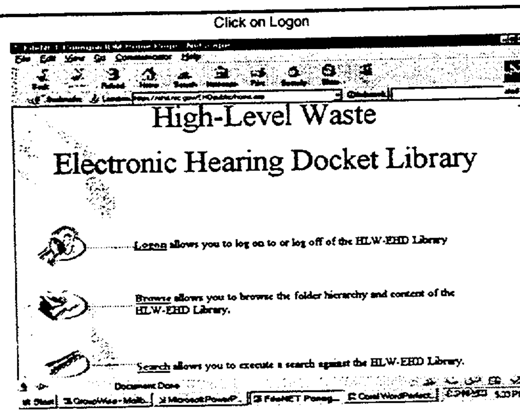


Figure 4

Logon as guest

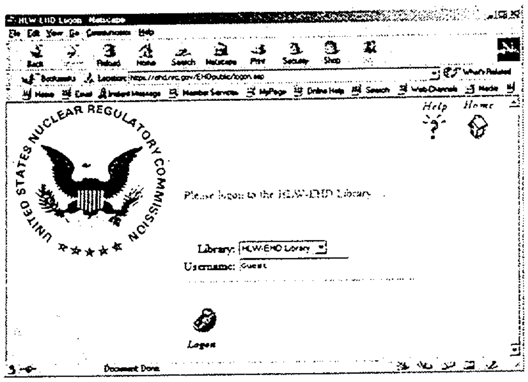


Figure 5

Browse folders

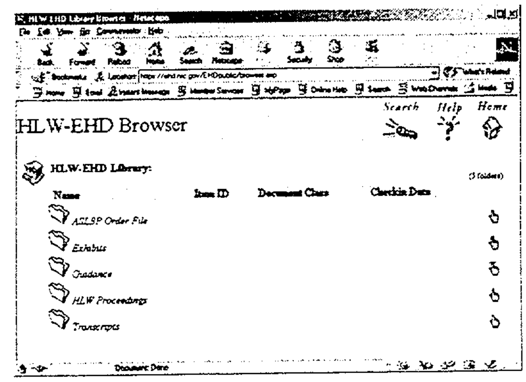


Figure 6

Samples of documents in Guidance folder

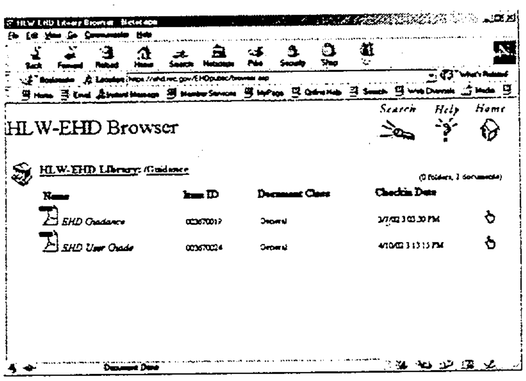


Figure 7

Perform searches

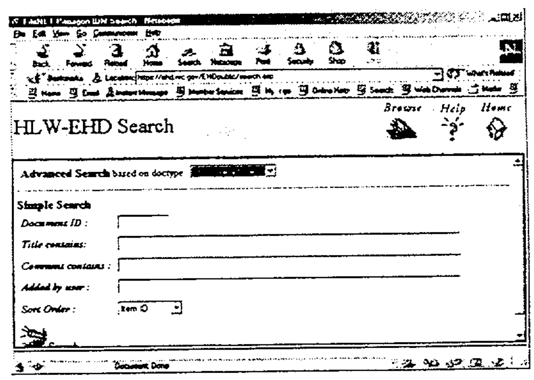


Figure 8

Agenda

- Introduction - NRC
- High-Level Waste (HLW) Information Architecture - NRC
- Use of NRC's Electronic Information Exchange (EIE) - NRC
- Electronic Hearing Docket (EHD) Document Format - NRC
- **Electronic Courtroom - NRC**
- Technical Issues with a Potential Electronic License Application (LA) Submission and Other Large Documents - NRC
- Status of DOE's Effort for Putting Its Documentary Collection Out on the Licensing Support Network (LSN) - DOE
- Status of NRC's Effort for Putting Its Documentary Collection Out on the Licensing Support Network (LSN) - NRC
- Status of Search and Retrieval Testing - DOE
- Status of NRC LSN Test Server - NRC
- Meeting Summary – NRC

Electronic Courtroom

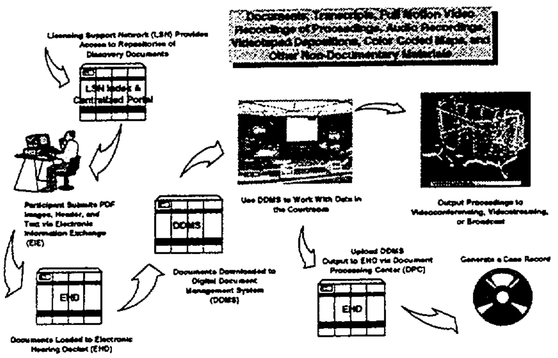
Dan Graser

Atomic Safety and Licensing Board Panel

(301) 415-7401

DJG2@NRC.GOV

Digital Proceeding From Start to Finish

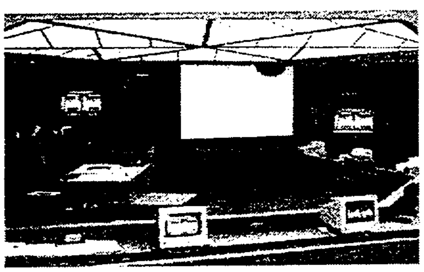


Electronic Courtroom Objectives

- Integrated environment for judges and counsel
- Electronic Hearing Docket (EHD)
- Case/hearing management
- Access to legal information (Lexis, Westlaw, etc.)
- Audio/visual presentation tools
- Real-time transcription and captioning
- Multi-point videoconferencing
- Webcasting (under consideration)

June 25-26, 2002 Page 62

Vision



June 25-26, 2002 Page 63

Key Features

- Audio/visual presentation devices to leverage electronic documents
- Voice activated cameras/microphones
- Video-conferencing
- Recordation via real-time court reporter
- Recordation of audio and video feeds
- Outputs to other media (webcasting) are under consideration
- Daily updates to Electronic Hearing Docket (EHD)
- General docket materials and transcripts searchable
- Protective order files selectively available and similarly searchable
- Docket file contents accessible in courtroom and remotely via web
- Generation of case record for appellate review

June 25-26, 2002 Page 64

Operational Approach

- Prefiled potential exhibits downloaded from EHD into the courtroom database
- Courtroom database updated with markings for pre-filed items as introduced
- Items first introduced in courtroom incorporated into and marked in courtroom database
- Courtroom database uploaded daily to refresh EHD
- Transcripts processed overnight

June 25-26, 2002 Page 65

Electronic Media Issues

- Introduction of large documents
  - Full document
  - Sections or pages of documents
- Courtroom retrieval and use without interrupting flow of proceedings
- Navigation within document to display page(s) counsel is presenting
  - Quickly
  - Clear image for large screen projections
- Formats
  - To be digitally recorded for case file
  - Must meet rules/criteria for evidence
  - Capable of being retired to National Archives
  - Courtroom case management and documents to be web accessible (download issues)

June 25-26, 2002 Page 66

Electronic Media Issues Impact

- Requires skilled Clerk of Court and paralegals
- New paradigm for NRC judges
- Parties' counsel need to become familiar
- Training/dry runs/shake-out during pre-hearings are all essential to success

June 25-26, 2002 Page 67

Agenda

- Introduction - NRC
- High-Level Waste (HLW) Information Architecture - NRC
- Use of NRC's Electronic Information Exchange (EIE) - NRC
- Electronic Hearing Docket (EHD) Document Format - NRC
- Electronic Courtroom - NRC
- **Technical Issues with a Potential Electronic License Application (LA) Submission and Other Large Documents - NRC**
- Status of DOE's Effort for Putting Its Documentary Collection Out on the Licensing Support Network (LSN) - DOE
- Status of NRC's Effort for Putting Its Documentary Collection Out on the Licensing Support Network (LSN) - NRC
- Status of Search and Retrieval Testing - DOE
- Status of NRC LSN Test Server - NRC
- Meeting Summary – NRC

June 25-26, 2002 Page 68

Technical Issues with a Potential Electronic License Application (LA) Submission & Other Large Documents

Lynn Scattolini
Office of the Chief Information Officer
(301) 415-8730
LBS@NRC.GOV

June 25-26, 2002 Page 69

Presentation Outline

- Electronic format and resolution standards
- Ability to move very large documents through Electronic Information Exchange (EIE)
- Issues related to processing and making very large documents with special attributes available in a usable way
- Approach to addressing issues

June 25-26, 2002 Page 70

Electronic Format and Resolution Standards

- 10 CFR Part 2, Subpart J, allows for a number of electronic formats and for gray scale and color images at a resolution of 150 dpi (dots per inch)
- These formats and resolution are authorized for the documentary materials that reside on participants' servers for LSN searching purposes, and for electronic document production and service

June 25-26, 2002 Page 71

Electronic Format and Resolution Standards (cont.)

- NRC plans to adopt PDF and a minimum resolution of 200 dpi for gray scale and color images as a standard for electronic submission to the HLW proceeding

June 25-26, 2002 Page 72

Ability to Move Very Large Documents Through EIE

- Current design of NRC's EIE process does not accommodate large, segmented HLW files such as DOE's Final Environmental Impact Statement
- The way in which a large document is submitted electronically can impact its capture, usability to support different work processes, and public access

June 25-26, 2002 Page 73

Issues Related to Processing and Making Very Large Documents with Special Attributes Available in a Usable Way

- Need to ensure fidelity, integrity, and currency of documents
- Records management issues
- Size of files – technology limitations of users
- Ability for user to search and navigate
- Some file types and electronic media pose problems
- NRC's document management system does not support external hyperlinks

June 25-26, 2002 Page 74

Approach to Addressing Issues

- Identify scope, characteristics, timing, number of very large submissions
- Identify needs of stakeholders and their different usages of documents and associated processes
- Evaluate solutions available to meet requirements
- Implement proposed solution(s) and issue special guidance on the electronic submission of textual, graphical, and other documentary materials in the HLW proceeding

June 25-26, 2002 Page 75

Approach to Addressing Issues (cont.)

- Consider the range of capabilities that exist in stakeholder community
- Technology limitations
- Agency/stakeholder costs associated with meeting any guidance, such as document conversion costs

June 25-26, 2002 Page 76

Agenda

- Introduction - NRC
- High-Level Waste (HLW) Information Architecture - NRC
- Use of NRC's Electronic Information Exchange (EIE) - NRC
- Electronic Hearing Docket (EHD) Document Format - NRC
- Electronic Courtroom - NRC
- Technical Issues with a Potential Electronic License Application (LA) Submission and Other Large Documents - NRC
- **Status of DOE's Effort for Putting Its Documentary Collection Out on the Licensing Support Network (LSN) - DOE**
- Status of NRC's Effort for Putting Its Documentary Collection Out on the Licensing Support Network (LSN) - NRC
- Status of Search and Retrieval Testing - DOE
- Status of NRC LSN Test Server - NRC
- Meeting Summary – NRC

June 25-26, 2002 Page 77

Agenda

- Introduction - NRC
- High-Level Waste (HLW) Information Architecture - NRC
- Use of NRC's Electronic Information Exchange (EIE) - NRC
- Electronic Hearing Docket (EHD) Document Format - NRC
- Electronic Courtroom - NRC
- Technical Issues with a Potential Electronic License Application (LA) Submission and Other Large Documents - NRC
- Status of DOE's Effort for Putting Its Documentary Collection Out on the Licensing Support Network (LSN) - DOE
- **Status of NRC's Effort for Putting Its Documentary Collection Out on the Licensing Support Network (LSN) - NRC**
- Status of Search and Retrieval Testing - DOE
- Status of NRC LSN Test Server - NRC
- Meeting Summary – NRC

June 25-26, 2002 Page 78

Status of Putting NRC's High-Level Waste Document Collection on the Licensing Support Network (LSN)

Jeff Ciocco
High-Level Waste Branch
Office of Nuclear Material Safety and Safeguards
(301) 415-6391
JAC3@NRC.GOV

Efforts Underway to Place Document Collection on LSN

- At this time, no NRC documents available to the LSN server
- Identifying and adding documentary materials
- As a result of 9/11, NRC is screening documents prior to making them available to the LSN server

Adding Documentary Material to Document Mgmt System

- Staff directed to expeditiously make their documents available to the LSN
- Staff directed to encourage other parties to expeditiously make their documents available to the LSN
- Staff are evaluating NRC and its contractor documents to determine what should be added to comply with Subpart J
- The documents under evaluation include graphic-oriented documents such as calibration records, data logs, scientific notebooks, and computer print-outs

Screening for Sensitive Homeland Security Information

- NRC has re-examined existing policies on the dissemination of information routinely provided to the public
- Recently, the Commission approved screening criteria and policy for withholding sensitive homeland security information
- Based on Commission guidance, expect majority of documents to become available

Summary

- NRC is taking steps to ensure compliance with the LSN requirements
- NRC will place documents on the NRC HLW server after screening for sensitive homeland security information

Agenda

- Introduction - NRC
- High-Level Waste (HLW) Information Architecture - NRC
- Use of NRC's Electronic Information Exchange (EIE) - NRC
- Electronic Hearing Docket (EHD) Document Format - NRC
- Electronic Courtroom - NRC
- Technical Issues with a Potential Electronic License Application (LA) Submission and Other Large Documents - NRC
- Status of DOE's Effort for Putting Its Documentary Collection Out on the Licensing Support Network (LSN) - DOE
- Status of NRC's Effort for Putting Its Documentary Collection Out on the Licensing Support Network (LSN) - NRC
- **Status of Search and Retrieval Testing - DOE**
- Status of NRC LSN Test Server - NRC
- Meeting Summary – NRC

Agenda

- Introduction - NRC
- High-Level Waste (HLW) Information Architecture - NRC
- Use of NRC's Electronic Information Exchange (EIE) - NRC
- Electronic Hearing Docket (EHD) Document Format - NRC
- Electronic Courtroom - NRC
- Technical Issues with a Potential Electronic License Application (LA) Submission and Other Large Documents - NRC
- Status of DOE's Effort for Putting Its Documentary Collection Out on the Licensing Support Network (LSN) - DOE
- Status of NRC's Effort for Putting Its Documentary Collection Out on the Licensing Support Network (LSN) - NRC
- Status of Search and Retrieval Testing - DOE
- **Status of NRC LSN Test Server - NRC**
- Meeting Summary – NRC

June 25-26, 2002 Page 85

Status of NRC Licensing Support Network Test Server

Dan Graser

Atomic Safety and Licensing Board Panel

U.S. Nuclear Regulatory Commission

(301) 415-7401

DJG2@NRC.GOV

June 25-26, 2002 Page 86

Status of NRC LSN Portal
Current Planning Dates

- LSN release 2.0 incorporates enhanced security and database/audit administration
  - Release 2.0 installed — 4/30/02
  - Enhanced spider target acceptance — 7/31/02
- Place test documents on production server to validate capabilities — ~6/30/02
- Production server acceptance testing — 8/9/02-9/13/02
- Flush test documents from production server — 9/13/02
- Production server resumes production of live documents — 9/16/02

June 25-26, 2002 Page 87

Agenda

- Introduction - NRC
- High-Level Waste (HLW) Information Architecture - NRC
- Use of NRC's Electronic Information Exchange (EIE) - NRC
- Electronic Hearing Docket (EHD) Document Format - NRC
- Electronic Courtroom - NRC
- Technical Issues with a Potential Electronic License Application (LA) Submission and Other Large Documents - NRC
- Status of DOE's Effort for Putting Its Documentary Collection Out on the Licensing Support Network (LSN) - DOE
- Status of NRC's Effort for Putting Its Documentary Collection Out on the Licensing Support Network (LSN) - NRC
- Status of Search and Retrieval Testing - DOE
- Status of NRC LSN Test Server - NRC
- **Meeting Summary – NRC**

June 25-26, 2002 Page 88

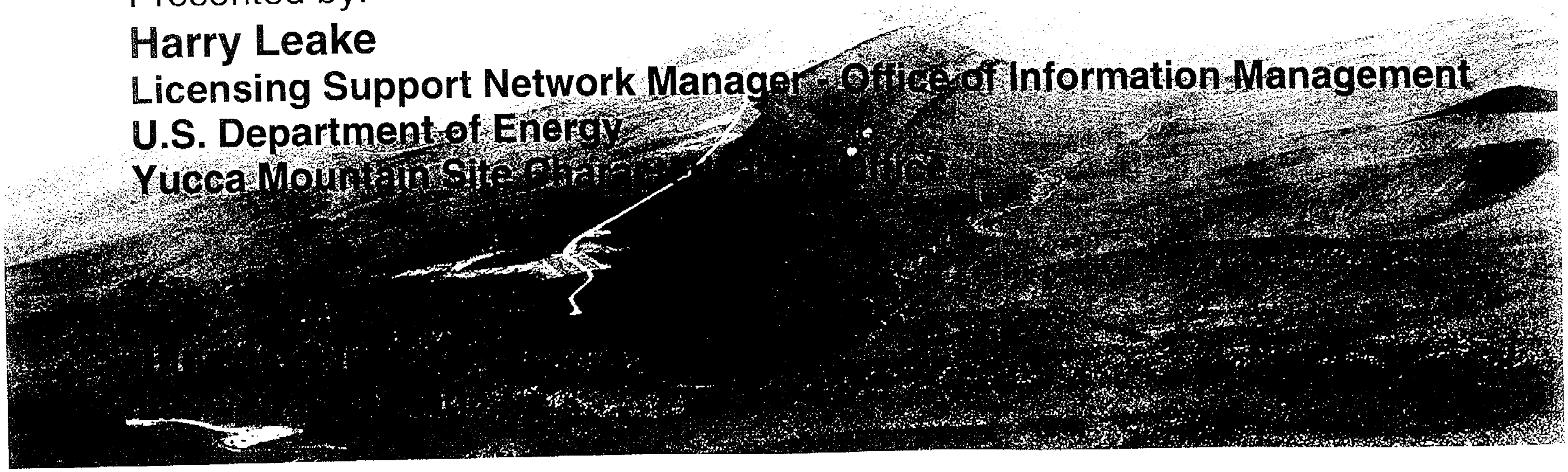
U.S. Department of Energy
Office of Civilian Radioactive Waste Management

YUCCA MOUNTAIN PROJECT

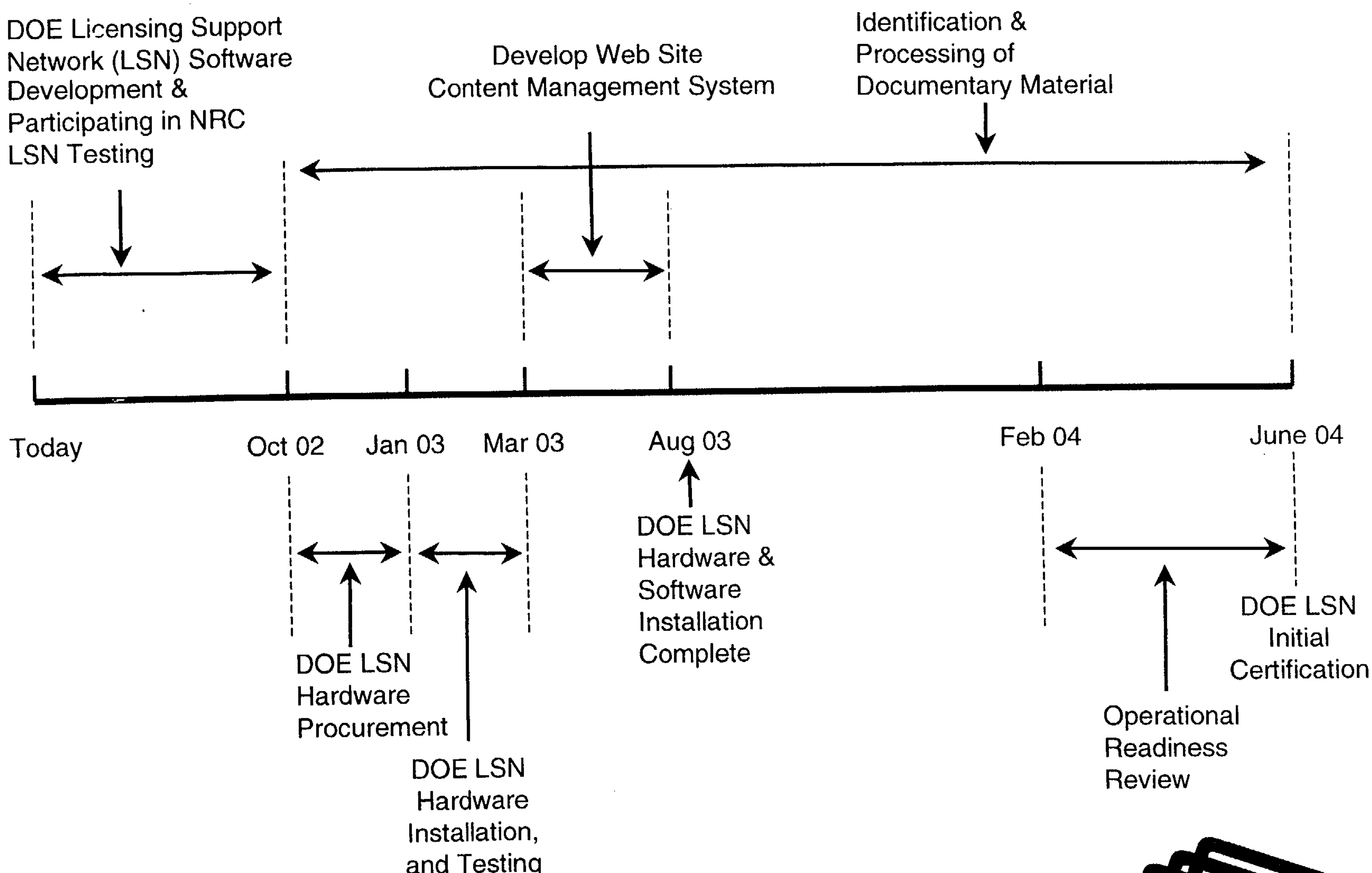
Status of DOE's Effort for Putting its Documentary Material on the Licensing Support Network

Presented to:
NRC/DOE Technical Exchange on Electronic Submissions

Presented by:
Harry Leake
Licensing Support Network Manager - Office of Information Management
U.S. Department of Energy
Yucca Mountain Site Characterization Office

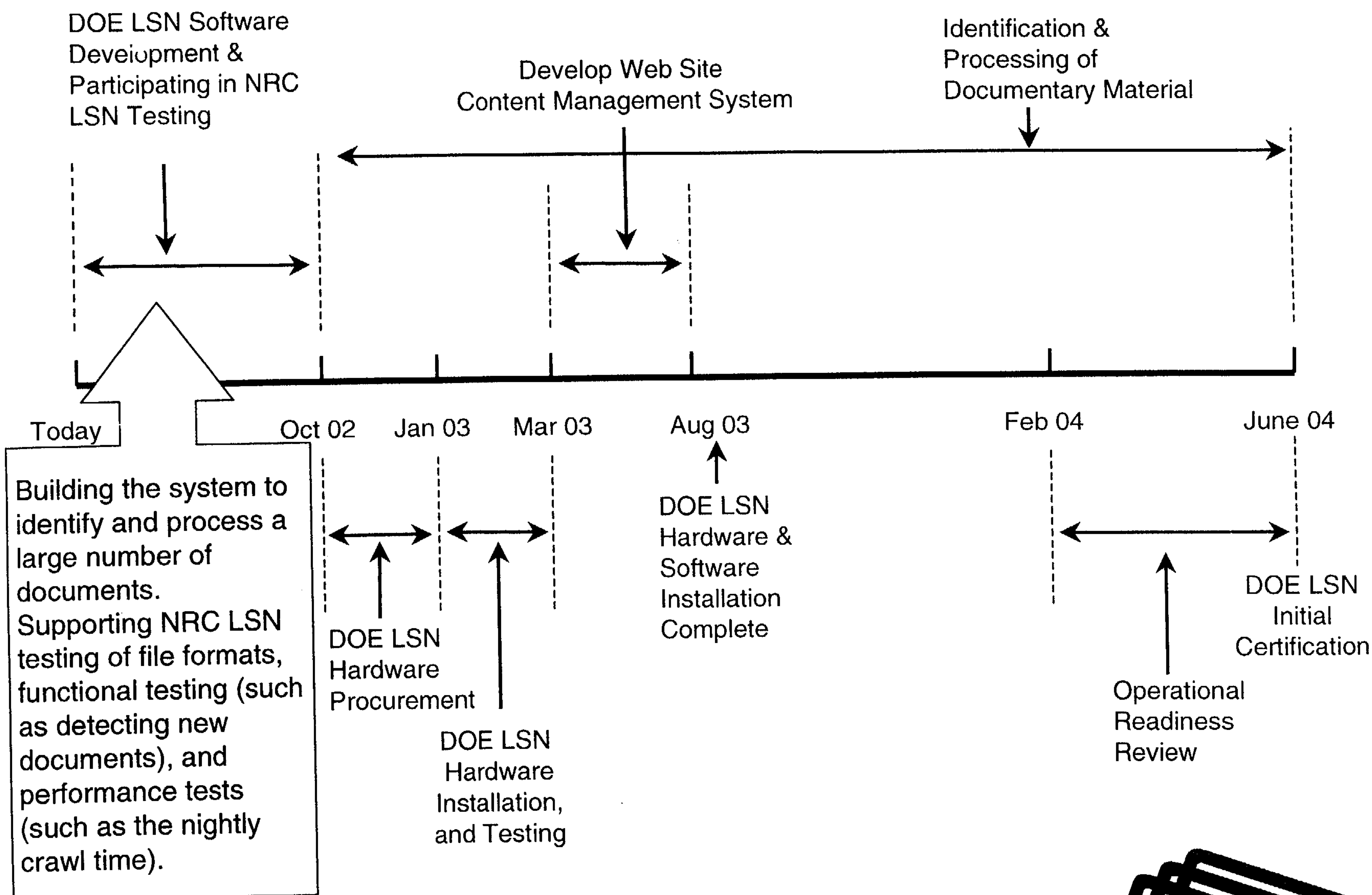


Current DOE Licensing Support Network Plan



Note: Dates above are approximate, represent current planning and may be subject to revision.

