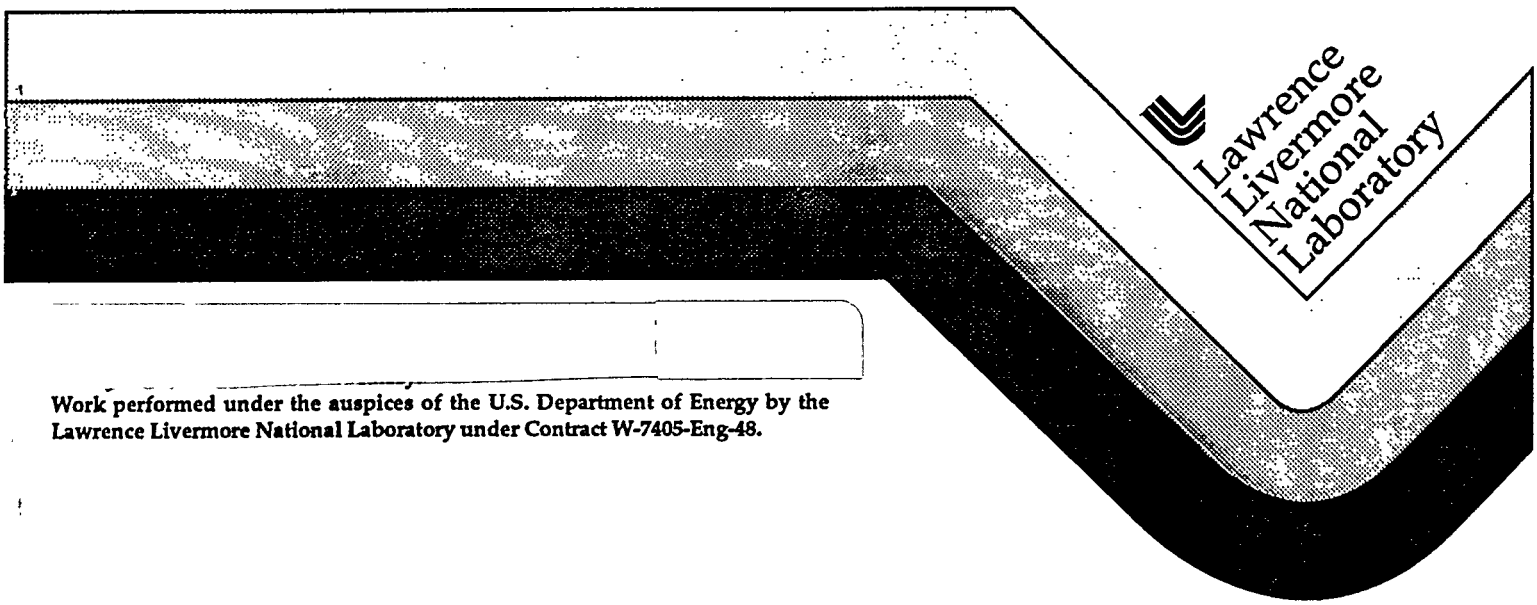
12/7-31-95 JSC (1)

UCRL-ID-121172

DOE Integrated Safeguards and Security (DISS) Historical Document Archival and Retrieval Analysis, Requirements and Recommendations

Howard B. Guyer
Craig A. McChesney

October 7, 1994



Work performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

DISCLAIMER

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

Table Of Contents

List of Figures ii

Section 1: Introduction 1
1.1 Purpose 1
1.2 Scope 1
1.3 Definitions, Acronyms and Abbreviations 1
1.4 References 5
1.5 Overview 6

Section 2: General Description 7
2.1 Product Perspective 7
2.2 Product Functions 10
2.3 User Characteristics 10
2.4 General Constraints 11

Section 3: Specific Requirements 13
3.1 Capture Module Functional Requirements 13
3.2 Retrieve Module Functional Requirements 19
3.3 Manage Module Functional Requirements 24