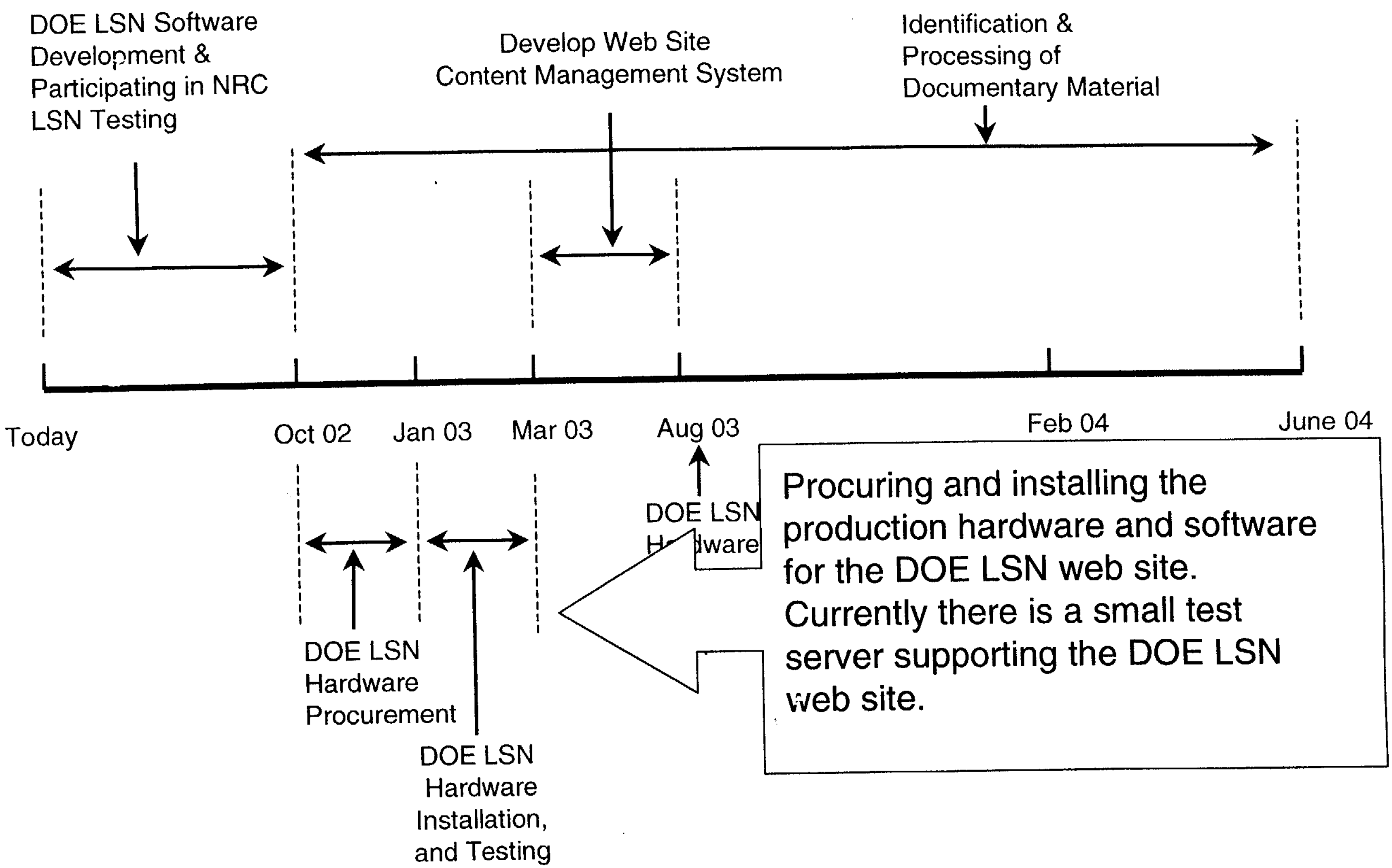
Current DOE Licensing Support Network Plan
(Continued)



Note: Dates above are approximate, and represent current planning and may be subject to revision.

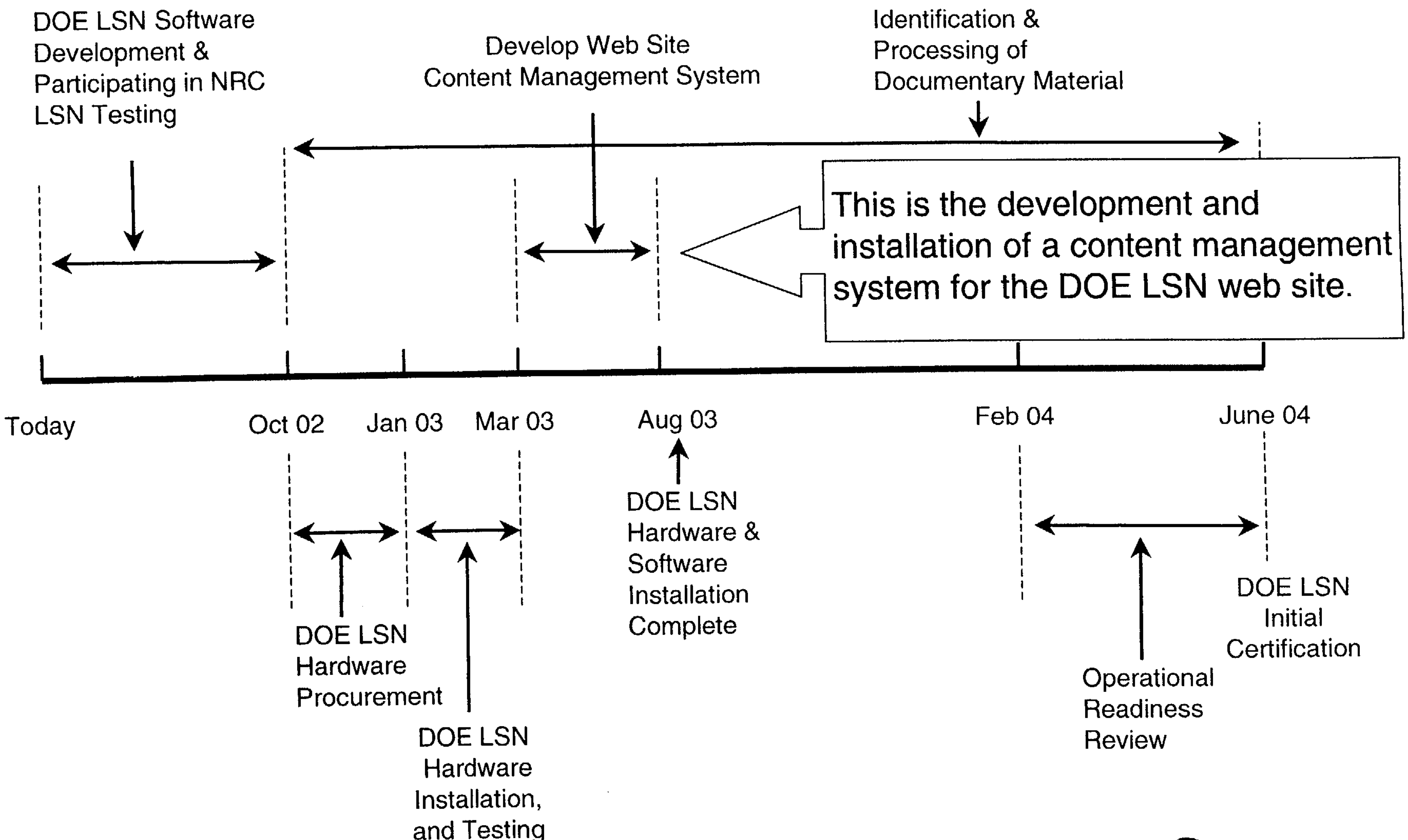
Current DOE Licensing Support Network Plan

(Continued)



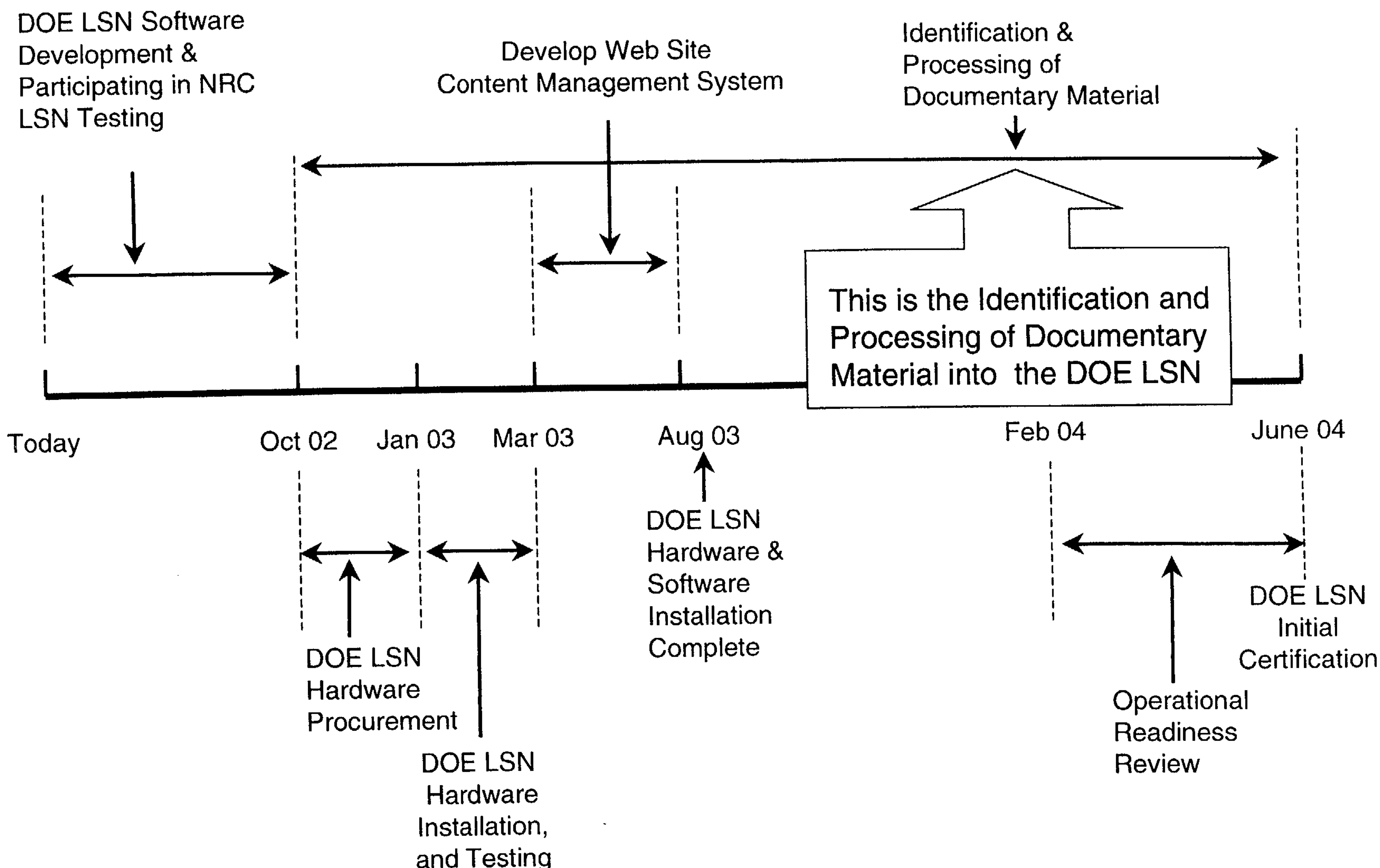
Note: Dates above are approximate, represent current planning and may be subject to revision.

Current DOE Licensing Support Network Plan
(Continued)



Note: Dates above are approximate, represent current planning and may be subject to revision.

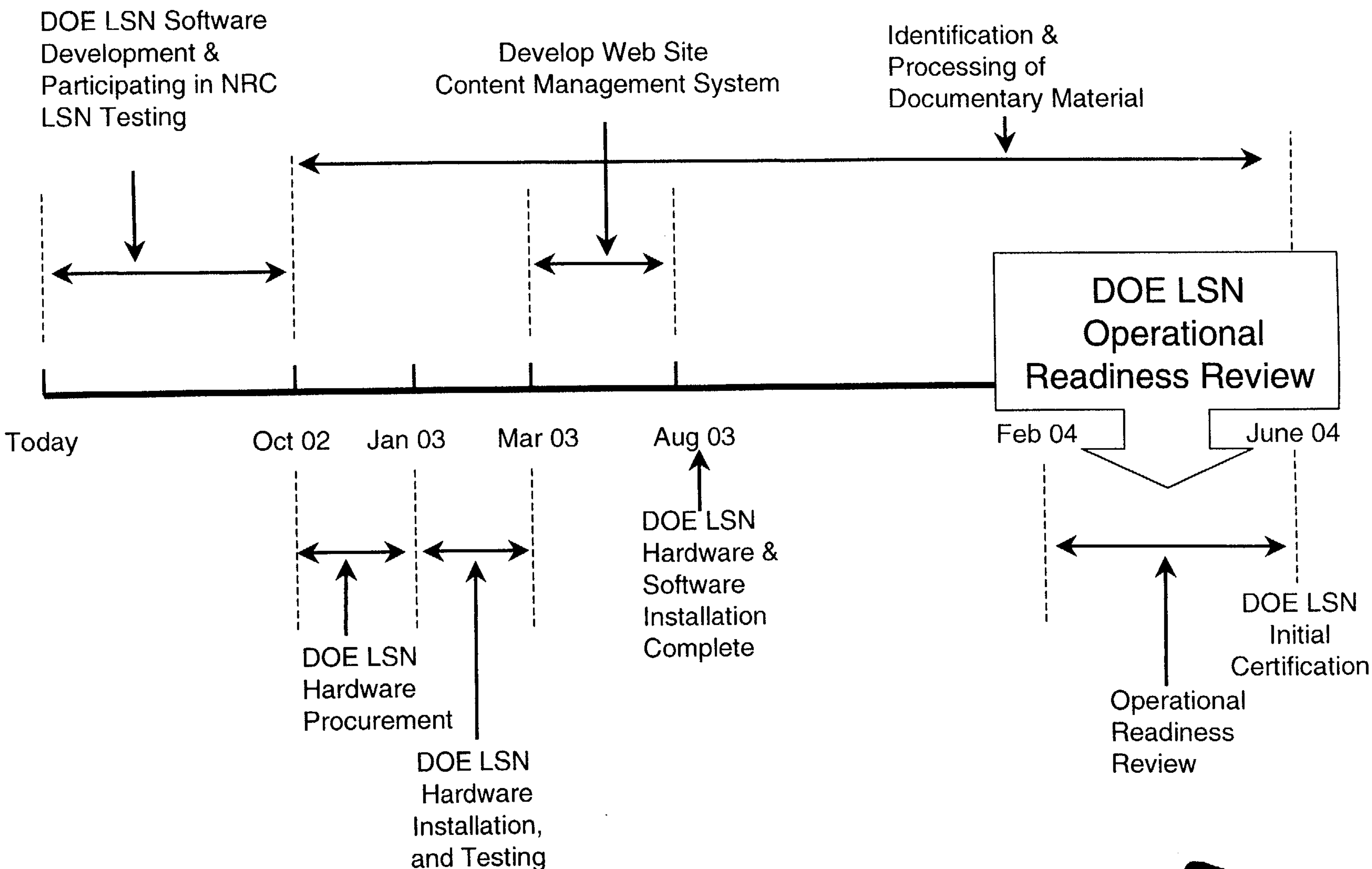
Current DOE Licensing Support Network Plan
(Continued)



Note: Dates above are approximate, represent current planning and may be subject to revision.

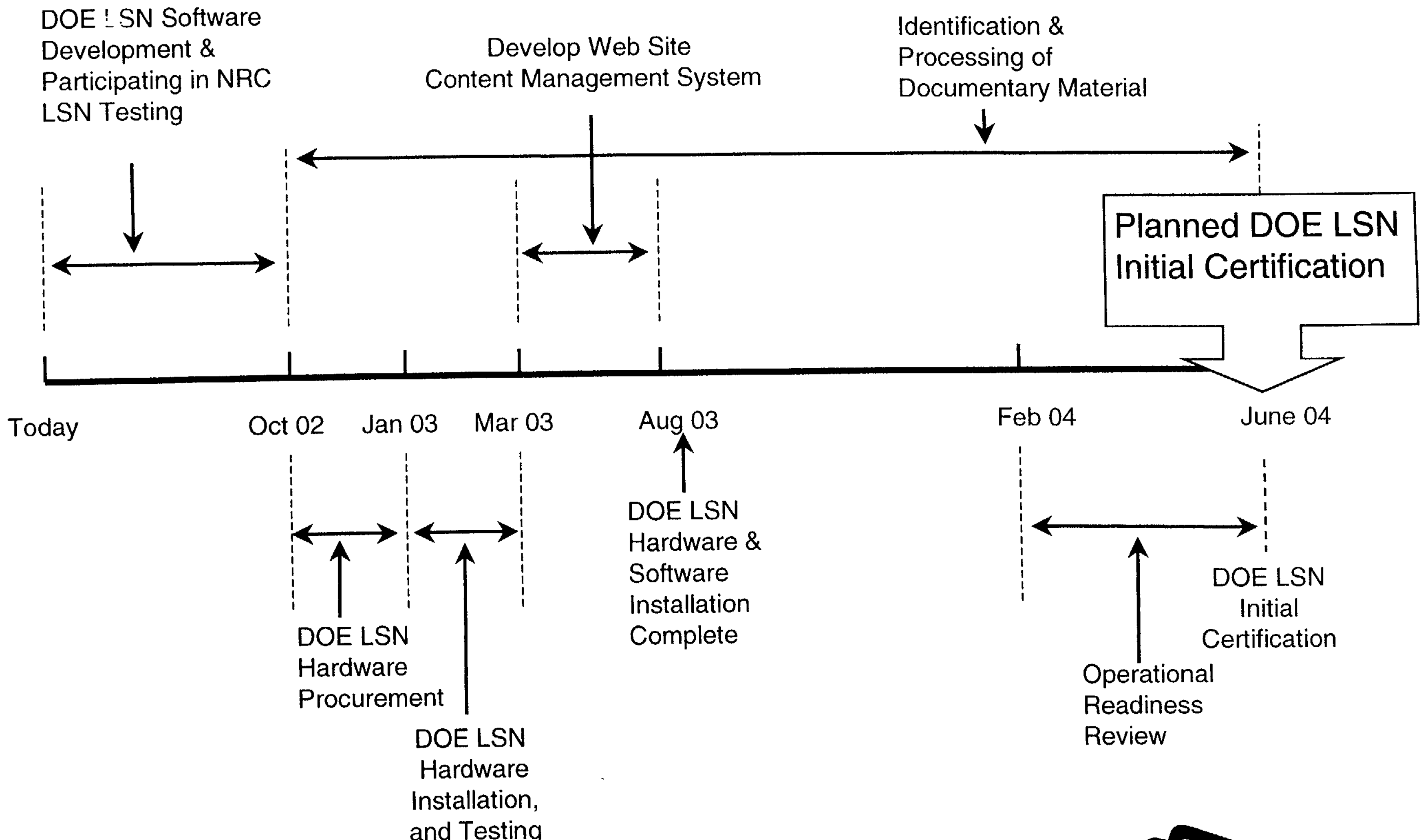
Current DOE Licensing Support Network Plan

(Continued)



Note: Dates above are approximate, represent current planning and may be subject to revision.

Current DOE Licensing Support Network Plan
(Continued)



Note: Dates above are approximate, represent current planning and may be subject to revision.

U.S. Department of Energy
Office of Civilian Radioactive Waste Management

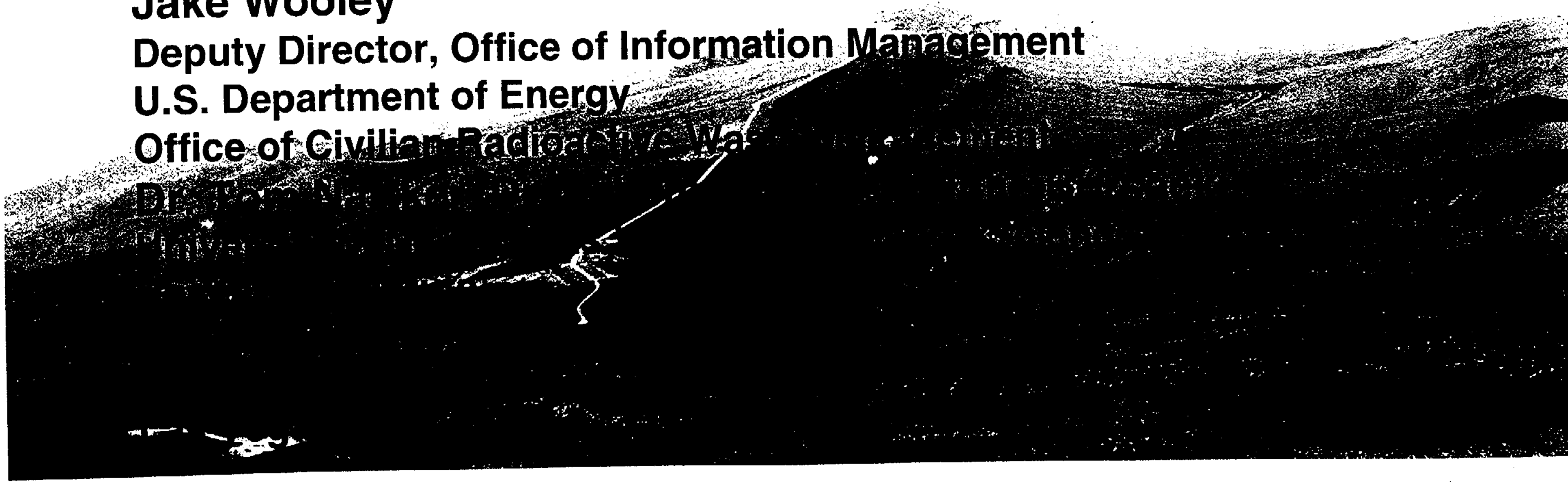
YUCCA MOUNTAIN PROJECT

Evaluation of the Current DOE Document Conversion System: A Study of Retrievability

Presented to:
NRC/DOE Technical Exchange on Electronic Submissions

Presented by:

Jake Wooley
Deputy Director, Office of Information Management
U.S. Department of Energy
Office of Civilian Radioactive Wa[illegible]
[illegible]
[illegible]



Study of Retrievability

- Agenda
  - Introduction
  - Background
  - Document conversion system recommendations
  - Tests to measure document conversion performance
    - Text accuracy
    - Retrievability tests
  - Conclusion

Study of Retrievability

(Continued)

- **Background**
  - Who is Information Science Research Institute (ISRI)?
  - Licensing Support System (LSS) (1990 - Current)
  - Optical Character Reader (OCR) Conferences (1991 - 1995)
  - Contracted by M&O (1996 - 1999)
  - Contracted by DOE (1990 - 1995, 2000 - Current)
    - Current tasks for FY02
      - Provide recommendations on DOE document conversion system
      - Evaluate performance of DOE document conversion system

Study of Retrievability

(Continued)

- Document conversion system recommendations
  - Retrievability is a better performance metric than character accuracy
    - Not all characters are used by a retrieval system
  - Automatic zoning, followed by MANICURE, will produce retrievability equivalent to manual zoning
    - Manually zoned text
    - Automatic zoned text
    - MANICURE

Tests to Measure Document Conversion Performance

- Text accuracy
  - NRC Licensing Support Network Administrator (LSNA) target accuracy for OCR created text (Licensing Support Network Guidelines provided 1/02)
    - Goal is to have 99.5% accurate text
  - Text accuracy test
    - 17 documents (1253 pages, 164,483 non-stopwords, and 1,361,124 characters)
    - Non-stopword accuracy tests
      - DOE word accuracy between 96.15% and 97.23%
    - Tests of character accuracy of non-stopwords
      - DOE character accuracy between 98.83% and 99.30%

Tests to Measure Document Conversion Performance

(Continued)

- Retrievability tests
  - Test data
    - 1055 documents containing 75,236 Pages
    - 40 queries
    - Average number of relevancy judgements per query - 100
    - Autonomy Server™ v2.2.0
  - Retrievability metrics
    - Precision and recall

Tests to Measure Document Conversion Performance

(Continued)

- **Automatic zoned text versus manually zoned text retrieval tests**
  - Average precision
    - Manually zoned - 37.9%
    - Automatic zoned - 39.2%

Tests to Measure Document Conversion Performance

(Continued)

- Ranking of retrieved documents compared between the manually zoned text and the automatic zoned text
  - Importance of ranking in retrieval systems
  - Results of ranking tests
    - Correlation factor - .97
  - Ranking problems in information retrieval systems

Tests to Measure Document Conversion Performance

(Continued)

- Automatic zoned text versus 99.8% correct text retrieval tests
  - 1058 documents containing 46,731 Pages
  - 62 queries
  - Average number of relevant documents per query = 17
  - Average precision
    - 99.8% accurate text - 24.5%
    - Automatic zoned text - 24.2%

Study of Retrievability

- Conclusion
  - Character accuracy produced by DOE document conversion system close to NRC LSNA goal
  - MANICURE improves word and character accuracy
  - Average character accuracy of non-stopwords on DOE documents is between 98.83% and 99.30%
  - Retrievability is equivalent for automatic zoned text and manually zoned text
  - Ranking of query results is equivalent for automatic zoned text and manually zoned text
  - Retrievability is equivalent for automatic zoned text and 99.8% accurate text

Backup

Zoned Page Image

1

Tertiary extension north of the Las Vegas Valley shear zone, Sheep and Desert Ranges, Clark County, Nevada

2

PETER L. GUTH* *Department of Earth and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139*

3

ABSTRACT

Detailed mapping reveals the presence of high-angle extensional faults and low-angle gravity slides on the west side of the Sheep Range. Three major high-angle faulting events each account for 20° of eastward rotation and accommodate extension between the Las Vegas Range and the Desert Range. Low-angle faults represent surficial slides in response to topography produced by extension on the high-angle faults. Faulting took place during the Miocene, synchronously with deposition of the Horse Spring Formation and with displacement on the Las Vegas Valley shear zone. The extension in the Sheep Range took place without volcanism, intrusion, or metamorphism of the Paleozoic sedimentary rocks.

Offset thrust faults suggest that the area west of the Sheep Range extended almost 100% during the Miocene, while the corresponding area south of the Las Vegas Valley shear zone did not extend significantly. The shear zone bounded the extending terrane on the south, acting as a transform fault. This extension west of the Sheep Range may in part balance that mapped by Anderson (1971) in the Eldorado Mountains. The Las Vegas Valley shear zone and the Lake Mead fault system may have acted together to compensate for areas of localized extension between the Colorado Plateau and the vicinity of the Specter Range.

INTRODUCTION

The importance of extensional faulting in the development of structure and physiography in the Great Basin has been evident since the pioneering observations of Gilbert (1874). Recently Stewart (1971, 1980) reviewed and summarized the horst and graben, tilted-block, and listric-fault models commonly applied to the Great Basin.

4

Stewart (1971) calculated that the horst and graben model requires about 10% extension across the entire Great Basin, assuming that 60° dips exist on range-front faults. The tilted-block model of Morton and Black (1975) seems to require 20-30% extension across the entire Great Basin, but local extension might exceed 100% (Stewart, 1980). Inferred listric-fault geometry leads to local estimates of 30% to over 100% extension (Anderson, 1971, Wright and Troxel, 1973; Proffett, 1977), although the model might not apply to the entire Great Basin.

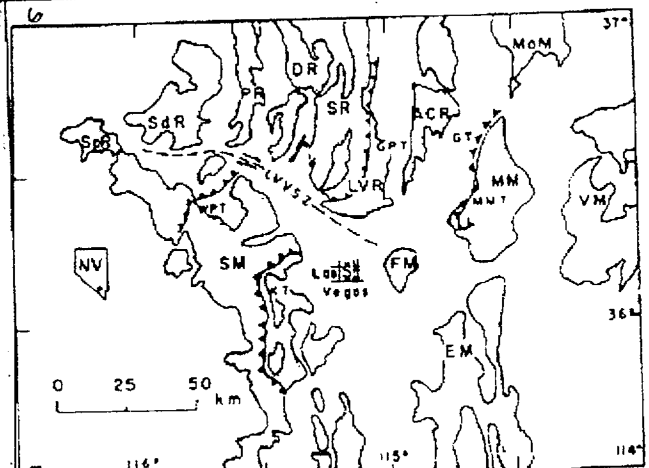
Strike-slip faults are related to extensional faulting as boundaries of domains with differences in style or magnitude of ex-

5

tension (Davis, 1979; Stewart, 1980). Strike-slip faults may be transform boundaries between regions of differential extension, such as the Garlock fault of southern California (Hamilton and Myers, 1966; Davis and Burchfiel, 1973). Recent work suggests this model for faulting in the Lake Mead area of southern Nevada (Bohannon, 1979a). Both strike-slip and extensional faulting appear to be dominantly late Tertiary events in the Great Basin.

Armstrong (1972) reviewed widespread low-angle denudational faulting in the eastern Great Basin. The faults generally place younger rocks on older, and Armstrong advocated a Tertiary gravity mechanism distinct from Mesozoic compression. In con-

6



7

Figure 1. Locality map of southern Nevada. Mountain ranges indicated and referred to in the text are: ACR, Arrow Canyon Range; DR, Desert Range; EM, Eldorado Mountains; FM, Frenchman Mountain; LVR, Las Vegas Range; MM, Muddy Mountains; MoM, Mormon Mountains; PR, Pintwater Range; SR, Sheep Range; SdR, Spotted Range; SpR, Specter Range; and VM, Virgin Mountains. Thrust faults indicated are the Wheeler Pass (WPT), Keystone (KT), Gass Peak (GPT), Muddy Mountain (MMT), and Glendale (GT).

8 * Present address: B Company, 34th Engineer Battalion (Combat) (Heavy), Fort Riley, Kansas 66442.

9 Geological Society of America Bulletin, Part I, v. 92, p. 763–771, 5 figs., 1 table, October 1981.

Example Queries

- Find all documents that discuss fracture frequency data (fracture density and radial fracture density) in boreholes at Yucca Mountain
- Find all documents that discuss strategies for environmental restoration and remediation (Hanford Site)

Precision and Recall

- Precision
  - # of relevant documents retrieved / # of retrieved documents
- Recall
  - # of relevant documents retrieved / total # of relevant documents
- Suppose there are 10 relevant documents for an example query
- And suppose the system returns 15 documents for this example query

Precision and Recall Example

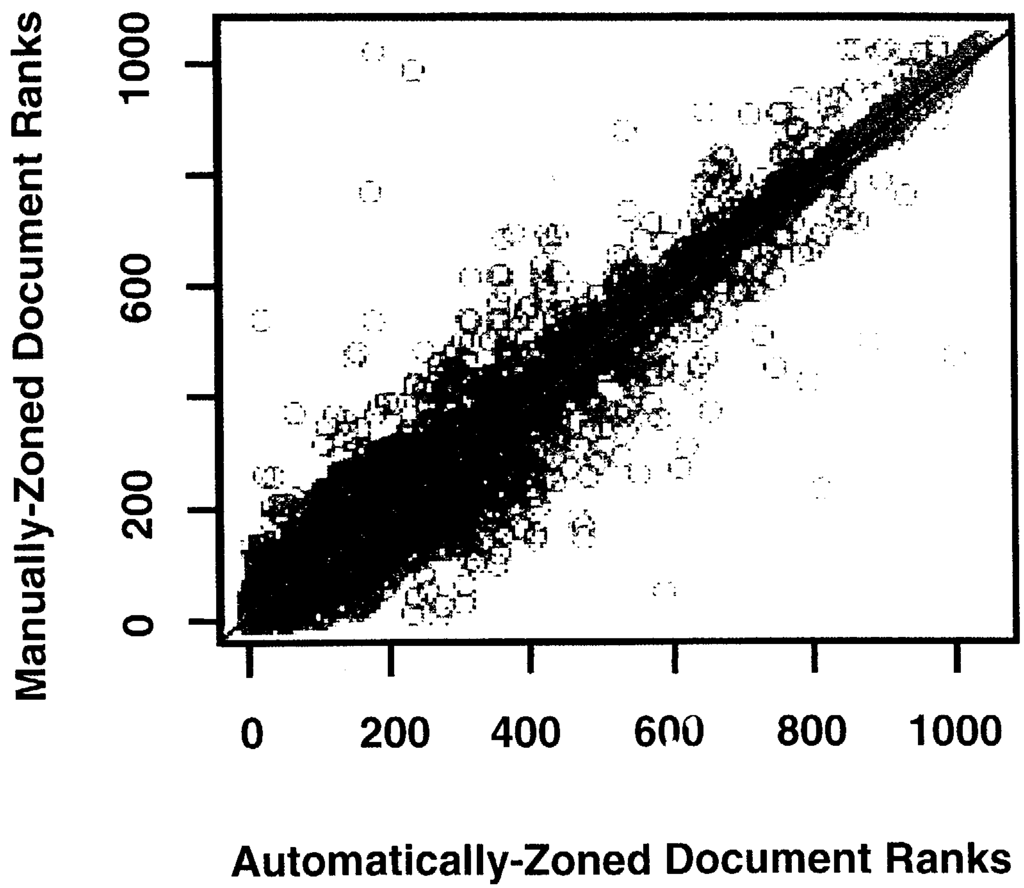
Docid	Relevant	Recall	Precision
1	R	10%	100%
2	--	10%	50%
3	--	10%	33%
4	R	20%	50%
5	R	30%	60%
6	--	30%	50%
7	R	40%	57%
8	--	40%	50%
9	R	50%	55%
10	R	60%	60%
11	R	70%	64%
12	--	70%	58%
13	R	80%	62%
14	R	90%	64%
15	R	100%	67%

NOTE: Docid - Document identified

Calculating Average Precision

- Add precision values at 10%, 20%, . . . , 100% recall levels
  - 100+50+60+57+55+60+64+62+64+67 = 639
- Average total precision / # of recall levels
  - 639 / 10 = 63.9%

Scatter Plot of Automatic Zoned and Manual Zoned Ranks



PUBLIC DOCUMENT ROOM FACT SHEET

LOCATION: One White Flint North
11555 Rockville Pike
Rockville, MD 20852

MAILING ADDRESS: U.S. Nuclear Regulatory Commission
Public Document Room
Washington, D.C. 20555

TELEPHONE NUMBERS: Voice: (301) 415-4737 or
(800) 397-4209
Fax: (301) 415-3548

INTERNET E-MAIL: pdr@nrc.gov

READING ROOM HOURS: 7:45AM-4:15PM Monday-Friday, Eastern Time, Closed Federal Holidays

TELEPHONE REFERENCE HOURS: 8:30AM-4:15PM Monday-Friday, Eastern Time, Closed Federal Holiday

HOLDINGS: The PDR maintains over 2.6 million agency documents in microfiche and, selectively, electronic full text (diskette and online). The majority of these documents relate to the licensing and inspection of nuclear facilities and to the use, transport, and disposal of nuclear materials. NUREG/BR-0052, "PDR File Classification System", gives a more detailed listing of PDR holdings. It is advisable to call the PDR in advance if a research project involves reviewing paper copies because most of the paper collection is now in off-site retrievable storage.

REFERENCE ASSISTANCE: The PDR is open to the public, and PDR reference librarians will assist on-site researchers. The reference staff also responds to requests received by telephone, Internet, fax, and mail. Although staffing constraints prohibit PDR staff from performing extensive research on behalf of users, the staff will assist the public in doing their own research.

AGENCY DOCUMENT MANAGEMENT SYSTEM (ADAMS): The ADAMS database contains citations and full-text to public documents. The ADAMS legacy library contains citations and some full-text to documents released prior to October 31, 1999. The ADAMS Publicly Available Records System (PARS) library contains full text documents from October 31, 1999 forward.

The PDR staff will assist clients in formulating search strategies and provide them with printouts of results. The PDR does not charge for computer searches, but does limit the size of free printouts. The staff will train clients in the use of ADAMS. A Users' Guide is available at the NRC website (http://www.nrc.gov/reading-rm/adams.html). Documents can be printed or downloaded from ADAMS by the user or ordered, for a fee, from the PDR staff.

DOCUMENT REPRODUCTION SERVICE: The PDR on-site reproduction contractor copies, with some exceptions, PDR documents in paper, microform and electronic full text for a fee. Fees and turnaround times vary depending on the amount and kind of material. See the current price schedule for more information (http://www.nrc.gov/reading-rm/pdr/fee-schedule.html). Reproduction orders can be picked up on-site or mailed to the requestor by regular or express mail. Facsimile service is also available. A self-service duplicating machine and coin-operated reader-printers are available at the PDR for users who need immediate copies of documents.

PDR USERS' GUIDE: For more information about the PDR, contact the reference librarians to order a free copy of the "Public Document Room Users' Guide", or access the most recent version on the Internet at: http://www.nrc.gov/reading-rm/pdr.html.

NRC HOME PAGE: For information about NRC and its mission as well as electronic versions of many NRC documents, visit the NRC web site, http://www.nrc.gov .

PDR Document Reproduction Fee Schedule

The Nuclear Regulatory Commission (NRC) Public Document Room (PDR) contractor, Qualex International, will provide document reproduction services at the PDR's Headquarters location at the rates listed below. Relevant sales tax and postage will be added. Please note that the hours indicated for turnaround times are work hours (7:45 a.m. to 4:15 p.m. Eastern time) unless otherwise noted as clock hours. On this page:

- Standard Processing
- Rush Processing
- Payment Instructions

Standard Processing

Standard charges and turnaround times are as follows:

CD-ROM to paper and paper to paper (up to and including 11"x14") reproduction is $0.15 per page with a 4 hour turnaround time for orders of 1-3000 pages. For 3000+ pages the turnaround time is 24 clock hours.

Pages 11"x17" (paper to paper) are $0.30 per page with a 4 hour turnaround time for orders of 1-

1000 pages. Pages larger than 11"x17", including drawings, are $2.50 per square foot. For 1-100 drawings the turnaround time is 24 clock hours. Pages greater than legal size, 81/2"x14", but smaller than or equal to 11"x14", shall be reduced to legal size unless the order specifically requests full size reproduction.

Diskette to diskette reproduction (1-100 diskettes) is $3.00 per diskette with a 1 hour turnaround time.

Microfiche to paper reproduction is $0.15 per page. There is a 6 hour turnaround time for reproduction of 1-3000. More than 3000 pages is a 24 clock hour turnaround time. Aperture card to paper reproduction is $2.50 per square foot. Turnaround time for 1-500 cards is 72 clock hours.

ADAMS full-text electronic files to paper reproduction are $0.15 per page with a 4 hour turnaround time for 1-3000 pages and 24 clock hours for anything above.

ADAMS full-text electronic files to diskette (1-10 diskettes) are $3.00 per diskette with a 4 hour turnaround time.

Microfiche or aperture card (1-300 cards) duplications are $0.75 for each card with a 3 hour turnaround time. Over 300 cards is a 24 hour clock time turnaround.

Color drawings (81/2"x11") are $2.00 per page. Turnaround time for 1-100 drawings is 24 hours. Larger drawings are $12.00 per square foot. Turnaround time for 1-10 larger drawings is 72 clock hours. It is the **customer's responsibility** to indicate which pages are to be reproduced in color. If no preference is indicated, drawings will be reproduced in black and white.

ADAMS electronic documents to CD-ROM are $10.00 for the first document and $5.00 for each additional document on the same CD. The turnaround time is 48 hours. CD-ROM to CD-ROM duplication is $10.00 per CD. The turnaround time for 1-10 CD's is 24 clock hours.

Photographic reproduction of slides and negatives is $5.00 per slide (1-10 is a 72 clock hour turnaround) and photographs are $10.00 per print (1-10 is a 24 clock hour turnaround).

Audio tape reproduction is $3.00 per tape (1-10 tapes is a 6 hour turnaround) and video cassette reproduction is $15.00 per cassette with a 6 hour turnaround time.

Rush Processing

Rush Processing is offered for standard size paper to paper reproduction, microfiche to paper and electronic file to paper reproduction, excluding pages reproduced from bound volumes. The charge is $0.20 per page. A request for rush processing must be indicated when placing the order. For ***rush order processing*** the charges and turnaround times are as follows:

CD-ROM to paper (1-1000 pages) and paper to paper (1-3000 pages)(up to and including 11"x14") reproduction is $0.20 per page and the turnaround time is 1 hour.

Diskette to diskette reproduction (1-50 diskettes) is $5.00 per diskette with a ½ hour turnaround time.

Microfiche to paper reproduction (1-3000 pages) is $0.20 per page with a 3 hour turnaround time. Aperture card to paper reproduction (1-500 pages) is $3.50 per square foot with a 24 hour clock time turnaround.

ADAMS full-text electronic files to paper reproduction (1-1000 pages) are $0.20 per page with a 1 hour turnaround time.

ADAMS full-text electronic files to diskette (1-10 diskettes) are $5.00 per diskette with a 1 hour turnaround time.

Microfiche or aperture cards duplications (1-300 fiche/cards) are $1.00 for each card with a 1 hour turnaround time.

Color drawings (1-100 pages) (81/2"x11") are $2.50 per page with a 4 hour turnaround time. Larger drawings (1-10) are $15.00 per square foot with a 24 hour clock time turnaround. It is the **customer's responsibility** to indicate which pages are to be reproduced in color. If no preference is indicated, drawings will be reproduced in black and white.

ADAMS electronic documents to CD-ROM are $15.00 for the first document and $7.50 for each additional document on the same CD with a 6 hour turnaround time. CD-ROM to CD-ROM (1-10 CD's) duplication is $15.00 per CD with a 2 hour turnaround time.

Audio tape reproduction (1-5 tapes) is $5.00 per tape and video cassette reproduction is $30.00 per cassette with a 3 hour turnaround time.

Facsimile Service (unlimited pages) is available. The turnaround time is 1 hour and the fees are: Local fax-$0.30 per page; Long Distance-U.S.-$0.50 per page; and Long Distance-Foreign-$1.00 per page.

Orders completed for mailing by the U.S. Postal service or by private delivery service will include an additional fee for the actual mailing, shipping or delivery service rate. Unless a user requests special packaging materials, there is no additional charge by the contractor for wrapping materials and handling.

Payment Instructions

Payment for document reproduction services is made directly to the contractor. Customers may pay by any of the following methods: cash; Visa, MasterCard, or Discover credit card. Orders can be submitted for customers who wish to pay by check. The PDR staff will inform the customer of the total charges for the order and a check may be submitted to Qualex International. Upon receipt of the check, the customer's order will be delivered as requested.

Repeat customers are encouraged to establish a deposit account. Invoices are issued with each order. Customers who order frequently and have a deposit account will receive a monthly statement. This monthly usage report will be mailed showing the customer name, billing period, status of account, and the usage activity for that period. **Customers with accounts should include their contractor assigned customer number on every order.** Service will be suspended for those customers with zero amounts in their deposit accounts. Returned checks will place accounts on hold and will incur a $25.00 penalty. ***Payment should be mailed to***: Qualex International at P.O. Box 255, Rockville, MD 20848-0255.

For information regarding billing and opening a deposit account, contact Accounts Receivable, Qualex International, at 301-610-5374 (fax: 301-610-5375). For information regarding the status of an order, customers may call 301-770-8952 (fax: 301-770-8954).

The contractor does not provide reference services and does not directly accept orders by U.S. Mail, fax, e-mail, or telephone from the public. Orders may be placed through the PDR reference librarians or through the online order module. Please note that information collected through the online ordering of documents from the NRC Public Document Room (PDR) is not maintained in a system of records. The requested information is used solely for shipping and billing purposes and notification of any contractor/rate changes.

If you have Questions or comments related to document duplication services please contact the Public Document Room Staff.

Site Help | Site Index | Contact Us [Search] Advanced Search

U.S. Nuclear Regulatory Commission

The Nuclear Regulatory Commission regulates U.S. commercial nuclear power plants and the civilian use of nuclear materials

Home | Who We Are | What We Do | Nuclear Reactors | Nuclear Materials | Radioactive Waste | Public Involvement | Electronic Reading Room

The **Nuclear Regulatory Commission (NRC),** which is empowered to protect public health and ~~safety as it~~ regulates the various commercial and institutional uses of nuclear energy, has redesigned its Web site (www.nrc.gov) to provide a consistent, easy-to-understand presentation of information about the agency and its activities.

The redesigned Web site has two major objectives:

Increase Public Confidence by providing information that:

- **Enhances the ability of stakeholders to participate** effectively in the regulatory process. Examples include public meeting notices, documents for comment, rulemaking information, enforcement actions, reportable events, and much more.
- **Broadens the public's understanding of NRC's mission, goals and performance.** Examples include information about the Commission's activities, the agency's strategic and performance plans, its organization and functions, and educational materials about nuclear topics.

Make Doing Business with the NRC Easier by:

- **Enhancing access to site information** through an improved search engine, consistent page design, expanded site topic indices, standardized navigational capabilities, and improved access for users with disabilities.
- **Making tools available for conducting business electronically at** www.nrc.gov. For instance, the site permits secure electronic transmittal of documents, lists job vacancy announcements, procurement plans for prospective contractors, licensing information for applicants, and NRC forms for submitting information.

Redesign Highlights

Overview

The new Web site's consistent look and feel throughout make navigation simple and enhance access to the multitude of agency documents and information available at the site. Information about the agency and its regulatory activities is arranged in a general-to-specific pattern. Diagrams, maps, and photographs augment the text to aid clarity. The timeliness and accuracy of information at the site are monitored to keep our commitment to the quality and integrity of the site.

Home Page

The redesigned home page permits one-click access to topics of high stakeholder interest. For example, a Key Topics box highlights site-wide topics of current interest. The buttons across the top of the home page include drop-down menus with direct access to more than 50 pages at the site. On the left side of the home page, you will find short cuts for accessing important links and in the center you'll find links to "What's Happening" and recent press releases.

The site also provides real-time Webcasts of Commission meetings open to the public. This feature is designed to foster familiarity with the agency and its scope of responsibilities, as well as to encourage increased public participation in the regulatory process.

Major Site Content

The redesigned site is organized into **seven** major areas to help users find agency information of interest quickly:

Who We Are - describes the NRC's mission, functions, organization, history, office locations, and funding, as well as information for those who wish to apply for employment or compete for NRC contracts.

What We Do - provides details about NRC's regulatory activities, Congressional and public affairs programs, and programs of nuclear safety involving States and the international community.

Nuclear Reactors - features a wealth of information on the types of nuclear power plants regulated by the NRC, the oversight process used to ensure safe plant operations, plant performance data, and agency licensing requirements to operate a reactor or to close a plant permanently.

Nuclear Materials - provides information on the nuclear materials regulated by the NRC that are used in industry, medicine, and academic settings and the regulations for using, transporting, and storing these materials.

Radioactive Waste - describes the types of radioactive waste regulated by the agency including low-level waste (such as contaminated protective clothing, tools, filters, rags, medical tubes), high-level waste (such as "irradiated" or used nuclear reactor fuel), and uranium mill tailings (the residue that remains after the processing of natural ore to extract uranium and thorium). It also includes information about the storage, transportation, and disposal of these wastes.

Public Involvement - tells users about how to become involved in the agency's regulatory process. This section conveniently links to pages on the site that provide opportunities to comment on documents, request enforcement action by the agency, or request that the NRC change or establish a regulation. Site users can find notices of public meetings and ways to contact someone at the NRC.

Electronic Reading Room - provides access to collections of current and archived documents pertinent to NRC's regulatory activities. (The NRC releases several hundred documents to the public each work day.) Many of these documents are available directly on the Web while the rest are available either from the NRC's document retrieval system, ADAMS, or from the agency's Public Document Room. This site also contains information on obtaining documents in accordance with the Freedom of Information Act (FOIA), as well as an index to all Frequently Asked Questions pages throughout the site.

These seven areas, as well as the home page, can be accessed from virtually every Web page. In addition, there are a series of links at the top of each page that show users where they are on the site and provide a path back to the NRC home page.

For example, the links at the top of the Operating Reactors page would read:

Home > Nuclear Reactors > Operating Reactors

Found at the top of each page above the site banner, the *Contact Us*, *Site Help*, and *Site Index* links and the *Search box* provide additional aids for locating information and getting help from NRC staff. Users will also find helpful icons throughout the site, such as **exit** EXIT, to inform them they are leaving the site, and **PDF** to indicate that they will need to download a "portable document format" reader (free of charge) to read the linked document.

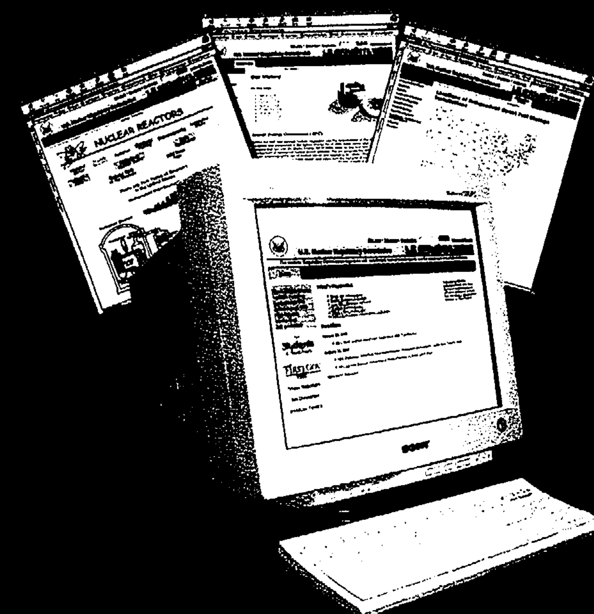
The NRC encourages you to visit our redesigned Web site, at http://www.nrc.gov, and welcomes your ideas for improving the site.

Send your comments to nrcweb@nrc.gov

Office of Public Affairs
U.S. Nuclear Regulatory Commission
Washington, D.C. 20555-0001
301-415-8200
opa@nrc.gov

NUREG/BR-0294
March 2002

Visit NRC's Redesigned Web Site



www.nrc.gov

Evaluation of the Current DOE Document Conversion System: A Study of Retrievability

Information Science Research Institute
University of Nevada, Las Vegas

May 2002

CONTENTS

Evaluation of the Current DOE Document Conversion System: A Study of Retrievability

Information Science Research Institute
University of Nevada, Las Vegas

May 2002

1.0 Background

During the 2001/2002 fiscal year, the Information Science Research Institute (ISRI) at the University of Nevada, Las Vegas (UNLV) has been tasked to suggest improvements and evaluate the performance of the current DOE document conversion system.[1] This report gives a summary of the recommendations made by ISRI staff and a summary of the results of two types of performance tests.

There are two approaches to evaluating the performance of document conversion systems. One approach is to measure the accuracy of the textual output (i.e., average character accuracy) of the system. A second approach is to measure the performance of the system that will make use of the output text. In this case, textual output will be used to build the index for an Information Retrieval (IR) system that will aid in the task of finding documents of interest. The appropriate performance measure for IR systems is retrievability (i.e., precision and recall). Thus, to provide a thorough evaluation of system performance, two different studies (a character accuracy study and a retrievability study) have been conducted. [1, 2] Section 3.1 below, gives a summary of the results of accuracy tests and Section 3.2 summarizes the results of the retrievability tests.

2.0 Document Conversion System Recommendations

The task of document preparation for the LSN has two major components: *character recognition* and *page zoning*. The task of loading the text produced into an information retrieval system is, by comparison, straightforward and not error prone. Thus, in any document conversion system, character recognition and page zoning are performance-controlling operations.

2.1 The Importance of Measuring Retrievability

Although character recognition is typically measured by standard character accuracy, many characters in a document's text have no role in its retrievability. For example, punctuation marks, end-of-line hyphenation, and characters in stopwords[2] are ignored by an IR system. The top ten standard stopwords account for about 20 to 30 percent of all words in any collection. Thus, while character accuracy is related to retrievability, it is not a good measure of retrievability. For these reasons, and because retrievability is more important to users of the LSN, the retrievability testing described in Section 3.2 was recommended by ISRI staff.

1 The system of concern is a computer driven character recognition system used to create an electronic (i.e. electronic text) copy of paper documents. This system includes the Scansoft SDK 2000 OCR system running on IBM compatible equipment and is operated for the DOE by the Bechtel-SAIC Company, LLC (BSC) in Summerlin, Nevada.

2 Stopwords are common English words with no information content such as "the" and "and."

2.2 The Use of Automatic Zoning

Page zoning can be done either manually by drawing a box around text to be captured or automatically by an OCR engine. Manual zoning not only necessitates thousands of hours of manpower, it also requires a pre-defined set of *zoning rules* which we have found to be error prone. [3] Automatic zoning on the other hand is performed by the recognition system and, although not always 100% accurate, captures all the data required for information retrieval.

During the 1990's, ISRI conducted a series of experiments comparing retrievability from manually zoned collections to retrievability from automatically zoned collections. In every experiment, the use of automatic zoning followed by MANICURE [4] gave retrieval results equivalent to what one could expect from manually zoned pages, even from a nearly perfect collection. Based on this experience, ISRI recommended that DOE employ automatic zoning and the MANICURE post-processing system. The studies reported in Section 3 were performed to compare both automatic and manual zoning with and without MANICURE post-processing to support the ISRI recommendation.

3.0 Results of Tests to Measure Conversion Performance

The LSN is the discovery database to be used in the licensing proceedings and the DOE documents will be an important component of the LSN. The performance of the conversion system is therefore very important to the operation of the LSN. It is therefore appropriate to thoroughly test the DOE system to insure that users of the LSN will have the best technology available to enable them to find documents of interest.

To be realistic, such tests must be based on sample sets of actual documents that will be submitted to the LSN. To be reliable, tests must use well-accepted standards and scientific methods. To be statistically significant, tests must be based on reasonably sized, random samples of DOE documents.

3.1 Accuracy Tests

The technical requirements for document collections submitted to the LSN processed through an OCR system have included a "target character accuracy of 99.5%." ISRI has designed and conducted a test to measure average character accuracy of documents processed by the conversion system. Because IR systems ignore stopwords and punctuation marks, this test focused on non-stopword accuracy and on the character accuracy of non-stopwords. MANICURE post-processing of the OCR output is a part of the conversion system. The text accuracies were measured both before and after application of the MANICURE system.

All accuracy tests were conducted with a random sample of 17 Microsoft word documents from the current DOE collection. The total number of non-stopwords in these documents is 164,483. The total number of characters in these words is 1,361,124. Because these documents were native Microsoft word files, they provided almost perfect images not typical of the DOE collection. To reflect realistic image quality, we also added first through fourth generation photocopies of these documents to our study. The error counting programs used are a modification of the OCR performance metrics developed by ISRI staff in the early 1990's [5].

3.1.1 Word (Non-stopword) Accuracies

The most important result of our word accuracy tests was that non-stopword accuracies produced by MANICURE (see Table 1) were uniformly higher that those produced by the OCR system alone. Because the average non-stopword is eight characters long, word accuracies are always lower than character accuracies. (If the average character accuracy for this collection were 99.5%, and if each character error were in a different word, the average word accuracy would be 95.9%.)

Another result is that the increased accuracy of MANICURE output over OCR output improves as page quality decreases (i.e., MANICURE helps the poor quality images more). Because our best judgement of the average print quality of DOE documents is between that of a first and second generation photocopy, these results indicate that the word accuracy produced is between 97.23% and 96.15% correct. It is also important to note that a 1% increase in word accuracy for a 97% correct page corresponds to correcting 33.3% of the word errors.

System	% Word Accuracy of	Orig.	Gen 0	Gen 1	Gen 2	Gen 3	Gen 4
Raw OCR Output	All non-stopwords	97.44	97.03	96.45	95.05	92.78	91.91
MANICURE Output	All non-stopwords	98.01	97.54	97.23	96.15	94.64	94.14

TABLE 1. Average Non-stopword Accuracy for All 17 Documents

3.1.2 Character Accuracy (of Non-stopwords)

As with word accuracies, the character accuracies of non-stopwords produced by MANICURE (see Table 2) were uniformly higher that those produced by the OCR system alone.

System	% Character Accuracy of	Orig.	Gen 0	Gen 1	Gen 2	Gen 3	Gen 4
Raw OCR Output	All non-stopwords	99.37	99.40	99.20	98.67	97.89	97.79
MANICURE Output	All non-stopwords	99.50	99.47	99.30	98.83	98.16	98.06

TABLE 2. Average Character Accuracy for All 17 Documents

Because our best judgement of the average print quality of DOE documents is between that of a first and second generation photocopy, these results indicate that the character produced is between 99.30% and 98.83% correct.[3]

3.2 Retrievability Tests

The second series of tests [2], was designed to address retrievability of documents produced by the current DOE conversion system from the Autonomy search system. In part one of this test, we designed an experiment to compare retrievability from document collections that had been manually zoned (**manual zoned**) to retrievability from the identical document set that had been automatically zoned (**auto zoned**). The idea here is to determine if retrievability from auto zoned collections (i.e., zoned by the OCR system) is as good as retrievability from collections that were zoned by human operators.

Another consideration is the order in which documents are returned by the Autonomy system. In part two of this test, we compared the ordering of retrieved documents from the manually zoned and from the automatically zoned collections. In this study, and in the first retrieval test described above, we used a 1055 document subset of the DOE collection with 40 queries, "typical" of queries likely issued to the LSN, and relevance judgements for each query for each document.

Finally, we designed a test to compare retrievability from a collection of documents that were 99.8% correct to retrievability from the same set that were auto zoned with MANICURE post-processing. The idea here is to determine if retrievability from collections produced with auto zoning & MANICURE is as good as retrievability from collections that are close to 100% correct (i.e., in this case 99.8% character accuracy). In this study, we used 1058 documents from the LSS prototype collection with 68 queries, again "typical" of queries likely issued to the LSN, and relevance judgements for each query for each document.

3 In conducting the accuracy test [1], an additional level of character accuracy was measured. The accuracy of "unique" non-stopwords was slightly higher than these percentages.

3.2.1 Retrievability from Manually vs. Automatically zoned Collections

Table 3 shows the average precision for the manual zoned and the auto zoned collections. Although average precision from the auto zoned collection differs by 3.5% from the manual zoned collection, this difference cannot be considered statistically significant. Basically, we can only conclude that retrievability between these two collections is equivalent.

Precision	Manual zoned	Auto zoned
Average	0.379	0.392

Table 3: Average Precision for Manually Zoned vs. Automatically Zoned Collections

3.2.2 Ranking from Automatically Zoned Collections

Another consideration, related to retrievability, is ranking. This is important because it makes a difference to the user if a relevant document is ranked 3rd or 300th. Thus, we also performed a detailed study to determine if any rank variability exists between retrieval results from manual zoned and auto zoned collections. Some IR systems available in the mid-1990's showed variability in ranking output of documents with automatically zoned collections. In fact, ISRI research during this period played a role in correcting this problem.

3.2.2.1 Ranking Problems in Information Retrieval Systems

Ranking variability in optically recognized documents is typically due to the concept of *document length normalization*. IR engines use normalization to treat short and long documents equivalently. Long documents generally have more distinct words than short documents. Also, long and short documents about the same subject matter may have the same set of distinct words, but the frequencies of these words are much higher for longer documents. IR systems typically use the number of distinct words, or the maximum frequency of words, to adjust the weights of terms in the documents in order to give fair representation to words in shorter documents. In optically recognized documents, mis-recognized words inflate the number of distinct words and the maximum frequency. The fact that this inflation leads to ranking variability was first pointed out by Taghva in 1994 and 1996 [6, 7].

Fortunately, after this discovery, the concept of length normalization was revisited by Prof. Gerard Salton and his student Amit Singhal at Cornell University. Their efforts led to redefining length normalization [8]. New measures were defined that depend on the byte size of the document (i.e. the number of characters in the file) eliminating extreme rank variability. Modern IR engines either use the new measure, or a similar concept, which is not affected by misrecognized words.

3.2.2.2 Results of Ranking Test

We summarized the ranking for both manual zoned and auto zoned collections by average relevant document rank and standard deviation. We further calculated the correlation coefficient between the ranks of the same documents for both the manual zoned and auto zoned versions. These values appear in Table 4. From the Table, we note that average relevant document rank and standard deviation for these sets are exactly equal. This is the first indication that document ranking in these two sets is similar. But it's actually the correlation coefficient $r = 0.97$ that convinces us the ranking in these two sets is very close. The correlation coefficient ranges between 0 and 1, where the r value of 1 indicates identical correlation; $r = 0.97$ says there is a very strong association between the ranked lists. A scatterplot pictorially shows this relationship. Figure 1 shows how the ranks of the manual zoned and the auto zoned tightly cluster around the regression line that begins at the origin.

Average Rank Automatic	Standard Deviation Automatic	Average Rank Manual	Standard Deviation Manual
289	258	289	258
Correlation coefficient r = 0.97			

Table 4. Average Rank and Standard Deviation for Auto Zoned and Manual Zoned documents

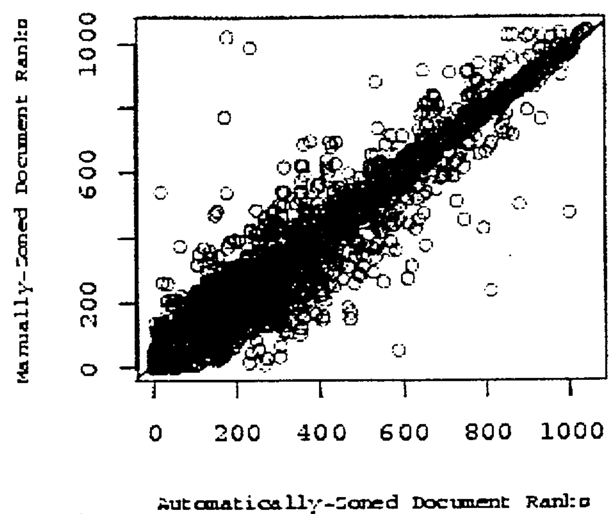


Figure 1. Scatter plot of Automatically Zoned vs. Manually Zoned Ranks for all Documents

3.2.3 Retrievability from 99.8% Correct vs. Auto zoned & MANICURED Collections

Our final comparison test ties the OCR accuracy tests to retrieval. As mentioned in [1], building a large collection of OCR ground-truth data is an arduous task. The requirement for exact duplication in ASCII of an image of a page is difficult to obtain for even a small set of pages.

Fortunately, ISRI has in its collection a 99.8% correct set of documents with queries and relevancy judgments that had been prepared for the LSS Prototype. This is not equivalent to OCR ground-truth since carriage-returns, spacing, and columnization are not equivalent to the hard copy page. Still, the typed text was measured to be 99.8% correct. Since this is higher than the current goals set by the NRC, we felt by comparing the results of the auto zoned version to this 99.8% correct version, we could help clarify if improving character accuracy would improve retrieval results. The average precision for both sets appears in Table 5.

Precision	99.8% correct	Auto zoned with MANICURE
Average	0.245	0.242

Table 5: Average Precision for 99.8% Correct Text vs. Automatically Zoned and Recognized Text

Note that average precision differs by only 1.2% (i.e. retrievability is statistically equivalent). We can therefore conclude that the process used by DOE to prepare documents for the LSN will return results equivalent to a collection whose character accuracy was corrected to 99.8%.

4.0 Summary

There are several things that can be concluded from these studies:

1. The character accuracy produced by the DOE conversion system is close to NRC requirements. For good quality images it is exactly 99.5% (see MANICURE character accuracy for the original image in Table 2).

2. The effect of MANICURE on character and word accuracies is uniformly positive.

3. Retrievability from automatically zoned collections is equivalent to retrievability from manually zoned collections.

4. Result ranking from automatically zoned collections is equivalent to ranking from manually zoned collections.

5. Retrievability from automatically zoned and MANICURED collections is equivalent to retrievability from 99.8% correct collections.

In conclusion, we believe that the combination of these accuracy tests and retrieval tests, demonstrate that the quality of the documents delivered by the DOE will give effective retrieval results for the users of the LSN.

References

[1] T. Nartker and R. Young, OCR Accuracy Produced by the Current DOE Document Conversion System, Technical Report 2002-06, Information Science Research Institute, University of Nevada, Las Vegas, May 2002.

[2] K. Taghva, J. Borsack, S. Lumos, and A. Condit, Retrievability of Documents Produced by the Current DOE Document Conversion System, Technical Report 2002-05, Information Science Research Institute, University of Nevada, Las Vegas, May 2002.

[3] K. Taghva, J. Borsack, A. Condit, and S. Erva, The Effects of Noisy Data on Text Retrieval, *J. American Society for Information Science*, 45(1):50-58, January 1994.

[4] K. Taghva, A. Condit, J. Borsack, J. Kilburg, C. Wu, and J. Gilbreth, The MANICURE Document Processing System, Proc. IS&T/SPIE 1998 Intl. Symp. on Electronic Imaging Science and Technology, 1998, San Jose, CA, January.

[5] S. Rice, F. Jenkins, and T. Nartker, The Fifth Annual Test of OCR Accuracy, ISRI TR-96-01, April 1996.

[6] K. Taghva, J. Borsack, and A. Condit, Results of Applying Probabilistic IR to OCR Text, Proc. 17th Intl. ACM/SIGIR *Conf. On Research and Development in Information Retrieval,* 1994, pp. 202-211.

[7] K. Taghva, J. Borsack, and A. Condit, Effects of OCR Errors on Ranking and Feedback Using the Vector Space Model, *Inf. Proc. and Management*, 1996, 32(3), pp. 317-327.

[8] A. Sinhal, G. Salton, and C. Buckley, Length Normalization in Degraded Text Collections, SDAIR 1996, pp. 149-162.

Retrievability of Documents Produced by the Current DOE Document Conversion System

Kazem Taghva, Julie Borsack, Steve Lumos, and Allen Condit

Technical Report 2002-05
Information Science Research Institute
University of Nevada, Las Vegas

April 25, 2002

1 Introduction

The Licensing Support Network (LSN), managed by the Nuclear Regulatory Commission (NRC), will provide information to all interested parties that is potentially relevant to the licensing of the high-level radioactive waste repository proposed for Yucca Mountain. There are several organizations contributing documents to the LSN, but DOE as the licensee, will submit the vast majority of these documents. Since the document collection will be very large, the identification and retrieval of documents must be timely and effective. Whatever search method is used, it should produce the expected results.

What's more, effective retrievability should be of the utmost concern. Whatever capture method is selected, it should produce the expected results. The retrievability studies we perform and explain in this report evaluate and compare automatic and manual methods that can be used for document conversion of text from images. These results will enable the DOE to make a well-informed decision for their document conversion processes. We report on optical character recognition's (OCR) impact on retrieval for the following:

- Manual zoning vs. automatic zoning (Section 2).
- Differences in ranked results between manually-zoned and automtically-zoned OCR text (Section 2.3).
- The DOE's document conversion methods and compare them to nearly perfect (99.8% correct) text (Section 3).
- Post-processing methods applied to improve the OCR text produced (Appendix B).

These tests, in conjunction with the results reported in [2], will provide an in-depth analysis of the performance of the current DOE document conversion system.

2 Manual Zoning vs. Automatic Zoning

2.1 Environment

The zoning experiments described in this report required that ISRI produce an environment that duplicates the systems and procedures applied by the DOE to prepare their documents for the LSN. Further, the documents and queries we use for these experiments should be a good representation of the expected LSN collection and its anticipated use. The system and procedures we applied in all the following tests are exactly what the DOE and NRC had in place at the date of testing:

OCR: Scansoft v10 with the MTX OCR module

Post Processing: MANICURE v1.7

Retrieval Engine: Autonomy Server v2.2.0

The test collection that we use to compare manual vs. automatic zoning consists of 1055 documents that were selected from documents in the RIS with the document type "Report," "Plan," "Design Document," or "Correspondence." Manual zoning information for each page was made available to us from the zoning procedures conducted by the Yucca Mountain Project management and operations contractor prior to 1999. Forty queries, with relevancy judgments were produced by UNLV geology students who were familiar with the RIS collection. Table 1 shows some statistics for this dataset.

Document count	1055
Number of pages	75,236
Query Count	40
Average number of relevant documents/query	100
Median number of relevant documents/query	64
Fewest relevant documents for a query	2
Most relevant documents for a query	608

Table 1: ISRI Collection Statistics for Auto-Zoned vs. Manual-Zoned Test

2.2 Recall/Precision Results

We produced two versions of the 1055 document collection described in Section 2.1. For one version, we applied the manual zoning information described above (call it `manual-zoned`). We also produced another version applying automatic zoning performed by Scansoft (call it `auto-zoned`). Both collections were loaded and indexed into Knowledge Server. All 40 queries were batch run against these two datasets in exactly the same way.

Recall Points	Manual-Zoned	Auto-Zoned
0.00	0.78	0.84
0.10	0.58	0.61
0.20	0.50	0.54
0.30	0.44	0.46
0.40	0.41	0.42
0.50	0.36	0.35
0.60	0.31	0.31
0.70	0.27	0.26
0.80	0.22	0.22
0.90	0.18	0.18
1.00	0.12	0.12
Avg	0.379	0.392

Table 2: 11-Point Precision for Manual-Zoned Vs. Auto-Zoned Sets

The objective is to compare these two result sets against each other. *Recall* and *precision* are the accepted measures applied in the IR community for comparing retrieval results. In Table 2, recall percentages are shown in the left most column with the corresponding precision values at these recall points. For example, when the system has returned 20% (0.20 recall) of the documents in the collection, 50% of those returned in the `manual-zoned` set were relevant and 54% of the `auto-zoned` were relevant. The precision values represent *averages* across all queries. The last row of Table 2 is the average of the precision values in the columns. For a more complete discussion of recall and precision, see Appendix A. Table 2 shows precision at 11 recall points for both the manual-zoned and the auto-zoned versions of this collection.

Note that the 11-point average for auto-zoned is 3.5% better than for manual-zoned. This higher average return for auto-zoned indicates that running these queries against this data set gives better results from automatic zoning than one could expect from manual zoning. This difference is not statistically significant though. To be statistically significant for this size collection, the difference would have to be 5% or greater. What we can learn from these results is that in general, automatic zoning gives results as good as those obtained from manual-zoned OCR.

2.3 Ranked Query Analysis

The 11-point precision average indicates that there is no difference in query results for the methods used for collection preparation. On the micro level, we felt it important to investigate what exactly happens to individual query rankings. In other words, if a relevant document was ranked, say 25, in the automatic results, what would be the rank of the same document in the manual version? By reviewing the query-by-query results for both versions, we observed that there was no significant variation between rankings. The following paragraphs are the technical details supporting our observation.

The collection has 1055 documents. Hence, for a specific query, a relevant document can be ranked between 1 and 1055. Obviously, we would like to see the relevant documents ranked as close to 1 as possible. Now consider all the relevant documents that were ranked, say 3, for the automatic version. We may ask, what are the rankings of these document in the manual version? Table 3 shows all the relevant documents ranked 3 in automatic and the corresponding rank of the same documents in the manual-zoned version. We can represent these points as as (3,6), (3,1), (3,5), etc.

Queryid	Docid	Auto-Zoned Rank	Manual-Zoned Rank
KW2-Q01	MOL.19990701.0270	3	6
KW2-Q03	MOL.19981008.0009	3	1
KW2-Q04	HQO.19950224.0009	3	5
KW2-Q06	NNA.19870625.0060	3	5
KW2-Q11	NNA.19920504.0221	3	35
KW2-Q13	MOL.19981009.0176	3	2
KW2-Q14	MOL.19980609.0061	3	10
KW2-Q17	MOL.19980123.0860	3	19
KW2-Q18	NNA.19870331.0563	3	3
KW2-Q19	MOL.19980122.0032	3	11
KW2-Q21	MOL.19990702.0236	3	5
KW2-Q23	MOL.19981008.0006	3	3
KW2-Q29	NNA.19920528.0154	3	135
KW2-Q31	MOL.19980716.0493	3	2
KW2-Q32	MOL.19981008.0002	3	3
KW2-Q33	MOL.19980729.0051	3	5
KW2-Q39	MOL.19980729.0047	3	3
KW2-Q40	MOL.19980724.0391	3	9

Table 3: Ranks of Auto-Zoned and Manual-Zoned Relevant Documents

If we continue this process for all the ranks and plot these points, we will discover the scatter plot in Figure 1. This graph exhibits the relationship between the corresponding rankings. In other words, for a fixed rank m on the x axis, the auto-zoned ranks, then the y values, the manual-zoned ranks, represent the correspon·ling ranks for the same documents for the manual-zoned. We can summarize this graph by average ranks, standard deviation (SD), and the correlation coefficient r as shown in Figure 1. The r value shows the strength of the association between the two variables. The r value ranges between 0 and 1. The

Auto Avg	Auto SD	Manual Avg	Manual SD
289	258	289	258
correlation coefficient $r = 0.97$			

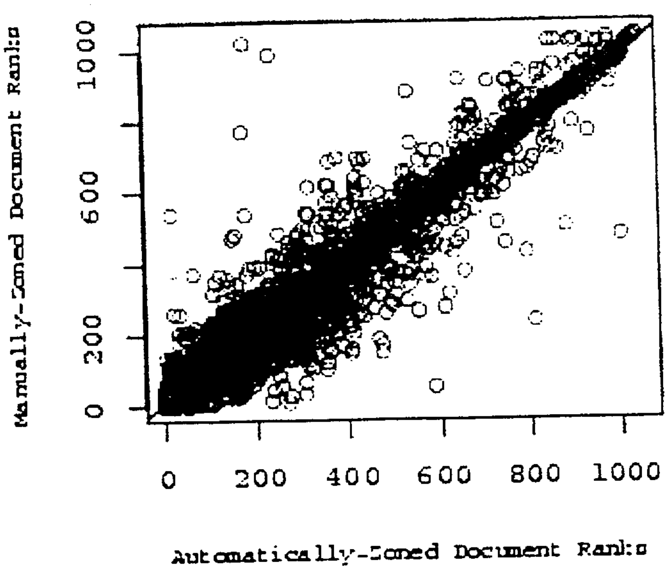


Figure 1: Auto-Zoned Ranks vs. Manual-Zoned Ranks Scatter Plot for All Document Ranks

closer r is to 1, the stronger the association.

We can use this plot to draw the *regression line* and use the *regression method* to predict the rank of the manual ranks from the automatic ranks. The solid line in Figure 1 represents the regression line.

$$manual\_average - \left[\left(\frac{auto\_average - auto\_rank}{auto\_SD} \right) (r)(manual\_SD) \right] = manual\_rank \qquad (1)$$

As can be seen in the scatter plot, the points in this graph are tightly clustered around the regression line. This clustering indicates a strong linear association between the two variables. In general, we use Equation 1 for calculating the predicted manual ranking of the same document in the manual version. Table 4 shows examples of the ranks of documents in `manual-zoned` when the `auto-zoned` rank is known.

Complete Collection, Predicted Ranks	
If Auto-Zoned Rank is:	Predicted Manual-Zoned Rank is:
200	203
150	154
100	106
50	56
10	19
350	349
400	397
First Quadrant, Predicted Ranks	
If Auto-Zoned Rank is:	Predicted Manual-Zoned Rank is:
50	57
25	32
5	13
150	155

Table 4: Predicted Manual-Zoned Ranks using the Regression Method

We can use the above analysis to extrapolate the rank performance of `manual-zoned` from the `auto-zoned` version. As can be seen from Table 4, the further away we are from the point of average the bigger difference we see in the ranks. This is the way the regression method works. To lend more credence to our analysis, we did the same calculation for the first quadrant ranks (i.e. only the ranks between 1 and 261), since these are the documents that the user will most likely evaluate. The graph in Figure 2 represents the correspondence between these ranks. The second half of Table 4 shows examples of the predicted ranks for the `manual-zoned` version for some rankings just in the first quadrant.

The regression method is a scientific way of comparing the ranking correspondence between the two collections, or in this case between two versions of the same collection. In our experiments, it can be seen that there is no significant difference in the ranking between the two methods of document preparation.

3 A Comparison Using the Prototype Collection

One might believe that the closer we get to 100% character accuracy, the better the retrieval results we will obtain from a search engine like Autonomy. In fact, one of the goals specified by the NRC is that collections submitted for the LSN should try to reach 99.5% character accuracy across the collection and 98.5% for any particular page. What this next experiment shows (and several other experiments performed at ISRI have shown)[8, 5, 1, 6, 7] is that close to 100% character accuracy may not be necessary for good retrieval performance.

We have a collection of 1058 documents, 62 queries, and 1104 relevancy judgments that we will use to answer this question. This collection is particularly well-suited for determining how character accuracy may affect retrieval performance for a couple of reasons. First, we have two versions of the collection: one version

Auto Avg	Auto SD	Manual Avg	Manual SD
103	76	109	91
correlation coefficient $r = 0.82$			

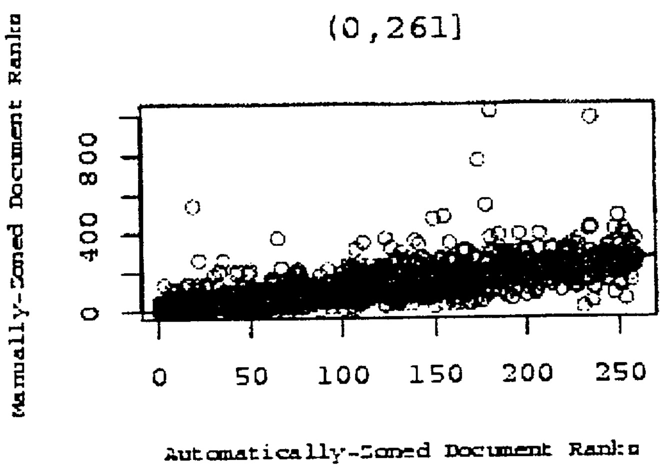


Figure 2: Auto-Zoned Ranks vs. Manual-Zoned Ranks Scatter Plot for First Quadrant Ranks

Document count	1058
Number of pages	46,731
Query Count	62
Average number of relevant documents/query	17
Median number of relevant documents/query	9
Fewest relevant documents for a query	1
Most relevant documents for a query	99

Table 5: LSS Prototype Collection Statistics

with 99.8% character accuracy[3] and another version that has been recognized and processed as describe in Section 2.1. Second, the documents and the queries in this collection were part of the original LSS Prototype Collection and so they should have similar characteristics and topic content as the planned LSN. Collection statistics appear in Table 5.

Again, as in our tests comparing manual vs. automatic zoning, we report on retrieval results using the standard measures of recall and precision. We loaded and indexed both collections into Knowledge Server and the 62 queries were batch run against these two datasets in exactly the same way. The recall/precision results appear in Table 6.

The difference between average precision for the two runs is less than 0.3%. As we pointed out for the manually-zoned vs. automatically-zoned runs, the difference is too small to be considered statistically significant. This test tells us that the process used by DOE to prepare the documents for the LSN will return results equivalent to a collection that was corrected to meet 99.8% character accuracy. With respect to retrievability, an artificially high character accuracy does not guarantee better results for the end user.

Recall Points	99.8% Correct	Auto-zoned w/MANICURE
0.00	0.55	0.54
0.10	0.46	0.45
0.20	0.35	0.34
0.30	0.29	0.30
0.40	0.26	0.26
0.50	0.22	0.22
0.60	0.18	0.18
0.70	0.14	0.14
0.80	0.12	0.11
0.90	0.08	0.07
1.00	0.06	0.05
Avg	0.245	0.242

Table 6: 11-Point Precision for 99.8% Correct Text vs. Automatically-zoned and Recognized Text

4 Conclusion

The tests in this report use well-accepted standards and scientific methods to measure and validate the current procedures used by the DOE to prepare documents for the LSN. We believe our investigation is unbiased and complete.

The aggregation of these results tells us in no uncertain terms that using automatic-zoning followed by MANICURE will give retrieval results equivalent to what one could expect from manually-zoned pages or even from a 99.8% correct collection. We have also shown that there is a strong linear association between ranked results. This association implies that for all practical purposes, the two ranked query result sets are statistically equivalent as well.

We believe these retrieval tests, in conjunction with the results reported in [2], demonstrate that the quality of the documents delivered by the DOE will give effective retrieval results for the users of the LSN.

References

[1] W. B. Croft, S. Harding, K. Taghva, and J. Borsack. An evaluation of information retrieval accuracy with simulated OCR output. In *Proc. 3rd Symposium on Document Analysis and Information Retrieval*, pages 115–126, Las Vegas, NV, April 1994.

[2] Tom Nartker and Ron Young. OCR accuracy produced by the current DOE document conversion system. Technical Report 2002-06, Information Science Research Institute, University of Nevada, Las Vegas, April 2002.

[3] Science Applications International Corporation. Capture station simulation lessons learned. Final report for the Licensing Support System, 1990. Final report for the Licensing Support System prepared under contract DE-AC01-87RW00084 for the U.S. Department of Energy, Office of Civilian Radioactive Waste Management, Washington D.C.

[4] Kazem Taghva, Julie Borsack, and Allen Condit. An expert system for automatically correcting OCR output. In *Proc. IS&T/SPIE 1994 Intl. Symp. on Electronic Imaging Science and Technology*, pages 270–278, San Jose, CA, February 1994.

[5] Kazem Taghva, Julie Borsack, and Allen Condit. Results of applying probabilistic IR to OCR text. In *Proc. 17th Intl. ACM/SIGIR Conf. on Research and Development in Information Retrieval*, pages 202–211, Dublin, Ireland, July 1994.

[6] Kazem Taghva, Julie Borsack, and Allen Condit. Effects of OCR errors on ranking and feedback using the vector space model. *Inf. Proc. and Management*, 32(3):317–327, 1996.

[7] Kazem Taghva, Julie Borsack, and Allen Condit. Evaluation of model-based retrieval effectiveness with OCR text. *ACM Transactions on Information Systems*, 14(1):64–93, January 1996.

[8] Kazem Taghva, Julie Borsack, Allen Condit, and Srinivas Erva. The effects of noisy data on text retrieval. *J. American Soc. for Inf. Sci.*, 45(1):50–58, January 1994.

[9] Kazem Taghva, Allen Condit, Julie Borsack, John Kilburg, Changshi Wu, and Jeff Gilbreth. The MANICURE document processing system. In *Proc. IS&T/SPIE 1998 Intl. Symp. on Electronic Imaging Science and Technology*, San Jose, CA, January 1998.

A Recall and Precision Explained

For these tests to provide a quantitative measure of retrieval effectiveness, we must know in advance which documents are relevant to which queries. The *relevancy judgments*, or list of relevant documents to each query, give us this a priori information. We then apply standard quantitative measures to compare the list of documents retrieved by the system to the relevancy judgments.

The standard measures we use are *recall* (2) and *precision* (3). Recall is the percentage of the relevant documents in the collection that are responses to a query. Precision is the percentage of the responses that are relevant to the query. If you think of it from a users perspective, these are the assessments he would use as well, "Have I received all the relevant documents that are in this collection?" (recall). And, "How many documents do I have to look through to find the ones that are relevant?" (precision). Following are the mathematical formulas that calculate these two values:

$$recall = \frac{\#\_of\_relevant\_retrieved\_documents}{total\ \#\_of\_relevant\_documents} * 100 \tag{2}$$

$$precision = \frac{\#of\_relevant\_retrieved\_documents}{total\ \#\_of\_retrieved\_documents} * 100 \tag{3}$$

Averaging the precision values at specific recall points gives us a better perspective of the overall retrieved results. For example, if we look at Table 7, the system returns on average 35% of the relevant documents when it has returned 20% of the collection. The collective average, in the last row of this table, is just an average of the precision values at each recall point. As you can see, precision tends to decrease as more documents are returned by the system.

There is also a notion of *statistical significance* that we introduce in this report. This is important because slight differences in precision results may not necessarily indicate a fundamental and consistent difference between the result sets.

For us to make a general statement like "automatic zoning will return more relevant documents than manual zoning" the results must be statistically significant. Statistical significance is related to collection size as well as the number of queries used and the number of relevant documents for the queries. For the datasets we've used in these tests, a difference of 5% would be considered significant.

Recall Points	Auto with rmgarbage	Auto w/o rmgarbage
0.00	0.84	0.83
0.10	0.61	0.60
0.20	0.54	0.53
0.30	0.46	0.44
0.40	0.42	0.40
0.50	0.35	0.35
0.60	0.31	0.31
0.70	0.26	0.26
0.80	0.22	0.22
0.90	0.18	0.18
1.00	0.12	0.12
Avg	0.392	0.385

Table 7: 11-Point Precision for Auto-zoned with/without rmgarbage

B Verification of Procedures

Based on ISRI's OCR and information retrieval research, several post-processing routines were built to improve the quality of OCR text loaded into a retrieval system. This set of processes eventually was streamlined into a system we call MANICURE[9]. Together with ISRI, DOE has been tuning this system specifically to LSN documents.

The major components of MANICURE include ppsys[4], a process that automatically corrects misspellings in the text, and rmgarbage, a process which removes "graphic text" and other non-retrieval strings from an automatically-zoned and OCR'd document. Previous experiments [5] have proven the effectiveness of ppsys but the efficacy of rmgarbage had yet to be tested. A simple means of testing its effects was to run the same experiment discussed in Section 2 except that the tested sets would be two versions of auto: auto with rmgarbage and auto without rmgarbage. All other processing steps remained the same. The results of this test appear in Table 7.

Reviewing these results, we see slightly increased precision at the highest recall values and average precision is nearly 40% when rmgarbage is used. And again, although not statistically significant, for this set of documents, a small but definite improvement is apparent.

These experiments show that with automatic zoning and MANICURE, users of the LSN will obtain retrieval results equivalent to what could be expected with manual zoning. Further, since in many cases non-stopwords [2] are corrected when MANICURE is applied, retrievability can potentially be improved.

OCR ACCURACY
PRODUCED BY
THE CURRENT DOE DOCUMENT CONVERSION SYSTEM

ISRI Staff
Information Science Research Institute
University of Nevada, Las Vegas
Las Vegas, Nevada

May 2002

CONTENTS

OCR ACCURACY PRODUCED BY THE CURRENT DOE DOCUMENT CONVERSION SYSTEM

ISRI Staff
Information Science Research Institute
University of Nevada, Las Vegas

May 2002

1. INTRODUCTION

The technical requirements for document collections submitted to the Licensing Support Network (LSN) have been set forth by the Nuclear Regulatory Commission (NRC). The current accuracy requirements for OCR'd document collections are a "target character accuracy of 99.5% with a 98.5% character accuracy target for each individual page." [1]

The Department of Energy (DOE) has selected and installed the best available OCR technology for converting its document collection for submission to the LSN. It is using the "Developers Kit 2000" (SDK2000) distributed by the Scansoft Corporation. SDK2000 is based on combined technologies developed by the Calera, Caere, and Recognita Corporations, and is the best available page reading engine for general purpose use. In installing this system, care has been taken in determining operating parameters that maximize the quality of the output.

Having setup a document conversion system based on SDK2000, the DOE has tasked the Information Science Research Institute (ISRI) at the University of Nevada, Las Vegas (UNLV) to measure the accuracy of the output produced by this system when converting documents from the current DOE collection. This report describes the tests conducted by ISRI staff and gives the average accuracy's measured.

Because the ultimate purpose for the text produced by the OCR engine was to enter it into an Information Retrieval (IR) system [2], DOE has chosen to use the ISRI designed MANICURE post-processing system. MANICURE was designed to accept the document text output from an OCR engine and to perform operations on it that would improve each documents retrievability. The operations applied by MANICURE include "garbage string removal" and spell checking and correction based on document level and collection level dictionaries that are built dynamically. Thus, the DOE document conversion system is a combination of the SDK2000 OCR engine and the MANICURE post-processing system. For this reason, both the accuracy's of OCR output text and of MANICURE output text are reported.

2. MEASURING CHARACTER ACCURACY OF TEXTUAL OUTPUT

The task of measuring the accuracy of textual output is complicated by several factors. First, in order to measure the accuracy of a text stream, it is necessary to have a "correct" text stream for comparison. In most cases, the cost of producing the correct, or ground-truth, character stream is very high [3]. Second, it is necessary to conduct such tests with large numbers of pages. No test of 5, or even 10 pages can be expected to produce statistically significant results. In general, it is preferable to have hundreds of test pages in order to insure significance of the measured results. Third, it is not a trivial issue to determine exactly which accuracy measure is most appropriate for a particular application. The standard measure, which we refer to as "conventional" character accuracy, measures the correctness of every ASCII character on each page. Correctness is defined as the number of total characters minus the number of character errors, divided by the total number of characters.

$$\text{Character Accuracy} = \frac{\text{Total Characters} - \text{Character Errors}}{\text{Total Characters}} \quad \text{(E1)}$$

(1) Final Licensing Support Network Guidelines, April 2002, NRC

(2) The NRC has chosen the Autonomy search system for use in the LSN

(3) Not only must each character of the document be manually retyped, but each character must be checked and rechecked for correctness.

Character errors are the sum of character insertions, deletions, and substitutions that are necessary to convert an output character string into the exact ground-truth string.

In Section 3 of this report, we discuss the relevance of "conventional" character accuracy as a measure of goodness of the output of an OCR system. In Section 4, we describe a test to measure a different set of accuracy metrics. In particular, we measure word accuracy's and the accuracy of the characters in words, produced by the current DOE system. For both of these metrics, we measure the accuracy of both the OCR output text and of the MANICURE output text.

3. CONVENTIONAL CHARACTER ACCURACY AS A MEASURE OF RETRIEVABILITY

The most important part of measuring the accuracy of any document conversion system is to determine what accuracy metric is most appropriate. There are many different performance metrics of conversion systems. The appropriate choice is the metric (or metrics) that best reflect improvement in the usage of the textual output. In this case, the output text will be used to build an index for the Autonomy search engine. Subsequently, the Autonomy engine will be used to retrieve documents of interest. Thus, it is the "retrievability" of documents that is most important.

Although the character accuracy of output text is related to retrievability, the conventional definition of character accuracy is not a good measure of retrievability. For example, OCR technologies typically output one or more characters for any set of black pixels on a page, even though these pixels do not resemble an ASCII character. Manufacturers of these technologies take the position that the user can easily delete such characters if they were generated because of stray marks. If these pixels were ignored by the system, it most certainly would not be noticed by the user. Just in case important information is represented, it is deemed better to draw the users' attention (and require a delete operation) rather than risk loosing important information. This phenomenon is especially noticeable when converting documents that are photocopies.

The overall result can be that a large number of delete operations are required to convert the output character string into the exact ground-truth string. Remember that each such delete operation is counted as a character error (see equation E1). Although the MANICURE system was designed to remove such noise, "conventional" character accuracy of OCR output will be affected by these delete operations.

Furthermore, since the conversion output is to be used to build the index in an IR system, it is the accuracy of the <u>words</u> to be indexed that better reflects retrievabiity. "Word" accuracy is defined as follows:

$$\text{Word Accuracy} = \frac{\text{Total Words} - \text{Number of Incorrect Words}}{\text{Total Words}} \qquad \text{(E2)}$$

In fact, since IR systems normally are setup to ignore some specific words, called stopwords (such as "the" & "and"), "non-stopword" accuracy is yet a better measure of retrievability. Equation E2 can also be used to calculate non-stopword accuracy by substituting non-stopwords for words.

The major point here is that print noise (or any stray marks), numbers, and punctuation marks in a document are NOT indexed by IR systems and thus, do not affect retrievability. Since "conventional" character accuracy can be profoundly affected by these kinds of characters, it is clear that "non-stopword" accuracy is a much better measure of retrievability. One possible alternative is to measure just the accuracy of the characters used to make up non-stopwords as an alternative to the "conventional" definition of character accuracy. We refer to this measure as the "character accuracy of non-stopwords" and use equation E1 replacing "characters" with "characters in non-stopwords."

We thus undertook the task of conducting a test to measure the average "non-stopword" accuracy (and the character accuracy of these non-stopwords) that is produced by the current OCR/MANICURE system. To ensure that the full benefit of using MANICURE is measured, complete documents must be used in these tests. Our goal was to measure "non-stopword" accuracy from a set of "documents" selected at random from the DOE collection.

4. NON-STOPWORD ACCURACY OF DOE DOCUMENTS

The major impediment to document level tests of OCR accuracy is the cost of producing the "correct," or ground-truth, copy of each page to use in calculating accuracy's. The cost of producing accurate ground-truth for even two or three 80 page documents is extremely high. Thus, finding a low cost method of producing the ground-truth needed was a dominant part of conducting document level tests.

To solve this problem, we selected 17 documents at random from the DOE collection that had Microsoft Word based native files. The accession number, the total number of non-stopwords, and the number of characters in each of these non-stopwords are shown in Table 1. We developed a process to capture the correct output text directly from the Microsoft Word system. We also parsed this text to remove all punctuation, most of the digits, and all stopwords (1). A concerted effort was made to retain document identifiers and other "project words" containing digits.(2) Thus, the text remaining contained only English non-stopwords (and project related non-stopwords that might not be in a normal dictionary) and formed the basis for computing accuracy's (3).

Table 1. Number of Non-stopwords and Characters in the 17 Document Sample

Document Accession Number	Total Number of Non-Stopwords	Number of Characters in Non-Stopwords	Number of Unique Non-Stopwords	Characters in Unique Non-Stopwords
mol199907200407	6974	56611	912	7880
mol199911010207	9641	80684	1720	15199
mol200002170216	9595	79777	1796	16064
mol200002280529	5572	47005	1088	9674
mol200004130692	4115	33855	778	6641
mol200004140874	12379	101435	2131	18691
mol200005230155	4381	37387	769	7033
mol200005250378	13920	113193	2465	21669
mol200005260336	6318	51210	1193	10743
mol200006090266	5112	44172	1175	10750
mol200006270254	7792	62586	1307	11738
mol200007250453	36713	302763	4198	39005
mol200011220005	8440	70020	1611	14221
mol200012080086	4523	37402	959	8568
mol200101250233	14247	120653	1883	16925
mol200103160002	8441	69501	1338	12360
mol200104160088	6320	52870	907	7874
Average of all 17 documents	**9675.47**	**80066.12**	**1542.94**	**13825.59**

Because the images extracted from native Microsoft Word documents were never printed or scanned, they were completely free of defects associated with either the printing or scanning process. Although the cost of generating this ground-truth data was reasonable, tests of OCR output accuracy from these images would not produce results that were typical of the current DOE conversion operation. Even if each document were printed and scanned, the images produced would be of higher quality than the average image from the DOE collection.

(1) The stopwords removed were the Brown Corpus list of 450 stopwords.

(2) The criterion for "project-words" that were retained was the same as that used by the MANICURE system. Equations, tables, graphs, and other non-textual material were manually removed.

(3) Note that the total number of characters of test data is over 1.3 million characters.

We therefore chose not only to print and scan each page of these documents, but to produce several generations of photocopies of each document. Our best judgement of the average quality of images in the DOE collection is somewhere between a first and second-generation photocopy. Thus, we chose to measure non-stopword accuracy's not only from original images, but also from the first printed and scanned image and from the first, second, third, and fourth generation photocopies of these images. Care was taken to use the same photocopy engine to produce all successive copies. To insure that all accuracy's measured were typical of current DOE conversion operations, all image copies were processed by BSC operations staff and the resulting OCR & MANICURE output was transmitted to UNLV on CD-rom

In designing this test of OCR output accuracy based on the characters of non-stopwords in documents, one other important issue was considered. If an OCR engine mis-recognized any of the characters of a non-stopword in a document, that document might still be retrieved by a retrieval engine. Since most non-stopwords exist several times in a document, if any one of these were recognized correctly, the document could still be retrieved by a word search. Thus, since retrievability is the important issue, it is also desirable to measure non-stopword accuracy based on "unique" non-stopwords. The idea is that character errors made in recognizing non-stopword A are not significant as long as one correct occurrence of A is generated. The number of unique non-stopwords in each document is also shown in Table 1.

Therefore, we constructed a program to measure both the average non-stopword accuracy (and the average accuracy of the characters of non-stopwords) for all non-stopwords and for "unique" non-stopwords from input documents. Although only 17 documents were involved, six different images of each document were tested. The first image was the "original" image extracted from Microsoft Word. The second image was the first-printed and scanned image and we refer to this as "generation 0". The third through sixth images are the first through fourth generation photocopies of generation 0 of these documents. The average OCR output accuracy's measured for all of these images for each document is shown in Appendix A. The average MANICURE output accuracy's measured for all of these images for each document is shown in Appendix B.

4.1 Summary of the Results of Non-stopword Accuracy Tests

The average character accuracy's for all 17 documents are shown in Table 2 below. The top two rows show accuracies from raw OCR output. The bottom two rows show accuracies from MANICURE output. The first and third row show accuracy's for all non-stopwords and the second and fourth rows show accuracies for unique non-stopwords.

TABLE 2. AVERAGE CHARACTER ACCURACY FOR ALL 17 DOCUMENTS

System	Character Accuracy Of	Orig.	Gen 0	Gen 1	Gen 2	Gen 3	Gen 4
RAW OCR OUTPUT	All non-stopwords	99.37	99.40	99.20	98.67	97.89	97.79
	Unique non-stopwords	99.65	99.58	99.46	99.22	98.93	98.80
MANICURE OUTPUT	All non-stopwords	99.50	99.47	99.30	98.83	98.16	98.06
	Unique non-stopwords	99.66	99.57	99.44	99.27	98.93	98.86

Table 2 shows that the character accuracy of all non-stopwords from the MANICURE output is slightly better than the raw OCR output. This improvement is more profound for higher generation photocopies (i.e., 97.79% for raw OCR and 98.06% for MANICURE for the fourth generation copy). It is interesting that this improvement is not reflected in the unique non-stopword results. Again the "improvement" of MANICURE output accuracy over raw OCR output accuracy is greatest for the fourth generation copy, but even then is only 0.06% (i.e., 98.80 for raw OCR and 98.86 for MANICURE). In general, this result shows that the MANICURE post-processing system does improve the accuracy of "all" non-stopwords in a document but does not significantly improve the accuracy of "unique" non-stopwords.

If the average image quality from the DOE collection is between a first and second-generation photocopy, then the most appropriate character accuracy (i.e., unique non-stopwords) is between 99.44% and 99.27%.

TABLE 3. AVERAGE NON-STOPWORD ACCURACY FOR ALL 17 DOCUMENTS							
System	Word Accuracy Of	Orig.	Gen 0	Gen 1	Gen 2	Gen 3	Gen 4
RAW OCR OUTPUT	All non-stopwords	97.44	97.03	96.45	95.05	92.78	91.91
	Unique non-stopwords	97.89	97.56	97.28	96.77	95.64	95.46
MANICURE OUTPUT	All non-stopwords	98.01	97.54	97.23	96.15	94.64	94.14
	Unique non-stopwords	98.74	98.36	98.15	97.65	96.61	96.51

The results shown in Table 3 above are much more significant. The word accuracy improvement of MANICURE output over raw OCR output for all non-stopword output ranges from 0.51% for generation 0 to 2.55% for generation 4. A one percent improvement in this accuracy measure is very significant (1). The word accuracy improvement of MANICURE output over raw OCR output for unique non-stopword output ranges from 0.80% for generation 0 to 1.05% for generation 4.

These results show the benefit of applying the MANICURE post-processing system. Even an 0.8% improvement in word accuracy is significant. In addition, these results show that the improvement provided by MANICURE post-processing increases as page quality decreases. This is exactly as we expected, since MANICURE was designed to improve the accuracy of the non-stopwords in OCR output leading to improved overall document retrievability.

5. SUMMARY

We believe the unique non-stopword accuracy's between 98.15 for generation 1 and 97.65 for generation 2, as shown in Table 2, to be very high. It is important to note that word (and non-stopword) accuracy's are always lower than character accuracy's. (2)

In terms of character accuracy, the character accuracy of unique non-stopwords for first and second-generation copies, as shown in Table 1, is between 99.44% and 99.27%. Although these accuracies are not quite at the 99.5% level, they are carefully measured results based on realistic documents and over 235,000 characters.

Overall, using the accuracy metrics ISRI believes are most appropriate, these results indicate that the character accuracy's produced by the current DOE document conversion system are very close to NRC requirements.

Finally, because retrievability of documents from the LSN will be the primary use of the text produced by the DOE, it seems clear that a test measuring retrieval effectiveness is at least as important; in fact, it is more important than the tests described above. ISRI has recommended that tests measuring retrievability (i.e., precision and recall) of documents from the Autonomy retrieval system be conducted.

(1) A one percent improvement over 97% correct words corresponds to eliminating 1/3 of the word errors.

(2) Because one incorrect character in a word causes the whole word to be in error, and because character errors tend to be spread among different words, character accuracies are uniformly higher than word accuracy's. This phenomena has been repeated in every OCR test conducted by ISRI over the past 10 years.

APPENDIX A.

Non-stopword accuracy's and unique non-stopword accuracy's for OCR output

addresses the application of analytical methods and examines the stress controlled modes of failure. The analyses performed in this report, which simulates excavation of the opening in a gravity-stressed rock medium, will serve as an analytical basis for assessing the opening shape and determining the general stress distribution around the tunnel opening. These analyses will be complemented with a rock mass classification study to aid in the design of a ground support system for the TS North Ramp opening. The analyses incorporating far field thermal loads and seismic effects will be addressed in this report. The ground support system recommended using empirical methods (Reference 8.3) will be analyzed in this report. The quasi-static seismic analysis are performed on an unsupported ramp to determine the stress distribution and potential failure zones around the opening. Dynamic analysis are performed on an unsupported and a supported ramp to determine the effects on the ground support system. The results of computer analyses and the rock mass classification study will be incorporated to establish the final ground support system. The final design will be complemented by a monitoring program during construction. Adjustments to the ground support system will be made during construction as required due to field conditions.

Analysis are performed at six different stations along TS North Ramp. Station 1+50 and 7+00 represent Tiva Canyon member (TCw) and Station 2+65 represents the material behind the Bow Ridge Fault (Rainier Mesa). Stations 10+00, 18+00, and 27+00 represent the upper Paint Brush Tuff (PTn), uppermost Topopah Spring (TSw1), and middle Topopah Spring (TSw2) members respectively.

Rock and joint properties are required to perform the analysis. In this report the rock properties from the NRG drilling program are used in the analysis. The joint properties from Reference 8.8 are used in performing the jointed rock analysis.

FLAC version 3.22 and UDEC Version 2.0 computer software are used to perform the analyses in this report. FLAC is based on a Lagrangian calculation scheme for continuum modeling while UDEC is a numerical program based on the distinct element method for discontinuum modeling. These two computer programs and their use in the analysis are explained in later sections in this report.

10.2 DESIGN INPUT DATA

The design of effective ground support systems for underground openings requires site-specific design input data. The mechanical rock properties and characteristic joint parameters are required to perform reliable computer modeling. In this report, data from NRG drill core testing has been used to perform the analyses. The mechanical properties from the NRG drilling program used in this analysis are presented in Attachment I. The mean values of the measured parameters from testing the NRG core are used in the analysis. Joint properties are not available from the NRG drilling program, therefore, the best available data from Reference 8.10 is used. Rock mass the shift will have to be designed into the trailing gantry of each TBM. These items include the following:

A. Precast concrete segments and rail
B. Conveyor components
C. Roof bolts, plates and resin
D. Steel sets

An additional provision would be required to haul ventilation pipe. Explosives haulage would require a special flat car equipped to carry explosives.

A more detailed study of the actual planned ESF construction plan is required to specify actual car storage needs and the actual number of locomotives required..

4.1.3.2 Nonroutine Materials

There are no known nonroutine loads which are too large to fit the transportation corridor or the capabilities of a 22.7 tonne locomotive.

4.3.2 Trackless Haulage

4.3.2.1 Routine Materials

Flat bed vehicles with 10 tonne capacity would be used to haul routine supplies such as conveyor components, roof bolts, roof bolt plates, resin, steel liner plates and wire mesh. Special tractors and low boy trailers would be used to haul the precast concrete floor segments. Special vehicles would be required to haul explosives if used and ventilation duct. Cycle times for trackless equipment are shown in Appendix G. A typical vehicle can make two or four trips per shift depending on the location of the TBM.

4.3.2.2 Nonroutine Supplies

Nonroutine loads would be hauled with tractors and low boy trailers.

lists the experts on the panel and their affiliations. Brief biographies for members of the expert panel are provided in Appendix A.

Technical Specialists: Numerous technical specialists participated in the project by providing the experts with specialized data, interpretations, or training through workshops and a field trip. A list of the technical specialists and their affiliations is given in Table 1-3. Members of both the MDT and the expert panel also acted as technical specialists.

1.4 PRODUCTS OF STUDY AND STRUCTURE OF REPORT

The UZFMEE study was conducted in approximately eight months. The project began with developing a plan for the course of the study and identifying the goals to be accomplished and methodologies to be implemented in meeting these goals. Next, the MDT developed and implemented a process for selecting the members of the expert panel, resulting in the selection of seven experts. The bulk of the study was centered around three workshops and one field trip. These activities were designed to facilitate interaction among the experts, provide all data needed for their assessments, and provide a forum for discussing a range of technical interpretations. Following the third workshop, the interpretations of each expert were elicited in individual interviews and documented in elicitation summaries. After reviewing the elicitation summaries of each member of the expert panel and the sensitivity analyses provided by LBNL, the experts finalized their assessments. The MDT performed the final calculations to show the individual and aggregated distributions on percolation flux at the proposed repository horizon at Yucca Mountain.

This report contains the products of the activities of the UZFMEE project outlined above. Section 2 describes in detail the process followed in eliciting the expert interpretations. Appendices B and C provide summaries of the references provided to the experts, and of the three workshops and the field trip. This information provides written documentation of the technical data discussed by the panel, the formats and content of interpretations presented by outside technical specialists during the study, and the expert panel's preliminary interpretations.

Section 3 of this report presents in detail the final interpretations provided by the expert panel and the results of the study. Both the results for each of the seven individual experts and the cross-hole seismic profiling, coupled hydraulic-mechanical characterization, radial-convergent tracer tests, two-well recirculating tracer tests, injection-pumpback tracer tests, and development of transport models.

Two alternative designs for the C-Hole multiple-packer hydraulic and tracer testing system were completed by the U.S. Bureau of Reclamation (USBR) in cooperation with the USGS. Input for design was obtained from hydrologists and engineers at USGS, USBR, LBL, Los Alamos, University of Nevada/Las Vegas (UNLV), and INTERA Corporation at the Waste Isolation Pilot Plant in Carlsbad, New Mexico. The final design is the less complicated, less expensive of the two alternatives. Although the final design will be easier to fabricate and assemble, it will require more drill-rig support for field use because packer strings will have to be disassembled and moved more often from hole to hole. Current plans call for fabrication and assembly of the prototype packer string during the second half of FY 90 and field testing during the first half of FY 91.

Final technical plans were formulated for conducting cross-hole seismic surveys in the C-Holes. Because the seismic surveys are designed to determine fracture characteristics in the saturated zone, results of these surveys are needed to finalize hydraulic and tracer testing strategies for the C-Hole complex.

Activity 8.3.1.2.3.1.5 - Testing of the C-Hole Sites with Conservative Tracers. Work began on the evaluation of various analytical techniques and computer codes for interpretation of the planned C-Hole tracer tests. A reaction flow-path model (CHILLER) developed by the Colorado School of Mines was evaluated for possible use. In addition, it was concluded that no additional analytical tools for interpretation of conservative tracer tests are needed at present because of the capabilities of the LBL TRINET and Los Alamos FEHMS flow and transport codes.

Initial literature research conducted by hydrochemists at UNLV suggests that pyridone derivatives probably are viable tracers for use in planned C-Hole tests. Based on specific technical aspects of both conservative and reactive tracer tests, scientists at UNLV and Los Alamos provided input to the design of the multiple-packer system being developed for conduct of the planned C-Hole hydraulic and tracer tests. Informal discussions also took place with scientists from SKB Sweden (Swedish repository program) to become familiar with SKB experience and plans in regard to use of tracer tests to characterize saturated, fractured rocks.

Activity 8.3.1.2.3.1.7 - Testing of the C-Hole Sites with Reactive Tracers. The development, testing, and documentation of the computer code FEHMN (Finite Element Heat and Mass) has been carried out. The code will be used to design the reactive tracer tests and to simulate the results.

Several detailed technical procedures referenced in the study plan for this activity were either written or revised.

Detailed specifications for the drilling of an additional borehole at the C-Wells complex have been laid out. The drilling of a new hole will supply core collected under an approved QA program for carrying out the laboratory

1. INTRODUCTION

This predictive geotechnical report was developed specifically for the Enhanced Characterization of the Repository Block (ECRB) project, which is intended to help determine the suitability of the Yucca Mountain site for the construction of a potential underground high-level nuclear waste repository. The ECRB project includes the design and construction of the Cross Drift, a new drift in the Exploratory Studies Facility (ESF). The ECRB project owner is the Department of Energy (DOE). The ECRB design team is the Civilian Radioactive Waste Management System (CRWMS) Management and Operating Contractor (M&O), which is a cooperative effort of companies, national laboratories, and government agencies.

The *ECRB Predictive Geotechnical Report* includes two volumes, (1) the Geotechnical Baseline Report (GBR) and (2) the Geotechnical Data Report (GDR), and is based on the relevant data and information collected at the Yucca Mountain Site Characterization Project (YMP).

1.1 PURPOSE OF THE GBR

This GBR was developed to establish an understanding of the subsurface conditions anticipated during construction of the ECRB Cross Drift and as such, sets baselines for the geotechnical conditions anticipated for the ECRB project. The GBR presents a description of the range of ground conditions expected during construction of the Cross Drift and provides the rationale for setting the geotechnical baselines.

1.2 SCOPE OF THE GBR

The scope of the GBR is limited to interpretive discussions and baseline statements that describe the geotechnical conditions expected during construction of the ECRB Cross Drift. The GBR is the primary document of the *ECRB Predictive Geotechnical Report*. Its companion document, the GDR (CRWMS M&O 1998b), is a summary of the factual information gathered during YMP investigations relevant to the proposed construction. The GBR is based on the data and information presented in the GDR. The scope and technical content of the GBR were established in the *Technical Document Preparation Plan for the ECRB Predictive Geotechnical Report* (CRWMS M&O 1998a), and were based on the American Society of Civil Engineers (ASCE) recommended practice and guidelines (Essex 1997).

1.3 BACKGROUND

The Nuclear Waste Technical Review Board has consistently expressed the position that high-priority site characterization activities include a full east-west traverse of the proposed repository block (CRWMS M&O 1997j). The repository block is located primarily within the Topopah Spring crystal poor lower lithophysal zone (Tptpll) of the TSw2 thermal mechanical unit (CRWMS M&O 1997a, p. 76). The Main Drift of the ESF is a north-south excavation located along the eastern edge of the potential repository block. The Main Drift is constructed almost entirely within the Topopah Spring crystal poor middle nonlithophysal zone (Tptpmn) of the TSw2 unit (Albin et al. 1997). Only a very small portion of the Main Drift is in the Tptpll. The ECRB Cross Drift will provide enhanced characterization of the repository block by exposing a greater portion of the Tptpll (see Section 4). The ECRB Cross Drift will also provide an

A. LIFE CYCLE PLAN

1. *Introduction*

a. Overall nature and purpose of the software
Computer program, called GUMBEL, was written to perform statistical analysis of meteorological data so that the statistical model can be used to make extreme value projections, which can be used in repository design.

b. Summary of functional requirements for the intended use of the software
The program uses a non-linear regression technique that produces a least squares best fit of meteorological data to statistical distributions known as the Gumbel and Weibull Distributions. This best fit is then used to make projections of the data for return periods of 50, 100, and 200 years.

2. *Software Requirements*

a. Functional requirements for the intended use of the software
Data input is interactive except for the sequential data to be analyzed. The input data set consists of ascending values of meteorological data (such as rainfall or wind speeds). The data are fit to a statistical model by minimizing the square of the difference between the observed points and calculated values (residual sum of squares) over a range selected by the user. Details can be found in reference 1.

b. Planned mathematical models and numerical methods
Initially two mathematical models were coded into the program, but others may be added at a later time. The statistical models are the Gumbel and Weibull Distributions (details on these distributions can be found in reference 1). The mathematical models are linearized. The program makes estimates of the statistical parameters. Corrections to the original estimates are applied iteratively until the successive estimates of the residual sum of squares are within 10^{-5}, or until 10 iterations have been made. (For more information on non-linear regression methods, the reader is advised to consult with a standard mathematical or statistical textbook.)

c. Performance requirements with respect to range of applicability and accuracy
For plotting accuracy, five significant figures is sufficiently accurate. A good quality printer can print about 1000 dots per inch, thus five-figure accuracy is considered sufficient.

d. Planned software language and version
The program is written in Microsoft's Visual Basic version 3.0 (Microsoft, 1993) programming language.

f. Planned computer operating system and hardware
The program runs on a personal computer with a Windows 95 operating system.

6.3.6.7 Portable Radiation Shields

This alternative may need to include occasional human access by using "portable" radiation shields that would be placed over waste package(s) by a remotely controlled gantry. EDA V also would require a gamma shield. Although this shield is unlikely to be portable and there are no details on the operations envisioned to install and remove the shield, it is noted that they could introduce new DBEs relative to the VA reference design. Potential DBEs associated with using the shield include dropping of the shield onto one or more WPs, and impact to the WP by malfunctions of the gantry system. If the design proceeds with the concept of portable shields, the WP design bases will have to be adjusted to assure no breach of the WP can occur.

6.3.7 Off-normal Event Recovery

Recovery equipment for off-normal conditions could be used to clean up a rockfall, while emplacement equipment could be used to recover the waste package. The equipment for emplacement and recovery of the waste packages is consistent with that proposed for the VA reference design. It is not anticipated that an off-normal event will require any additional considerations for EDA V.

6.4 FLEXIBILITY

This criterion expresses the degree to which a design would be capable of remaining viable and/or able to change in the face of future regulatory or other changes. Possible changes to consider are included in the following sections.

6.4.1 Increased Disposal Capacity

EDA V design is extremely flexible in regards to an increased capacity for the repository. The high AML enables the waste to be placed in a considerably higher waste package density, resulting in a substantial decrease in the area required for waste emplacement. Two scenarios for increase repository capacity are discussed in the following sections.

6.4.1.1 Disposal Scenario One

This acreage calculation is presented to indicate the flexibility of EDA V with respect to an increased disposal capacity. The acres required for disposal scenario one is based on the AML of 150 MTU/acres (Section 4.1.16) and the CSNF of 87,000 MTU (Section 4.1.6). The drift spacing can be adjusted to accommodate all waste including CSNF and HLW canisters.

$$\text{Acreage} = \frac{\text{CSNF (MTU)}}{\text{AML}}$$

Where: CSNF = MTU of commercial spent nuclear fuel and
AML = areal mass loading in MTU/acre.

chloride concentrations in water samples collected during hydraulic tests in the perched water bodies are not very variable because the water is probably well mixed. Their modeling exercise also showed that perched water compositions are best matched by a mixture of Pleistocene age water with variable amounts of modern water (Sonnenthal and Bodvarsson 1999, p. 151).

6.2.6 Summary and Further Discussions

Based on the discussions of transport issues and processes for the UZ (Sections 6.2.1 through 6.2.5), we have constructed a conceptual model of transport in the UZ that can be summarized as follows. Advective transport pathways coincide with flow pathways. Matrix diffusion is a major mechanism for mass transfer between fractures and matrix, and is expected to contribute to the retardation of the radionuclide transport when fracture flow is dominant. Sorption may retard the movement of radionuclides in the UZ, but this retardation is limited by the reduction of fracture-matrix interfacial area resulting from fingering flow in fractures. However, sorptive interactions may enhance radionuclide transport if the aqueous species sorbs to colloids that subsequently may be transported through the UZ. Dispersion is not expected to be a major transport mechanism in the UZ.

It is useful to emphasize that flow is a major driving force for transport. As a result, the conceptual model of transport in the UZ is closely tied to the conceptual model of flow. Alternative conceptual models of flow give rise to different transport behavior from that discussed above. If the liquid-water flow primarily occurs in the matrix—as hypothesized by Wang and Narasimhan (1993, pp. 327-339), Peters and Klavetter (1988, pp. 416–430) and Nitao and Buscheck (1991, pp. 2099-2112)— matrix diffusion will be insignificant for UZ transport, and colloid-facilitated transport may not need to be considered given that it mainly occurs in fractures. In contrast, if most liquid-water flows through structural features (Montazar and Wilson 1984, p. 51) or sparse flow paths (Pruess 1999, pp. 1040–1051), transport will be primarily determined by flow in fractures and the effects of the matrix, such as matrix diffusion and matrix sorption, become insignificant. However, as discussed in Section 6.1.9 of this report, liquid-water flow in the UZ is considered more likely to be consistent with the current conceptual model of flow rather than those proposed by these authors.

6.3 COUPLED PROCESSES: EFFECTS ON FLOW AND TRANSPORT

If the Yucca Mountain site is determined to be suitable and is licensed, the DOE is planning to emplace, in a geologic repository at the site, radioactive wastes that will emit a significant amount of radioactive decay heat. This heat will influence hydrologic, mechanical, and chemical conditions in both the near field (drift-scale) and far field (mountain-scale). This subsection discusses the effects of the corresponding coupled processes, including thermo-hydrologic (TH), thermo-mechanical (TM), and thermal-chemical (TC) processes, on flow and transport within the UZ at Yucca Mountain. Note that TH, TM and TC processes are still coupled among themselves, although they are discussed separately (for reasons of simplicity) in this subsection.

6.3.1 TH Processes

The expected TH response of the unsaturated, fractured tuff to potential radioactive decay heat involves a number of key processes (Buscheck and Nitao 1993, pp. 418-448; Tsang and Birkholzer 1999, pp. 389–390). As the formation temperatures rise around waste packages, pore

Data Distribution and Unit Geometry–The distribution of modeled alluvium is illustrated in Figure 15. Alluvial thickness was interpreted with the use of the site area geologic map (DTN: GS970808314221.002) and available borehole data (DTN: MO9811MWDGFM03.000), including the UZN boreholes as discussed in Section 6.1.1. The areal extent of alluvium is well constrained by geologic mapping; however, because some boreholes did not penetrate to bedrock, the alluvial thickness is constrained by limited subsurface information. The map, therefore, should be considered more representative of a minimum alluvial thickness or an interpretation based on sparse data rather than of an absolute thickness.

As shown in map view (Figure 13), the post-Tiva rock units are only sparsely encountered in the modeled area. The distribution is based on the geologic map (DTN: GS970808314221.002) and borehole data (DTN: MO9811MWDGFM03.000). South of Yucca Wash, these units are typically preserved in wedges on the downthrown sides of faults. For example, in Figure 14, a wedge of the Tiva Canyon Tuff Crystal-Rich Member and post-Tiva unit is shown on the downthrown side of the Solitario Canyon fault.

6.4.1.2 Tiva Canyon Tuff (Tpc)

Overview–In the GFM, the Tiva Canyon Tuff (Table 5) consists of the Crystal-Rich Member (Tpcr, grouped with post-Tiva rocks) and the Crystal-Poor Member (Tpcp), which is undivided in the GFM except for the three basal vitric subzones (Tpcpv1, Tpcpv2, and Tpcpv3) and a low-density zone (TpcLD). The Tiva Canyon Tuff makes up most of the exposed bedrock in the modeled area (Figure 13).

Because the Tiva Canyon Tuff makes up most of the exposed bedrock on Yucca Mountain, it is important in hydrologic infiltration modeling. The distribution of the lower vitrophyre (Tpcpv3) may be important in hydrologic modeling because, like the other vitrophyres, the lower vitrophyre is one of the layers in the mountain having the lowest porosity (Rautman and McKenna 1997, p. 142).

Data Distribution and Unit Geometry–The distribution and thickness of Tpcpv3 are illustrated in Figure 16. The model interpretation for this unit is based on borehole data (DTN: MO9811MWDGFM03.000) and abundant geologic map data (DTN: GS970808314221.002). Because the top of the Tiva Canyon Tuff is extensively eroded in the model area, none of the input boreholes penetrate the entire formation, and a true thickness map cannot be produced. The Tiva Canyon Tuff is thickest in the center of the modeled area and thins to the east, west, and south. The crystal-poor densely welded vitric subzone (Tpcpv3) is present only in the southwestern part of the area and appears to be distributed as pods or in a web-like pattern (Figure 16).

6.4.1.3 Paintbrush Tuff Nonwelded (PTn) Unit

Overview–The PTn unit (defined in Table 5) is a grouping of rock layers used in hydrologic and thermal-mechanical modeling. Stratigraphically, it consists of the rock units Tpcpv2, Tpcpv1, Tpbt4, Tpy, Tpbt3, Tpp, Tpbt2, Tptrv3, and Tptrv2.

The EDA II base case model was modified for this calculation by combining the modifications described in Sections 5.1.1.6 (UZ neutralization) and 5.1.1.7 (SZ neutralization). Also, all radionuclide-specific information associated with the additional 30 radionuclides was added to the model. The radionuclide inventory and solubility information was obtained from Chapter 6 of the TSPA-VA Technical Basis Document (CRWMS M&O 1998g, Table 6-6, and Table 6-32, respectively). Radionuclide decay rate information was obtained from the Chart of the Nuclides (General Electric 1984).

5.1.1.11 Neutralization of the Unsaturated Zone and Saturated Zone Transport Barriers, and the Overlying Rock Barrier

This scenario investigates the importance of the combined effect of the UZ and SZ transport barriers and the overlying rock on the performance of the repository. The EDA II base case model was modified for this calculation by combining the modifications described in Sections 5.1.1.6 (UZ neutralization), 5.1.1.7 (SZ neutralization), and 5.1.1.1 (overlying rock neutralization). In addition, the 30 additional radionuclides added to the above case were also added to this case.

5.1.1.12 Neutralization of Thermal Effects

This scenario investigates the importance of thermal effects on the performance of the repository. The only modifications to the EDA II base case that were necessary for this simulation were to modify the external RIP files that specify waste package temperature histories and waste package degradation rates.

The external RIP tables with the file name extensions t02 and t05 contain waste package temperature histories for commercial spent nuclear fuel and high-level waste, respectively. These two dimensional tables contain 12 columns of waste package temperature values versus time. The first six columns represent the temperature histories for the current climate in all repository regions. The next six columns represent the temperature histories with climate change for the same six repository regions. To modify these two tables to represent ambient temperatures at all times, all the time and temperature entries were removed from the table. Then, two rows were added, one for time zero and one for 1,000,000 years. All 12 columns in both rows were given ambient temperature (25 °C).

The waste package degradation rates were changed by obtaining new waste package degradation time histories (RIP input tables t20 and t35) and new drip shield failure time histories (RIP input table t38) for ambient temperature and humidity conditions. Table t20 is the degradation history for waste packages that always experience dripping conditions and table t35 is the degradation history for waste packages that do not experience dripping conditions. These new degradation time histories were generated by the WAPDEG code (TBV 568) (CRWMS M&O 1999d). The WAPDEG result file NE1a5s5EDA2-2-wp_did.rip was used for RIP external table t20, NE0a5s6EDA2-2-wp_did.rip was used for RIP external table t35, and NE1a5s5EDA2-2-ds_did.rip was used for RIP external table t38 (DTN: MO9906MWDWAP90.000 under directory ./elect/Constant_History /Post308outputs_ConstHis).

The active fracture concept accounts for the contact area between the fracture and the matrix (Table 4), as well as the frequency of fractures (Table 4). The AFC is that fracture flow only occurs through some of the fractures. This is more conservative than assuming the influx flows evenly through all fractures. The flux through a fracture is greater when it has higher saturation and, therefore, focusing flow through a portion of the fractures (i.e., to active fractures) maximizes flux and results in fast pathways for flux through the mountain.

The rock properties in DTN: LB990861233129.001 were calibrated using an inverse modeling technique that assumes the properties will only be used in DKM employing AFC. Therefore, the DKM and AFC are appropriate NUFT options.

3.1.2 YMESH

YMESH is classified as an unqualified software program (per AP-SI.1Q, *Software Management*), and is under configuration management (Table 1). YMESH is used in this model to interpolate the thickness of the stratigraphic units as documented in Attachment VI (file: LBL99-YMESH) at given locations (Section 5.1.5). YMESH is appropriate software for this task. YMESH was run on a Sun Ultra 2 workstation with SunOS 5.5.1 operating system.

3.1.3 CONVERTCOORDS

CONVERTCOORDS is classified as an unqualified software program (per AP-SI.1Q, *Software Management*), and is under configuration management (Table 1). CONVERTCOORDS is used to convert from Universal Transverse Mercator coordinates to Nevada State Plane coordinates, as well as to reformat the data (see Attachment VI, files: *.inf). The desired format is columns of data, with the input files in a matrix format. CONVERTCOORDS is appropriate software for this task. CONVERTCOORDS was run on a Sun Ultra 2 workstation with SunOS 5.5.1 operating system.

3.1.4 XTOOL

XTOOL is classified as an unqualified software program (per AP-SI.1Q, *Software Management*), and is under configuration management (Table 1). The output from XTOOL is graphical (no actual data is produced with XTOOL). XTOOL is tracked in accordance with AP-SI.1Q because it is not commercial off the shelf software, and it is under configuration management (Table 1). XTOOL is used to develop graphical representations (Figures 2 through 4) of the results in the NUFT output files (VI-files: *.out). XTOOL is appropriate software for this task. Software programs used to produce figures are exempt from AP-SI.1Q requirements. XTOOL was run on a Sun Ultra 10 workstation with SunOS 5.6 operating system.

3.2 DESCRIPTION OF ROUTINES USED

All routines used in the preparation of this document are qualified within this document as follows: Chim_Surf_TP V1.1 (Chim_Surf_TP) and Chim_wt_TP V1.1 (Chim_wt_TP) are qualified in Attachment II, ColumnInfiltration V1.1 (ColumnInfiltration) is qualified in Attachment III, Cover V1.1 is qualified in Attachment IV, and rme6 V1.1 (rme6) is qualified in Attachment V.

of waste. He stated that the Committee could help by determining the greatest contributors to risk. John Larkins (NRC) noted that the General Accounting Office has been critical of the NRC staff for going forward in the risk area without a plan.

A letter to the Committee from Bob Budnitz (consultant) was discussed. Garrick noted that the letter states that Budnitz does not understand the Committee's views on defense in depth. Garrick stated that he thought Budnitz misinterpreted the Committee's view. The Committee agrees that the contribution of each level of protection should be understood, but that the contribution should not be prescribed. Dr. Hornberger noted his opinion that defense in depth has not been clearly articulated except for the multiple-barrier concept. He stated that Budnitz is asking how to judge the sufficiency of a given barrier. And he stated that he did not think apportionment to natural and engineered barriers is appropriate. Ted Sorensen (NRC staff) stated that the fundamental goal should be to replace defense in depth as a fundamental objective with risk.

Regarding the importance analysis, Andy Campbell (ACNW staff) discussed the rapid-dissolution model the DOE uses, and he questioned how the rapid dissolution calculated may be driving other considerations.

Agendas for future meetings were then discussed. As part of the September 14 through 16 meeting (in Rockville), there will be full-day session (including training) on risk communication on September 16. The staff will discuss Part 63 and the DEIS. The October meeting will be in Nevada. It will include a half-day tour of Yucca Mountain, a discussion of ongoing science activities, and meetings with the public. The Committee will plan to brief the Commission in November. The December meeting will be the 13th through 15th and will likely include a retreat.

During a discussion of the previous meeting, Hornberger indicated that he discussed the volcanism issue with the Center for Nuclear Waste Regulatory Analysis (CNWRA). He noted that the Committee did not hear anything new during the meeting on this subject. He agreed that some orderly additional work on the issue would be appropriate, but that it should not be raised to DOE as a "hot-button" issue.

After lunch, the Committee held an extended discussion on risk communication. Deering suggested that the Committee have two public meetings in Nevada to get public opinion on the regulatory agencies and to discuss the regulatory process applicable to Yucca Mountain. Garrick noted that the Committee should not present itself as experts on risk communication; rather, the Committee is there to communicate better with the public. Mike Scott (M&O) passed on a remark from Judy Treichel that meetings in the September-October time frame should be avoided because the public will be heavily involved in reviewing the EIS. Tim McCartin noted that the paper to the Commission on defense in depth will recommend a future public meeting prior to issuance of the final Part 63 rule. Dr. Wymer noted that the discussion should include the subject of risk. Hornberger responded that the public does not want to hear about relative risk. Larkins said that the success of a discussion would depend on the audience.

6.2.6.2 Drip Shield Emplacement Error

Probability

The current engineered barrier design (CRWMS M&O 1999c, p. O-13) includes a titanium drip shield that would be placed over the WPs at the time of repository closure to provide defense-in-depth for postclosure performance. The drip shield will be continuous down the entire length of the drift, and will be fabricated and emplaced in 1.8 meter long segments (see Attachment V). Emplacement of the drip shield will be accomplished remotely by using a mobile gantry (CRWMS M&O 1998c, p. 24). Each segment will slightly overlap the previously emplaced segment. Installation of the drip shield segments will occur just prior to closure of the MGR. Once the drip shield has been installed down the entire length of the drift, backfill will be placed over the drip shield using a belt conveyor on a mobile gantry (CRWMS M&O 1998c, p. 15).

The benefits of the drip shield could be diminished for a particular package if the operator fails to overlap the drip shield with the previously emplaced segment, such that a large separation exists that would allow any dripping water above it to directly fall onto the package below. To estimate the probability of this occurring, the event sequence tree shown in Figure 6.2-10 was developed. The assumptions used to develop the event sequence tree are summarized in Section 5.8. The following human error probabilities have been used to quantify the tree:

- The probability that the operator fails to properly place the drip shield such that it overlaps the previously placed drip shield is based on the HEP for improperly mating a connector, 0.003 (Swain and Guttmann 1983, pp. 20-28). Since there are two drip shield joints per package for 1.8 m drip shield segments, the probability of having at least one improperly mated joint over a WP is 0.006. Since the drip shield is larger than the type of connector for which the HEP was developed, and is being mated remotely, the maximum error factor of 3 (Swain and Guttmann 1983, pp. 20-28) is applied, for a final probability of 0.0178 per WP for a misplaced drip shield.
- Failure of the operator self-check, and the QA check of emplaced drip shields, are approximated by a check failure using written procedures with an HEP of 0.1 (Swain and Guttmann 1983, pp. 20-38).

Table 6.2-7 provides detailed descriptions of the actions in the drip shield emplacement error event sequence tree. This tree was quantified in the "Drip Shield" sheet of the Excel 97 spreadsheet Seq-Trees.xls (see Attachment II).

Three series of experiments were run:

- Sorption of Pu(IV) and Pu(V) on 1,000 mg/L suspensions of each colloid (hematite, goethite, montmorillonite, two types of silica)
- Sorption of ^{243}Am on 200 mg/L suspensions of each of the above colloids except goethite
- Sorption of Pu(V) and ^{243}Am on suspensions of each of the above colloids except goethite for concentrations of 10, 50, 100, 150, 200, 1,000, and 5,000 mg/L

Not all of the combinations listed were reported. The ^{243}Am experiments in Series 3 apparently were not complete at the time the report was prepared. Partition coeffiient (K_d) values obtained in Series 1 and 2 are shown in Table 3-25.

In addition to the sorption measurements, data were reported on the desorption of Pu(IV) and Pu(V) from Pu-loaded samples of hematite, goethite, montmorillonite, and silica colloids after 96 hr of sorption. After 150 days of agitation in natural or synthetic J-13 water, desorption was far from complete for any of the colloids, per the following:

- Essentially no Pu was desorbed from hematite.
- Less than 1 percent of Pu was desorbed from goethite.
- About 8 percent of Pu(IV) and less than 1 percent of Pu(V) was desorbed from montmorillonite.
- About 20 percent of Pu(IV) and 6 percent of Pu(V) was desorbed from silica colloids.

Based on this laboratory work, a bounding sorption coefficient value of $K_d = 7 \times 10^5$ mL/g is used in this model. These laboratory results show that sorption is irreversible for Pu. It is possible that sorption on montmorillonite (and possibly other clays) and silica could be reversible over a time of decades, but complete reversibility is not evident from the reported data. However, given the large amount of steel in the emplacement drifts, iron colloids are likely to be present in greater amounts. Also, they have greater affinity for Pu, so behavior of the siliceous colloids is less important to radionuclide transport.

3.1.2.6.4 Bounding Values of Radionuclide Solubility Enhancement Factor

The enhancement factor for colloidal transport represents the factor by which the total amount of a radionuclide in groundwater is increased over the amount in solution at the time the radionuclide was adsorbed. For irreversible adsorption, the amount in solution may be equal to, less than, or greater than the solubility limit at any point along the flow path after the colloid is "loaded" with radionuclide. Mathematically, the enhancement factor E is defined by

$$E = \frac{C_{Total}}{C_0} = 1 + K_d M_C \qquad \text{(Eq. 3-19)}$$

The following description is derived from deliverable RPA176M3, *Summary of Work in Progress for Reconfiguration of the Waste Handling Building*/letter LV.SFO.GWG.9/99-070 w/enclosures, 9/30/99, Attachment VI, Dry Vault Storage Concept (CRWMS M&O 1999g).

Since the staging or inventory of waste may be required in varying amounts with time, depending on the heat output of the fuel arriving, a modular system was developed. A 500 MTHM, 300-basket inventory facility was selected as the smallest vault. For operational efficiency, a two-vault dry fuel inventory building concept was developed as the module size. The building would have an inventory capacity of 1,000 MTHM, therefore five such buildings would be interconnected to provide the maximum calculated requirement of 5,000 MTHM.

Each dry inventory building would be 290 ft by 252 ft (88.4 m x 76.8 m) with a height of 63 ft (19.2 m) and contain two parallel vaults, each 290 ft by 70 ft (88.4 m x 21.3 m), separated by an operating corridor 290 ft by 40 ft (88.4 m x 12.2 m), Figure V-1. The remaining space is taken up by HVAC equipment and HVAC support areas. The vaults have 5-ft (1.5 m) thick reinforced-concrete walls and ceiling. Each vault has two levels. The first level is 21 ft (6.4 m) high and contains 300 vertical commercial spent nuclear fuel (CSNF) basket sleeves in a 30 sleeve by 10 sleeve array, Figure V-2. The cylindrical sleeves have an external diameter of 42 in. (107 cm). The sleeves contain a 28 in. (71 cm) square inner column that receives the 25 in. (63.5 cm) square CSNF basket. The inner square sleeve and the outer circular sleeve serve as primary and secondary contamination control barriers, respectively. The second level of the vault is a 30-ft (9.1 m) high basket-handling cell that contains the crane for handling fuel baskets. The two levels of the vault are separated by a 2 ft (61 cm) concrete barrier with penetrations at each sleeve. Each of the vaults in the dry fuel inventory building, as well as other dry fuel inventory buildings, are interconnected by fuel basket transfer tunnels. These tunnels extend to the assembly transfer system (ATS) in the Waste Handling Building (WHB).

CSNF assemblies will be extracted from shipping casks in the ATS hot cells and placed in fuel baskets for further handling. The fuel baskets are designed to hold 4 pressurized water reactor (PWR) assemblies or 8 boiling water reactor (BWR) assemblies (Figure V-3). The fuel baskets will be placed in a transfer cart in the ATS and transferred through transfer tunnels to the vault in one of the dry fuel inventory buildings. In a vault, individual baskets will be removed from the transfer carts by the fuel basket crane. The crane will transfer the basket to a sleeve. When the basket is in the sleeve, the crane will install a thermal plug into the sleeve over the basket. The remote fuel basket crane is tele-operated or automatically controlled from an operating gallery adjacent to the basket handling cell.

The sleeves are designed to transfer heat from the baskets to the outer circular barrier. Air will be forced through the sleeve level to remove heat from the outer surfaces of the sleeves.

The heat-loads of the stored basket assemblies will be recorded. As the heat load of assemblies in individual, stored baskets match those required for blending in waste packages, the baskets will be removed from the sleeves by the fuel basket crane and placed in a transfer cart for transportation through the fuel transfer tunnels back to the ATS. There, individual assemblies will be removed from the basket and placed in a waste package.

To simulate the impact of the EBS random release on system performance at the Yucca Mountain site, the FEHM EBS random release model was developed to perform the following tasks:

- Locate the M_fine early failed package nodes in repository sub-regions based on given failed package coordinates. If no node matches a given coordinate, then select the nearest node to the given coordinate.

- Randomly select the failed package nodes in the designated sub-region i.

The existing FEHM subroutine *getrip* was modified to handle EBS random release. From FEHM particle-tracking subroutine *part_track*, subroutine *getrip* is called to determine the particle release locations. First, the subroutine obtains information passed by GoldSim in an input array called in[ ]. The structure of the in[ ] array is shown in Figure 7.

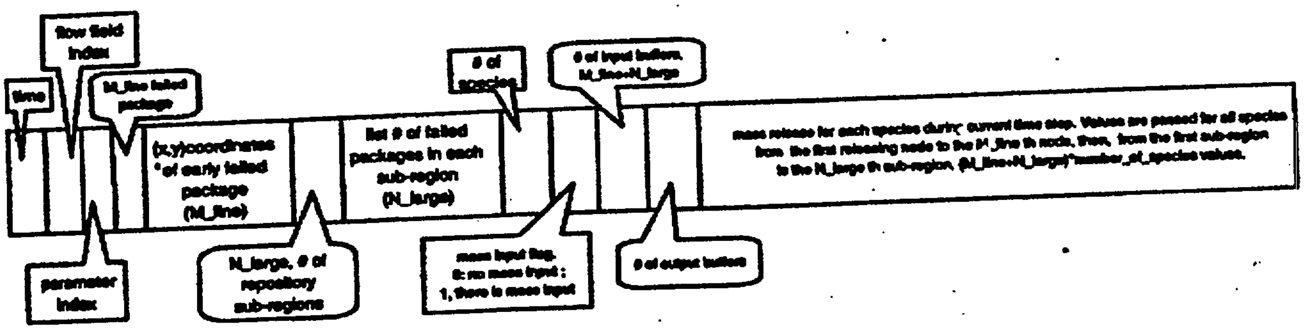


Figure 7. The Structure of the in[ ] Array Passed to FEHM from GoldSim

The algorithm used in FEHM EBS random release model is summarized in Figure 8, the flow chart of the EBS random release model.

Starting with the M_fine early failed packages, *getrip* extracts the (x,y) coordinates of the early failed packages and loops through each repository sub-region node to select the one that is closest to the given coordinates. To prevent a node being selected more than once for two or more given coordinates, getrip checks the selected nodes for overlapping. If overlapping is found, *getrip* prints out error messages to the error file fehmn.err, then stops the program.

6.2.2.6 Codes and Standards Criterion

6.2.2.6.1 Comply With Assumptions in "Monitored Geologic Repository Project Description Document" (1.2.6.2, CRWMS M&O 2000k)

The only assumption identified in the *Monitored Geologic Repository Project Description Document* (MGR PDD, CRWMS M&O 1999b) that is significant to emplacement pallet design is CPA 039 – Enhanced Design Alternative II Design Definition for Performance Assessment, Waste Package Operations, and Engineered Barrier System Operations. The applicable portions of this assumption are as follows:

"In addition, performance assessment, Waste Package Operations, and Engineered Barrier System Operations will assume for SR that:

- the invert ballast material is crushed tuff,
- the backfill material is Overton sand,
- the free-standing drip shield is of "mailbox" shape and with uninterrupted coverage for the entire length of the emplacement drift, and
- the average heat output per waste package for pressurized water reactor commercial SNF at the time of emplacement will be 11.3 kW, and the average heat output per waste package for all waste packages at the time of emplacement will be 7.9 kW."

This assumption, specifically that the invert ballast material is crushed tuff, is consistent with the parameters used when evaluating the emplacement pallet under static loading in CRWMS M&O 2000f.

6.2.2.7 Criteria From Other SDDs

These criteria, while not directly specified for the emplacement pallets, should be applied since the emplacement pallet is being used for lifting of the waste package. Future revisions to the *Emplacement Drift System Description Document* (CRWMS M&O 2000k) should include these criteria.

6.2.2.7.1 Retrieval for Up to 300 Years After Start of Emplacement (1.2.1.8, CRWMS M&O 1999e) (1.2.1.9, CRWMS M&O 1999f) (1.2.1.6, CRWMS M&O 1999g)

Supporting retrieval requires the emplacement pallet to stay in a condition that can be lifted at the end of, or during, the prescribed period of time. This is verified in the *Structural Calculations for the Lifting of a Loaded Emplacement Pallet* (CRWMS M&O 2000d), see Section 6.4.

6.12 HORIZONTAL ANISOTROPY

Anisotropy occurs when the permeability is directionally dependant. A recent study by Winterle and La Femina (1999) concluded that a conceptual model of horizontal anisotropy in the tuff aquifer is reasonable and flow in the tuff aquifer is believed to occur in a fracture network that exhibits a preferential north-south strike azimuth. In addition, north to north-northeast striking structural features are optimally oriented perpendicular to the direction of least principal horizontal compressive stress, suggesting tendency toward dilation and potentially higher permeability (Ferrill et al. 1999, p. 5). They recommended that hydrogeologic conceptual models for site-scale flow should consider the potential effects of horizontally anisotropic transmissivity. Performance assessment could be impacted by the inclusion of horizontal anisotropy because the flow could be diverted to the south causing transported solutes to remain in the fractured volcanic tuff for longer distances before moving into the valley fill/alluvial aquifer. The importance of this is that a more southward flow path would increase travel distances in the tuff and reduce the amount of flow in the alluvium (Ferrill et al. 1999, p. 7). A reduction in the flow path length in the alluvium would decrease the amount of total radionuclide retardation that could occur for those radionuclides with greater sorption coefficients in alluvium than in fractured volcanic rock matrix. In addition, potentially limited matrix diffusion in the fractured volcanic units could lead to shorter travel times in the volcanic units relative to the alluvium.

There is significant uncertainty in the appropriate model parameter value for horizontal anisotropy due to lack of data on the variability in the horizontal anisotropy over the scale of the transport path length. Winterle and La Femina (1999) estimated values for anisotropic transmissivity using data from the pumping tests at the C-well complex. These estimates are poorly constrained and may not be representative of the anisotropy on the scale of the transport model. The uncertainties include differences in pumping test analysis methods, the fact that only a minimum number of observation wells were used, and the additional uncertainty regarding the validity of assuming a homogenous effective continuum over the scale of the test (Winterle and La Femina 1999, p. 4-29). Given this uncertainty in anisotropy, and to simplify the model, the potential effects of anisotropy are bounded by setting the anisotropy ratio to 1 (isotropic) or 5 (based on the C-well data).

The stochastic parameter HAVO determines whether horizontal anisotropy is applied to a given realization for the TSPA calculations. The HAVO parameter is uniformly distributed from 0.0 to 1.0. For a value of 0.0 to 0.5 the isotropic groundwater flow field is used in radionuclide transport simulations. For a value of 0.5 to 1.0 the anisotropic groundwater flow field is used. The horizontal anisotropy ratio of 5:1 is imposed in the SZ site-scale flow model by multiplying values of horizontal permeability in the north-south direction by 2.24 and dividing values of horizontal permeability in the east-west direction by 2.24. This modification of the permeability field of the SZ site-scale model is applied to the volcanic units in an area beneath and to the south and east of Yucca Mountain.

Section 6.6.6, the results of Section 6.6.5 are simplified for a system with only one waste package. This form is useful in implementing a performance assessment.

6.6.1 Wet Unzipping Abstraction

Fuel rods with perforated cladding are expected to remain intact until the WP breaches and permits air and moisture to enter. While the humidity is low, dry unzipping could occur. Since the WP is expected to remain intact for at least 200 years, the fuel temperatures will be too low for dry unzipping (fuel conversion to U_3O_8) to occur. Wet unzipping of failed rods is analyzed to start when the WP breaches. Rods that fail after WP breach immediately start to unzip. The fuel matrix is dissolved at the intrinsic dissolution rate that is evaluated for the current temperature and in-package chemistry. The dissolved UO_2 precipitates locally as metaschoepite. This secondary phase isolates most of the fuel from the moisture and increases in volume compared to UO_2. In time, the cladding in the reaction region is torn as the reaction continues. This reaction region is cone shaped and the cone angle is based on experimental observations in dry unzipping and theoretical analyses.

The unzipping propagates along the rod at a rate that is proportional to the intrinsic dissolution rate. The ratio of unzipping speed to intrinsic dissolution rate is given by a triangular distribution. The distribution has a minimum of 1, a mode of 40, and a maximum of 240 (CRWMS M&O 2000e, Section 7). It is assumed that the perforation is at the middle of the fueled length of the rod. This maximizes the release rate. The unzipping time or velocity is also a function of local chemistry and pH. In TSPA-SR, the unzipping velocity and fraction of fuel exposed are evaluated at each time step because of the evolution of in-package chemistry and temperature. Section 6.6.6 gives equations for determining the rate at which fuel is exposed as a result of unzipping. These equations take into account the effect of having fuel rods fail at different times.

6.6.2 Intrinsic Dissolution Abstraction

The intrinsic dissolution rate is used in the unzipping calculations to determine the reaction rate velocity. The intrinsic dissolution equation is to be applied at each TSPA-SR simulation time step and is to be based on the local chemical conditions. At some times the pH could be basic and at other times it could be acidic (CRWMS M&O 2000i). The abstraction is therefore divided into regions of pH greater than and less than a neutral pH (pH = 7). CRWMS M&O 2000d (p. 82) develops the intrinsic dissolution equations that are recommended for TSPA-SR.

For basic conditions (pH > 7),

$$\mathrm{Log}_{10}\, Dr = a_0 + a_1 / Tk + a_2 \cdot PCO_3 + a_3 \cdot PO_2 \qquad \text{(Eq. 6.6-1)}$$

where $a_0 = 4.69$, $a_1 = -1085$, $a_2 = -0.12$, and $a_3 = -0.32$.

For acid conditions (pH < 7),

$$\mathrm{Log}_{10}\, Dr = a_0 + a_1 / Tk + a_3 \cdot PO_2 + a_4 \cdot pH \qquad \text{(Eq. 6.6-2)}$$

where $a_0 = 7.13$, $a_1 = -1085$, $a_3 = -0.32$, and $a_4 = -0.41$.

ATTACHMENT I NORMALIZED INFILTRATION RATES

The repository block model developed in Attachment IV, shape1.dat (see Figure I-1), is divided into 31 sections. The block model is composed of a rectangle with a smaller rectangle attached to the southern half of the west boundary of the repository. The 31 sections of the block model are derived by divided the block model into 4 columns with seven rows, plus one additional column (3 rows) in the extension on the southwest side of the repository (Table I-1 and Figure I-1). The location of the 31 elements (Table I-1) is easily checked with coordinate geometry. One example is given:

The Northern row of elements are L1c1-L1c4, as shown in the example below.. To check their spacing simply find the distance between the points and then verify that the slope of the line segments between points is similar. The similar distances and slopes between points verifies that the first row of points represent block elements of similar size. Calculations presented in Table I-1 verify that the repository block elements are similarly sized. The information in Table I-1 is in the file column.data (Attachment VI).

ID	Easting (ft)	Northing (ft)	Points	Distance (ft)	Slope (radians)
l1c1	171234.3	235534.8	c1-c2	236.7	-0.053
l1c2	170997.9	235547.3	c2-c3	236.7	-0.053
l1c3	170761.5	235559.9	c3-c4	236.7	-0.053
l1c4	170525.1	235572.4			

(Portion of Table I-1)

Note: Slope is the quotient of ΔY and ΔX.

The average infiltration rate in the modeled repository is different than the average infiltration rate in the actual repository. To offset this difference, the infiltration rates at the 31 locations are normalized (Table I-2). The normalized infiltration rate is the product of the estimated infiltration rate and a normalization factor. The normalization factor is the quotient of the average normalized infiltration and the actual infiltration. The average normalized infiltration is the average of the estimated infiltration at the 31 block element locations (Attachment VI, *.out). The average actual infiltration is included in the output from Columninfiltration (Attachment VI, *.out).

divided by $\nu^I(x|\theta^{0.95})$ to produce joint distributions for L^I_m and ϕ_m conditional on the 5th and 95th percentile values for the frequency of intersection.

In summary, the mathematical formulation for computing the conditional distribution for the length and azimuth of intersecting dikes within the potential repository footprint is developed directly from the PVHA formulation presented in CRWMS M&O (1996, Section 3 and Appendix E) without invoking any additional assumptions. The formulation for computing the conditional distribution for the number of eruptive centers occurring within the potential repository footprint requires additional assumptions in order to assess the number of eruptive centers per volcanic event and the spatial distribution of eruptive centers along the length of the dike. Five alternative approaches are developed to implement these assumptions to span the range of available approaches. Calculations are performed for all five approaches to indicate the sensitivity of the results. As a final step, relative weights are assigned to the five approaches in order that a composite result can be obtained. The five approaches are summarized below:

1. The Independent, Uniformly Distributed, Uncorrelated (*IUD-UC*) approach. The distribution for the number of eruptive centers per volcanic event is derived from the PVHA experts' interpretations. These distributions are uncorrelated with the distributions for dike length. The location for each eruptive center is defined by a uniform distribution over the total length of the dike, and if multiple eruptive centers occur in a volcanic event, the distributions for their locations are independent.

2. The Independent, Uniformly Distributed, Correlated (*IUD-C*) approach. The distribution for the number of eruptive centers per volcanic event is derived from the PVHA experts' interpretations. These distributions are completely correlated with the distributions for dike length. The location for each eruptive center is defined by a uniform distribution over the total length of the dike, and if multiple eruptive centers occur in a volcanic event, the distributions for their locations are independent.

3. The Uniformly Spaced, Randomly Distributed, Uncorrelated (*USRD-UC*) approach. The distribution for the number of eruptive centers per volcanic event is derived from the PVHA experts' interpretations. These distributions are uncorrelated with the distributions for dike length. The total length of the dike is divided into equal segments for each eruptive center. Within each segment, the location of the eruptive center is defined by a uniform distribution over the length of the segment.

4. The Uniformly Spaced, Randomly Distributed, Correlated (*USRD-C*) approach. The distribution for the number of eruptive centers per volcanic event is derived from the PVHA experts' interpretations. These distributions are completely correlated with the distributions for dike length. The total length of the dike is divided into equal segments for each eruptive center. Within each segment, the location of the eruptive center is defined by a uniform distribution over the length of the segment.

5. The Uniformly-Spaced, Randomly Distributed, Fixed Density (*USRD-FD*) approach. The number of eruptive centers per volcanic event is determined by dividing the total length of the dike by an average distance between eruptive centers derived from the PVHA experts'

7.3.3.2 Average Ballast Pressure – Scenario 2

In this scenario, the WP/pallet assembly is directly supported by compacted rock ballast with or without ties in between.

Case 1 – WP/pallet supported by ballast through ties

(a) Without drip shield and backfill

The maximum weight of each WP/pallet assembly is 88 MT (CRWMS M&O 2000g, Section 6.2), half of this weight, i.e., 44 MT, will be supported by ties through the contact area of each pallet. The bottom area of each pallet beam is 1845.2 (across the drift) x 552.4 (along the drift) mm (see Attachment III, p. III-1 of CRWMS M&O 2000k). It should be noted that the final tie spacing is not determined yet. For most railway track construction, it is about 20 to 25 inches (508 to 635 mm). For the emplacement drift invert design, the tie spacing may be selected to be smaller than 20 inches in order to have smaller load distribution factor. Therefore, based on the above dimensions, each pallet beam may be supported by more than one tie. For conservative design, the maximum load will be supported by one tie. It should also be noted that the above mentioned weights are only for static condition. In order to consider the effect of seismic force, the weights due to static loading should be multiplied by a factor of (1 + 0.182x1.5), in which 0.182 is a maximum underground vertical acceleration and 1.5 is a factor to account for seismic force to equivalent static loading (see Section 4.4.2). Tables 1 and 2 show the average ballast pressure under ties for different tie dimensions. Note that the ABP is derived by dividing the total weight including seismic force by the tie contact area.

Table 1. Average Ballast Pressure (psi) Under One Tie Subjected to WP/Pallet Assembly Weight

	L = 3.6 m	L = 3.8 m	L = 4.0 m	L = 4.2 m
C 9x20	96.8	91.8	87.1	83.0
C 10x20	87.1	82.6	78.4	74.7

Table 2. Average Ballast Pressure (psi) Under Two Ties Subjected to WP/Pallet Assembly Weight

	L = 3.6 m	L = 3.8 m	L = 4.0 m	L = 4.2 m
C 9x20	48.4	45.9	43.6	41.5
C 10x20	43.6	41.3	39.2	37.3

(b) With drip shield and backfill

Under this situation, the additional loads from the backfill bearing on the drip shield and the drip shield of 462.07 kips (i.e., 452.8+9.27, see Section 7.2.1) will be applied to the ties. For each drip shield with a length of 5.485 m (215.94 in.) (see Section 7.2.1), there will be 10.8 ties under the drip shield with a tie spacing of 20 in. Using rounded values of 10.8 ties, i.e., 11 ties, each tie will be subjected to an additional load of 42 kips (462.07/11 = 42). By using the same calculation method as in case (a), the average ballast pressure under one and two ties subjected to the weight of the WP/pallet assembly, the drip shield, and the backfill are shown in Tables 3 and 4.

Subsections 3.2 through 3.5 provide a general discussion of the roles and responsibilities for each M&O Management Team member. Further responsibilities related to meeting the requirements of the *Quality Assurance Requirements and Description* (DOE 2000) for each of the M&O Management Team members are described in appropriate M&O procedures.

3.2 M&O GENERAL MANAGER

It is the M&O General Manager's responsibility to negotiate and balance product scope to meet overall program funding and schedule targets. The M&O General Manager is responsible for providing the necessary support, guidance, and infrastructure. In addition, it is the responsibility of the M&O General Manager to:

- Ensure compliance with requirements to implement the *Civilian Radioactive Waste Management Strategic System Management Policy* (DOE 1998)
- Ensure compliance with planning guidance
- Ensure compliance with operating guidance
- Ensure compliance with technical, cost, and schedule baselines at the program level
- Ensure compliance with policies and procedures
- Define M&O Management Team expectations
- Define roles and responsibilities
- Ensure compliance with change control, funds management, property management, and contracts
- Ensure appropriate integration with DOE, the U.S. Geological Survey, and non-M&O entities under OCRWM control
- Ensure compliance with applicable regulatory and statutory requirements.

3.3 M&O FUNCTIONAL MANAGERS (DIRECTORS, DEPARTMENT MANAGERS, SECTION MANAGERS, AND LEADS)

M&O Functional Managers manage and perform the required technical work scope to support products, subproducts, and subproduct elements. It is the responsibility of M&O Functional Managers to assure the technical quality of products. M&O Functional Managers define the "how." The specific responsibilities of the M&O Functional Managers are planning, execution and implementation, monitoring and evaluation, and delivery.

1. INTRODUCTION

1.1 BACKGROUND

Since the late 1970s, Yucca Mountain has been studied for its suitability as a host for a potential Monitored Geologic Repository (MGR) to dispose of spent nuclear fuel (SNF) and high-level radioactive waste (HLW). Over this time, the design of the proposed repository has been developed and refined as additional information about the geologic setting and its waste isolation performance characteristics has been obtained and as the analytic modeling of the performance of the site as a repository has matured. Design changes have included improving the materials to separate the radioactive wastes from the natural system for longer periods of time, developing cost-effective ways to emplace the waste in larger disposal containers, and examining thermal management techniques that might create more favorable environments for the engineered components of the system and reduced impact on the natural system. The primary design variable used to control the natural environment to which the engineered system will be subjected has been the density at which radioactive material is loaded into the repository. This density is defined either in terms of metric tonnes of heavy metal (MTHM) per unit area, a mass density which does not vary with time, or in terms of a power density per unit area (or unit length within a given repository geometry), usually kilowatts (kW) per unit area. The power density is defined at the time the waste is emplaced and then decreases with time. The importance of this design variable is that it directly affects the temperature and temperature history in the emplacement areas and host rock.

In 1988, the *Site Characterization Plan Yucca Mountain Site, Nevada Research and Development Area, Nevada* (DOE 1988, Volume III, Chapters 6 and 7) was based on a design that created a thermal loading of nearly 57 kW per acre by packaging waste in relatively small, thin-walled containers made of stainless steel and Inconel and emplacing them in vertical boreholes. By the time of the *Mined Geologic Disposal System Advanced Conceptual Design Report* (CRWMS M&O 1996), the concept had changed to using larger waste packages comprised of a thick A516 carbon steel outer corrosion allowance barrier that provided the structural strength of the waste package and a nickel-based Alloy 825, corrosion-resistant, inner shell. These larger waste packages could contain large, thin-walled multi-purpose (storage, transportation, and disposal) canisters (MPCs) that could hold as many as 21 pressurized-water reactor (PWR) fuel assemblies. It was not practical to emplace these larger waste packages in vertical boreholes so the design changed to emplacing them in horizontal drifts. Although the MPC was no longer used in succeeding stages of development, the concept of large waste packages remained. The capability to handle MPCs, however, remains as a requirement. The *Viability Assessment of a Repository at Yucca Mountain* (DOE 1998, Volume 2, Section 4.2.1, p. 4-43) established a high thermal load of 85 metric tonnes of uranium (MTU)/acre with the waste contained in two-layer (corrosion allowance barrier made from A516 carbon steel, and corrosion-resistant inner shell, made from nickel-based Alloy 22) waste packages (DOE 1998, Volume 2, Section 5.1.2.1, p. 5-8). In the Viability Assessment design, waste packages are emplaced on average about five meters apart (creating a so-called point thermal load) in 5.5-m diameter drifts with a drift-center-to-drift-center spacing of 28 m (DOE 1998, Volume 2, p. 4-45). Once in the mountain, the decay of the radioactive waste heats the host rock and the water that resides in the interstices of the rock. The thermal load in the Viability Assessment design was high enough to boil the interstitial water and dry out the rock surrounding the entire

3.8 MINERALOGY DATA

The unqualified mineralogy data considered in this section were taken from DTN GS991299995215.001 and directly used in AMR U0085 (Fabryka-Martin et al. 2000, p 99). Although these data were collected and analyzed in accordance with the YMP-approved USGS QAPP; they are unqualified because they are preliminary and have not completed USGS internal reviews. These data sets have been assigned individual accession numbers within the data package cited in the DTN and are listed in Table 17.

Table 17. Types of Mineralogy Data

Data Set in DTN GS991299995215.001	Data Type
MOL.20000104.0008	Calcite cuttings concentrations
MOL.20000104.0009	Locations of microsamples collected for $^{207}Pb/^{235}U$ dating
MOL.20000104.0011	X-ray fluorescence elemental compositions
MOL.20000104.0012	Strontium isotope ratios and isotopic dilution data for rubidium and strontium

3.8.1 Calcite Cuttings Concentrations

Fabryka-Martin et al. (2000, p 100) use calcite cuttings data from data set MOL.20000104.0008 as one line of evidence in characterizing mineral coatings on fractures and cavities in the rock to infer relationships among infiltration, evaporation, and CO_2 mobility in fractures. Calcite concentrations were calculated according to a commonly used procedure. Samples were taken at five-foot intervals when drilling Borehole USW WT-24, which traverses welded Tiva Canyon (TCw) and the underlying PTn. The cuttings at each interval were homogenized and split, a representative 100 to 200 gram aliquot was acid-digested in a gas collection bulb, and the CO_2 concentration was determined by gas chromatography.

The resulting CO_2 data were plotted in AMR U0085, Figure 53 as calcite concentration (ppm) vs. depth (m). The results are discussed in AMR U0085, Section 6.10.1.1 (Fabryka-Martin et al. 2000, p 99). The data indicate larger concentrations of calcite in the near-surface TCw that decrease progressively downward in the PTn, but increase generally below the base of the PTn. This is interpreted as reflecting a progressively smaller amount of evaporation and CO_2 loss from open fractures with increasing depth in the PTn and, in addition, an apparent disparity between the amount of net infiltration and the amount of recharge to the PTn.

The data plotted in AMR U0085, Figure 53 are depicted with a linear depth scale and a logarithmic calcite concentration scale in a well-defined pattern. At a given depth the range of calcite concentration determinations varies from less than one order of magnitude to a maximum of approximately two orders of magnitude. The data are to be used to help evaluate a 3-D model of fracture mineral distribution, as well as to provide information for modeling fracture fluxes based on mineral data (Fabryka-Martin et al. 2000, p 100).

application, and was used within the range of validation in accordance with AP-SI.1Q, *Software Management*, as described in the software qualification report (CRWMS M&O 1998). The analysis was performed using a Gateway 2000 Personal Computer, CPU# 111161.

The biosphere model used by GENII-S was validated in accordance with AP-3.10Q and was used within the range of validation. The documentation of inputs, and outputs for the GENII-S code (Sandia National Laboratories 1998) follows in this attachment.

I-2 INPUT

Input parameters were developed in a series of analyses and documented in their respective Analysis and Model Reports (AMR). The data sets used in this calculation are listed in Table I-1. Selection of the values of input parameters, as well as justification of the applicability of the selected data to the specific exposure scenarios considered for the proposed Yucca Mountain repository, is described in the corresponding AMRs. These AMRs are also listed in Table I-1 together with the corresponding data tracking numbers (DTNs) for the developed data.

All relative behavioral factors were developed assuming that the near-equilibrium conditions exist in soil with regard to radionuclide buildup. Radionuclide removal mechanisms that were considered in the calculations were radioactive decay, leaching, and harvest removal. Radionuclide removal by soil erosion was not taken into account.

The general description of the GENII-S input can be found in CRWMS M&O (2000e, Figures 4 through 7 and Table 7). All of the input parameters were the same as those considered for the deterministic calculations of the Biosphere Dose Conversion Factors for the current climate condition except for the parameters related to the behavioral factor being evaluated, i.e., employment, recreation, diet, or drinking water. Input parameters which were modified to address a specific relative behavioral factor are listed below.

Employment factor
To develop employment factors only two pathways from the GENII-S Exposure Pathway Options menu were selected. They were External Ground Exposure and Inhalation Uptake. Inhalation exposure results from inhalation uptake of resuspended contaminated soil. External exposure results from radiation emitted by radionuclides in contaminated soil. Inhalation Exposure Time as well as Soil Exposure Time were set to 3120 hours/year (60 hours/week) (see Sections 5.1 and 6.1.1).

Recreational factor
Recreational factors use the same two pathways as employment factors, i.e., External Ground Exposure and Inhalation Uptake. The annual average time of recreation (Inhalation Exposure Time and Soil Exposure Time) used to calculate the recreational factor was set to a fixed value of 827 hours/year (136 minutes/day) (see Sections 5.2 and 6.1.2).

This next line steps down the 2000 lines of stress samples

2) For i = 0 To 1999

Zone 1

This next line copies the temperature shaping term from Sheet "Creep-WP", Row i+7, Column 4 (Column D) to Sheet "Creep-Rod", Location B3.

3) Sheets("Creep-Rod").Range("B3").Value = Sheets("Creep-WP").Cells(i + 7, 4).Value

This next line copies the free volume of each rod from Sheet "Creep-WP", Row i+7, Column 28 (Column AB) to Sheet "Creep-Rod", Location B2.

3b) Sheets("Creep-Rod").Range("B2").Value = Sheets("Creep-WP").Cells(i + 7, 28).Value

This next line copies the stress value from "Creep-WP", Row i+7 Column 3 (Column C) to Sheet "Creep-Rod", Cell B4.

4) Sheets("Creep-Rod").Range("B4").Value = Sheets("Creep-WP").Cells(i + 7, 3).Value

The next line shifts the resulting calculated strain from "Creep-Rod" Cell B5 to "Creep-WP", row i+7, Column 5 (Column E).

5) Sheets("Creep-WP").Cells(i + 7, 5).Value = Sheets("Creep-Rod").Range("B5").Value

For the next 5 WP radial zones the temperature shaping index is copied to "Creep-Rod" and the resulting creep strain is written into "Creep-WP". The same values of stress and free volume are used in all zones as was used for Zone 1 above.

Zone 2

6) Sheets("Creep-Rod").Range("B3").Value = Sheets("Creep-WP").Cells(i + 7, 6).Value
7) Sheets("Creep-WP").Cells(i + 7, 7).Value = Sheets("Creep-Rod").Range("B5").Value

Zone 3

8) Sheets("Creep-Rod").Range("B3").Value = Sheets("creep-WP").Cells(i + 7, 8).Value
9) Sheets("Creep-WP").Cells(i + 7, 9).Value = Sheets("Creep-Rod").Range("B5").Value

Zone 4

10) Sheets("Creep-Rod").Range("B3").Value = Sheets("Creep-WP").Cells(i + 7, 10).Value
11) Sheets("Creep-WP").Cells(i + 7, 11).Value = Sheets("Creep-Rod").Range("B5").Value

Zone 5

12) Sheets("Creep-Rod").Range("B3").Value = Sheets("Creep-WP").Cells(i + 7, 12).Value
13) Sheets("Creep-WP").Cells(i + 7, 13).Value = Sheets("Creep-Rod").Range("B5").Value

Zone 6

14) Sheets("Creep-Rod").Range("B3").Value = Sheets("Creep-WP").Cells(i + 7, 14).Value
15) Sheets("Creep-WP").Cells(i + 7, 15).Value = Sheets("Creep-Rod").Range("B5").Value

Bottom of the i "Do Loop".

16) Next i

Alarm to announce problem is complete

17) Beep
18) Beep
19) Beep
20) Beep
21) End Sub

Test: The macro can be tested the following ways:

profiles consisting of the fraction of waste packages failed versus time and the average (per waste package) number of penetration openings versus time. The degradation profiles are used as input into the TSPA model (see FEP 2.1.03.01.00 Corrosion of Waste Containers).

6.2.7.7 Supplemental Discussion:

Item intentionally left blank.

6.2.7.8 Relevant References:

CRWMS M&O 2000n. *Abstraction of Models for Pitting And Crevice Corrosion of Drip Shield and Waste Package Outer Barrier.* ANL-EBS-PA-000003

6.2.7.9 Treatment of Secondary FEPs:

The following is a discussion of the secondary FEPs addressed by Primary FEP 2.1.03.03.00, Pitting of Waste Containers and Drip Shields.

Secondary FEP Number and Name: 2.1.03.03.01, Localized corrosion (of Waste Container).

Relationship to Primary FEP: This secondary FEP and the associated primary FEP both address localized (pitting and crevice) corrosion on a waste package container.

Screening and Disposition: Redundant. See the Screening Argument and TSPA Disposition for primary FEP 2.1.03.03.00, Pitting of Waste Containers and Drip Shields (Sections 6.2.7.5 and 6.2.7.6 of this WP FEPs AMR). This secondary FEP is included in the TSPA analysis.

Secondary FEP Number and Name: 2.1.03.03.02, Pitting (of Waste Container).

Relationship to Primary FEP: This secondary FEP and the associated primary FEP both address localized (pitting and crevice) corrosion on a waste package container.

Screening and Disposition: Redundant. See the Screening Argument and TSPA Disposition for primary FEP 2.1.03.03.00, Pitting of Waste Containers and Drip Shields (Sections 6.2.7.5 and 6.2.7.6 of this WP FEPs AMR). This secondary FEP is included in the TSPA analysis.

Secondary FEP Number and Name: 2.1.03.03.03, Pitting corrosion develops on containers

Relationship to Primary FEP: This secondary FEP and the associated primary FEP both address localized (pitting and crevice) corrosion on a waste package container.

Screening and Disposition: Redundant. See the Screening Argument and TSPA Disposition for primary FEP 2.1.03.03.00, Pitting of Waste Containers and Drip Shields (Sections 6.2.7.5 and 6.2.7.6 of this WP FEPs AMR). This secondary FEP is included in the TSPA analysis.

3.2.4 October Census

For cool-season species at the October census, irrigated plots had significantly higher plant densities than control plots with the exception of those seeded in December (Table 4a, interaction between seeding date and irrigation, Figure 7a). Irrigated and control plots that were seeded in December had similar plant densities (Figure 7a). Densities decreased on irrigated plots that were seeded in March and April compared to those seeded earlier (Table 4a, Figure 7a). Plant density was about 3 plants/m^2 on control plots that were seeded in February, March and April. Without irrigation, cool-season species achieved a high density only when planted in December. These results are similar to those observed at the August census for cool-season species (Figure 6a and b).

Table 4. Results of a two-way analysis of variance examining the effects of seeding date and irrigation on plant density of a) cool-season species, b) warm-season, c) annual forbs, and d) annual grasses on multi-species mix study plots for the October 1997 census date at Yucca Mountain. Data were log transformed for these analyses. DTN: MO0006SEPPLDEN.002

a) Cool-season

Source	Sum of Squares	df	Mean Square	F-ratio	P
Seed date	10.636	3	3.545	31.950	< 0.001
Irrigation	12.379	1	12.379	111.601	< 0.001
Seed date x Irrigation	4.573	3	1.524	13.743	< 0.001
error	2.662	24	0.111		

b) Warm-season

Source	Sum of Squares	df	Mean Square	F-ratio	P
Seed date	0.252	3	0.084	1.256	0.312
Irrigation	2.066	1	2.066	30.834	< 0.001
Seed date x Irrigation	0.994	3	0.331	4.943	0.008
error	1.608	24	0.067		

c) Annual forbs

Source	Sum of Squares	df	Mean Square	F-ratio	P
Seed date	1.824	3	0.608	1.882	0.161
Irrigation	9.826	1	9.826	30.408	< 0.001
Seed date x Irrigation	12.939	3	4.313	13.348	< 0.001
error	7.432	23	0.323		

d) Annual grasses

Source	Sum of Squares	df	Mean Square	F-ratio	P
Seed date	17.426	3	5.809	4.716	0.010
Irrigation	12.754	1	12.754	10.355	0.004
Seed date x Irrigation	6.483	3	2.161	1.754	0.183
error	29.562	24	1.232		

alternative physical pathways for such transport processes. Since none of the likely degraded configurations are found to be critical, the additional effort of the more detailed calculations is not necessary.

Since none of the configurations have been found to be critical, there have been no probability calculations. Nevertheless the configurations described in Sections 6.2.1.2 are believed to be the most likely, since the N-reactor SNF is the most rapidly degrading material in the waste package.

6.2.1.2 Degraded Spent Nuclear Fuel with Nearly Intact MCO and Waste Package Shells

In this case, the spent nuclear fuel could be partially or fully degraded. As a variation, the other internal components of MCO could also be degraded. This configuration is a variation of configuration class 6 and can be reached from standard scenario IP-1. The results of criticality calculations for this configuration are given in Section 7.4.1.

6.2.1.3 Spent Nuclear Fuel Degrades After Degradation of MCO Basket

In this case, the spent nuclear fuel is intact and the degradation scenarios and configuration classes are applied to the MCO and its contents including the baskets and the center post. As a variation, there could be partial degradation of the MCO internals and partial-to-full degradation of the spent nuclear fuel. This configuration is a variation of configuration class 1 and can be reached from standard scenario IP-3. The results of the criticality calculations for this configuration are given in Section 7.4.2.

6.2.1.4 Nearly Intact MCO and Degraded Waste Package Internals

In this case, the concepts of scenario and configuration are applied to the entire waste package. The fuel element and the MCO shell retain their initial configuration. This configuration is a variation of configuration class 1 and can be reached from standard scenario IP-3. The results of the criticality calculations for this configuration are given in Section 7.5.1

6.2.1.5 Degraded MCO and Waste Package Internals, Intact Spent Nuclear Fuel

In this case, the degradation scenario and configuration are applied to the entire waste package. The MCO and waste package internals including the high-level waste glass canisters are degraded. The fuel elements are intact. This configuration is a variation of configuration class 1 and can be reached from standard scenario IP-3. The results of the criticality calculations for this configuration are given in Section 7.5.1.

6.2.1.6 Completely Degraded MCO and Waste Package Internals

In this case, the degradation scenario and configuration are applied to the entire waste package including the spent nuclear fuel. Degradation products from the MCO and the waste package and their contents mix uniformly inside the waste package. This configuration is a variation of configuration class 2 and can be reached from standard scenario IP-1, IP-2, or IP-3. The results of the criticality calculations for this configuration are given in Section 7.5.2.

APPENDIX B.

Non-stopword accuracies and Unique non-stopword accuracies for OCR output

Document Accession Number	ALL Non-Stopwords (N)	orig	gen0	gen1	gen2	gen3	gen4	Unique Non-Stopwords (M)	orig	gen0	gen1	gen2	gen3	gen4
mol199907200407	Non-Stopword Accuracy	97.75	97.33	95.80	95.14	93.20	87.81	Non-Stopword Accuracy	99.45	99.01	98.68	98.57	97.37	96.49
	Accuracy of Characters in Non-Stopwords	99.49	99.45	98.21	98.02	97.43	96.43	Accuracy of Characters in Non-Stopwords	99.89	99.77	99.00	98.97	98.73	98.34
mol199911010207	Non-Stopword Accuracy	97.37	97.26	96.57	96.11	93.53	88.39	Non-Stopword Accuracy	98.95	98.72	98.60	98.14	97.27	95.76
	Accuracy of Characters in Non-Stopwords	99.18	99.51	99.38	99.29	98.65	97.07	Accuracy of Characters in Non-Stopwords	99.74	99.75	99.78	99.64	99.44	99.18
mol200002170216	Non-Stopword Accuracy	98.25	96.96	96.27	93.35	88.14	89.84	Non-Stopword Accuracy	98.46	98.11	97.51	96.68	91.42	93.90
	Accuracy of Characters in Non-Stopwords	99.30	99.34	99.05	97.66	96.53	96.78	Accuracy of Characters in Non-Stopwords	99.59	99.59	99.29	99.02	97.69	98.34
mol200002280529	Non-Stopword Accuracy	98.49	98.80	98.60	97.43	95.73	95.82	Non-Stopword Accuracy	98.53	99.08	98.44	98.35	97.33	97.79
	Accuracy of Characters in Non-Stopwords	99.57	99.73	99.57	99.17	98.64	98.75	Accuracy of Characters in Non-Stopwords	99.66	99.81	99.58	99.72	99.20	99.45
mol200004130692	Non-Stopword Accuracy	97.81	97.50	97.28	94.51	92.93	92.47	Non-Stopword Accuracy	99.36	98.97	98.84	98.07	97.94	97.69
	Accuracy of Characters in Non-Stopwords	99.60	99.56	99.58	98.58	97.85	98.17	Accuracy of Characters in Non-Stopwords	99.85	99.82	99.77	98.80	99.61	98.71
mol200004140874	Non-Stopword Accuracy	97.96	97.58	97.26	95.89	94.14	93.26	Non-Stopword Accuracy	98.64	97.51	97.65	96.95	96.20	95.96
	Accuracy of Characters in Non-Stopwords	99.48	99.42	99.43	98.99	98.38	98.37	Accuracy of Characters in Non-Stopwords	99.46	99.23	99.27	99.13	98.82	98.51
mol200005230155	Non-Stopword Accuracy	97.67	96.74	96.53	95.48	93.04	92.67	Non-Stopword Accuracy	99.09	99.09	98.83	98.44	97.92	97.14

	Accuracy of Characters in Non-Stopwords	99.65	99.39	99.37	98.99	98.25	98.44	Accuracy of Characters in Non-Stopwords	99.80	99.82	99.80	98.92	99.69	98.66
mol200005250378	Non-Stopword Accuracy	98.02	97.64	97.35	96.22	94.85	93.53	Non-Stopword Accuracy	98.30	97.53	97.81	96.92	97.12	96.55
	Accuracy of Characters in Non-Stopwords	99.48	99.46	99.44	99.20	98.73	98.23	Accuracy of Characters in Non-Stopwords	99.47	99.28	99.38	99.24	99.29	99.10
mol200005260336	Non-Stopword Accuracy	98.42	97.67	97.29	95.43	92.85	92.04	Non-Stopword Accuracy	98.91	98.07	97.90	97.65	96.31	96.06
	Accuracy of Characters in Non-Stopwords	99.74	99.55	99.54	98.68	97.57	97.84	Accuracy of Characters in Non-Stopwords	99.76	99.66	99.67	99.27	98.73	98.86
mol200006090266	Non-Stopword Accuracy	98.63	98.63	98.44	97.52	94.70	93.84	Non-Stopword Accuracy	99.06	99.32	99.15	98.55	96.77	96.51
	Accuracy of Characters in Non-Stopwords	99.76	99.83	99.79	99.63	98.80	98.78	Accuracy of Characters in Non-Stopwords	99.77	99.85	99.80	99.68	99.19	99.00
mol200006270254	Non-Stopword Accuracy	98.47	98.14	97.86	96.97	95.66	95.38	Non-Stopword Accuracy	98.47	98.47	98.01	98.24	96.79	97.17
	Accuracy of Characters in Non-Stopwords	99.60	99.57	99.55	99.17	99.01	99.02	Accuracy of Characters in Non-Stopwords	99.52	99.54	99.47	99.48	99.29	99.34
mol200007250453	Non-Stopword Accuracy	96.07	95.99	95.56	93.36	91.17	90.68	Non-Stopword Accuracy	85.02	84.83	84.59	83.64	82.35	81.87
	Accuracy of Characters in Non-Stopwords	98.97	98.98	98.91	97.66	97.02	97.03	Accuracy of Characters in Non-Stopwords	99.42	99.40	99.36	98.60	98.28	98.17
mol200011220005	Non-Stopword Accuracy	96.82	96.33	96.14	95.43	92.54	91.99	Non-Stopword Accuracy	98.32	98.01	97.83	97.08	95.78	95.72
	Accuracy of Characters in Non-Stopwords	99.29	99.18	99.24	99.08	97.85	97.57	Accuracy of Characters in Non-Stopwords	99.52	99.34	99.50	99.33	98.76	98.92
mol200012080086	Non-Stopword Accuracy	97.10	98.61	96.46	95.87	92.13	92.42	Non-Stopword Accuracy	98.64	98.75	97.39	96.87	96.14	95.62

	Accuracy of Characters in Non-Stopwords	98.59	99.72	98.56	98.26	96.12	96.34	Accuracy of Characters in Non-Stopwords	99.79	99.72	98.81	98.93	97.41	97.61
mol200101250233	Non-Stopword Accuracy	97.54	96.44	95.91	94.61	92.50	91.86	Non-Stopword Accuracy	98.19	97.61	97.82	97.19	96.55	96.02
	Accuracy of Characters in Non-Stopwords	99.46	99.28	99.01	98.67	97.70	97.73	Accuracy of Characters in Non-Stopwords	99.57	99.47	99.48	99.31	99.28	99.11
mol200103160002	Non-Stopword Accuracy	98.98	97.96	97.18	95.79	95.11	94.79	Non-Stopword Accuracy	98.58	98.06	97.68	97.09	97.31	96.86
	Accuracy of Characters in Non-Stopwords	99.74	99.56	99.45	99.17	98.89	98.97	Accuracy of Characters in Non-Stopwords	99.56	99.47	99.42	99.26	99.32	99.27
mol200104160088	Non-Stopword Accuracy	91.14	89.94	89.07	86.80	85.06	85.63	Non-Stopword Accuracy	98.13	97.46	97.02	96.58	95.26	95.70
	Accuracy of Characters in Non-Stopwords	98.47	98.31	98.24	97.20	96.69	96.93	Accuracy of Characters in Non-Stopwords	99.67	99.40	99.43	99.36	99.11	98.98
Average of all 17 documents	**Non-Stopword Accuracy**	**97.44**	**97.03**	**96.45**	**95.05**	**92.78**	**91.91**	**Non-Stopword Accuracy**	**97.89**	**97.56**	**97.28**	**96.77**	**95.64**	**95.46**
	Accuracy of Characters in Non-Stopwords	**99.37**	**99.40**	**99.20**	**98.67**	**97.89**	**97.79**	**Accuracy of Characters in Non-Stopwords**	**99.65**	**99.58**	**99.46**	**99.22**	**98.93**	**98.80**

APPENDIX C.

Non-stopword accuracies and Unique non-stopword accuracies for Manicure output

Document Accession Number	ALL Non-Stopwords (N)	orig	gen0	gen1	gen2	gen3	gen4	Unique Non-Stopwords (M)	orig	gen0	gen1	gen2	gen3	gen4
mol199907200407	Non-Stopword Accuracy	99.00	98.52	97.19	96.62	95.87	92.83	Non-Stopword Accuracy	99.67	99.34	99.01	99.01	97.81	97.26
	Accuracy of Characters in Non-Stopwords	99.80	99.69	98.43	98.26	97.85	97.09	Accuracy of Characters in Non-Stopwords	99.94	99.85	99.05	99.05	98.79	98.49
mol199911010207	Non-Stopword Accuracy	98.68	98.36	98.03	97.80	96.44	92.57	Non-Stopword Accuracy	99.42	99.13	99.07	98.72	97.62	96.40
	Accuracy of Characters in Non-Stopwords	99.49	99.70	99.64	99.57	99.09	97.68	Accuracy of Characters in Non-Stopwords	99.84	99.80	99.84	99.57	99.35	99.23
mol200002170216	Non-Stopword Accuracy	98.49	97.47	97.07	94.66	91.71	91.27	Non-Stopword Accuracy	98.34	98.16	97.75	96.68	92.30	94.26
	Accuracy of Characters in Non-Stopwords	99.36	99.40	99.17	97.82	97.10	97.02	Accuracy of Characters in Non-Stopwords	99.59	99.60	99.31	99.01	97.73	98.46
mol200002280529	Non-Stopword Accuracy	99.37	99.34	99.01	98.19	96.61	97.29	Non-Stopword Accuracy	99.36	99.45	98.90	98.62	97.70	98.44
	Accuracy of Characters in Non-Stopwords	99.76	99.79	99.70	99.25	98.74	98.78	Accuracy of Characters in Non-Stopwords	99.84	99.86	99.67	99.70	99.18	99.47
mol200004130692	Non-Stopword Accuracy	98.13	97.67	97.64	95.07	94.43	94.51	Non-Stopword Accuracy	99.61	99.10	98.97	97.94	98.20	98.07
	Accuracy of Characters in Non-Stopwords	99.68	99.59	99.64	98.66	98.04	98.32	Accuracy of Characters in Non-Stopwords	99.91	99.83	99.79	98.77	99.59	98.75
mol200004140874	Non-Stopword Accuracy	98.64	98.02	98.04	96.90	95.69	95.72	Non-Stopword Accuracy	98.69	97.65	97.75	97.28	96.39	96.39
	Accuracy of Characters in Non-Stopwords	99.63	99.46	99.51	99.12	98.59	98.68	Accuracy of Characters in Non-Stopwords	99.46	99.25	99.27	99.12	98.83	98.84
mol200005230155	Non-Stopword Accuracy	98.56	97.38	97.58	96.62	95.39	95.50	Non-Stopword Accuracy	99.48	99.22	99.09	98.57	98.31	97.79

	Accuracy of Characters in Non-Stopwords	99.86	99.42	99.39	99.01	98.45	98.66	Accuracy of Characters in Non-Stopwords	99.89	99.84	99.83	99.82	99.69	99.62
mol200005250378	Non-Stopword Accuracy	98.88	98.20	98.25	97.54	96.52	95.50	Non-Stopword Accuracy	98.50	97.73	98.09	97.40	97.24	96.96
	Accuracy of Characters in Non-Stopwords	99.70	99.57	99.60	99.42	99.00	98.46	Accuracy of Characters in Non-Stopwords	99.53	99.32	99.37	99.22	99.24	99.10
mol200005260336	Non-Stopword Accuracy	99.03	98.13	98.12	96.47	94.52	94.30	Non-Stopword Accuracy	99.25	98.49	98.41	98.07	96.65	96.23
	Accuracy of Characters in Non-Stopwords	99.88	99.62	99.61	98.83	97.81	98.09	Accuracy of Characters in Non-Stopwords	99.82	99.74	99.64	99.32	98.76	98.83
mol200006090266	Non-Stopword Accuracy	98.69	98.81	98.73	98.34	95.97	95.34	Non-Stopword Accuracy	99.06	99.32	99.15	98.72	97.19	96.68
	Accuracy of Characters in Non-Stopwords	99.76	99.85	99.83	99.73	98.82	98.46	Accuracy of Characters in Non-Stopwords	99.79	99.85	99.81	99.70	99.10	98.23
mol200006270254	Non-Stopword Accuracy	98.64	98.33	98.77	97.81	96.95	96.71	Non-Stopword Accuracy	98.70	98.39	98.24	98.39	97.02	97.32
	Accuracy of Characters in Non-Stopwords	99.71	99.66	99.70	99.35	99.25	99.24	Accuracy of Characters in Non-Stopwords	99.63	99.58	99.55	99.55	99.37	99.37
mol200007250453	Non-Stopword Accuracy	96.94	96.70	96.57	95.12	93.75	93.61	Non-Stopword Accuracy	97.40	97.02	96.78	95.86	94.88	94.50
	Accuracy of Characters in Non-Stopwords	99.21	99.18	99.14	98.22	97.76	97.77	Accuracy of Characters in Non-Stopwords	99.28	99.21	99.15	98.73	98.54	98.33
mol200011220005	Non-Stopword Accuracy	97.39	96.60	96.77	96.17	94.25	93.99	Non-Stopword Accuracy	97.58	97.33	97.21	96.65	95.10	95.22
	Accuracy of Characters in Non-Stopwords	99.30	99.09	99.18	99.06	97.99	97.73	Accuracy of Characters in Non-Stopwords	99.31	99.11	99.27	99.20	98.60	98.74
mol200012080086	Non-Stopword Accuracy	97.66	98.96	96.99	96.60	93.30	93.74	Non-Stopword Accuracy	98.96	98.75	97.50	97.08	96.35	95.93

	Accuracy of Characters in Non-Stopwords	98.72	99.77	98.63	98.35	96.25	96.64	Accuracy of Characters in Non-Stopwords	99.87	99.73	98.82	98.97	97.48	97.75
mol200101250233	Non-Stopword Accuracy	98.09	97.28	96.81	95.80	94.22	94.13	Non-Stopword Accuracy	98.19	97.61	97.82	97.24	96.76	96.49
	Accuracy of Characters in Non-Stopwords	99.57	99.44	99.15	98.86	97.95	98.09	Accuracy of Characters in Non-Stopwords	99.54	99.43	99.45	99.29	99.21	99.17
mol200103160002	Non-Stopword Accuracy	98.89	97.99	97.64	96.74	96.54	96.52	Non-Stopword Accuracy	98.28	97.91	97.61	96.86	97.23	96.79
	Accuracy of Characters in Non-Stopwords	99.63	99.45	99.43	99.21	99.05	99.16	Accuracy of Characters in Non-Stopwords	99.35	99.32	99.15	99.10	99.20	99.13
mol200104160088	Non-Stopword Accuracy	91.16	90.41	89.62	88.02	86.79	86.90	Non-Stopword Accuracy	98.13	97.46	97.13	97.02	95.59	95.92
	Accuracy of Characters in Non-Stopwords	98.46	98.39	98.30	97.35	96.97	97.09	Accuracy of Characters in Non-Stopwords	99.61	99.40	99.45	99.44	99.21	99.06
Average of all 17 documents	**Non-Stopword Accuracy**	**98.01**	**97.54**	**97.23**	**96.15**	**94.64**	**94.14**	**Non-Stopword Accuracy**	**98.74**	**98.36**	**98.15**	**97.65**	**96.61**	**96.51**
	Accuracy of Characters in Non-Stopwords	**99.50**	**99.47**	**99.30**	**98.83**	**98.16**	**98.06**	**Accuracy of Characters in Non-Stopwords**	**99.66**	**99.57**	**99.44**	**99.27**	**98.93**	**98.86**