Section 4: Recommendations. 29
4.1 Backfile Conversion Options and Estimates 29
4.2 Scanner Requirements and Alternatives 32
4.3 Backfile Conversion and Capture Subsystem Recommendations 34
4.4 Document Imaging Software Alternatives 34
4.5 Database Schema 38
4.6 Long-Term Storage Subsystem: Alternatives and Recommendations 39
4.7 Document Imaging Peripherals: Alternatives and Recommendations 41
4.8 Possible Future Directions 41

Appendices:
A: Vault Analysis: DOE Oakland Operations Office 47

B: Vault Analysis: DOE Albuquerque Operations Office 57

C: Description of Document Types 69

List of Figures

Figure 2-1 9
(Conceptual Model: Historical Document Archival and Retrieval Subsystem)

Figure 4-1 43
(HDAR Architecture Using Image-enabling Approach)

Figure 4-2 44
(HDAR Architecture Using Imaging Application Development Toolkit Approach)

Figure 4-3 45
(HDAR High-level Document Repository Entity Relationship Model)

Figure VO-1 56
(DOE Oakland Operations Office Vault Layout)

Figure VA-1 67
(DOE Albuquerque Operations Office Vault Layout)

Section 1: Introduction

This section defines the scope of the Historical Document Archival and Retrieval sub-system (HDAR), describes the organization of this document and lists references to additional and related information and documentation.

1.1 Purpose

The purpose of this document is to provide software and hardware requirements specifications for the Historical Document Archival and Retrieval subsystem of the DOE Integrated Security Information System / Electronic Transfer system (DISS/ET).

The intended audience is primarily the DISS/ET project manager and appropriate team members.

1.2 Scope

The overall primary objective of HDAR is to create a repository of historical personnel security documents and provide the functionality needed for archival and retrieval use by other software modules and application users of the DISS/ET system. The software product to be produced from this specification is the Historical Document Archival and Retrieval Subsystem. The product will provide the functionality to capture, retrieve and manage documents currently contained in the personnel security folders in DOE Operations Offices' vaults at various locations across the United States.

The long-term plan for DISS/ET includes the requirement to allow for capture and storage of arbitrary, currently undefined, clearance-related documents that fall outside the scope of the "cradle-to-grave" electronic processing provided by DISS/ET. However, this requirement is not within the scope of the requirements specified in this document.

The anticipated benefits of this system are as follows:

- Decreased storage costs required for personnel security folders (PSF's);
- Improved accessibility: (speed, convenience, efficiency);
- Improved accountability and control;
- Eliminates need for physical handling (retrieving, refiling, etc.);
- Information within documents maintained at higher quality over longer period;
- Reduced or eliminated photocopy costs;
- Enables convenient electronic transmittal of historical documents;
- Contributes efficiency to overall security investigation process.

1.3 Definitions, Acronyms and Abbreviations

This section contains a list of terms used in this document and their definitions.

Adjudicative Security Investigation Support Tool

ASIST; The system module of the DISS/ET system that focuses on satisfying the requirements of the DOE analyst who is responsible for adjudicating the results of the clearance investigation.

annotate

The ability to attach notes or graphics to page images by typing them in, or with a lightpen, digitizing tablet or other device. This is useful, for example, for clarifying documents.

applet

A stand-alone application with limited functionality, such as an image viewer.

Applicant User Interface

AUI; The system module of the DISS/ET system that focuses on satisfying the requirements of the applicant who is the subject of a clearance investigation. This module will provide the functionality for the applicant to fill out and submit the QSP form electronically.

backfile conversion strategy

An alternative method of document acquisition for a document imaging system. For the HDAR system, a backfile conversion strategy must be selected to convert all paper documents held in the DOE Personnel Security vaults to electronic image form.

clearance investigation number

A number assigned to each person who has a clearance investigation conducted for them. The format of the number is aannnnnn (two alphabetic characters followed by six numeric characters).

Clearance Office User Interface

CUI; The system module of the DISS/ET system that focuses on satisfying the requirements of the Clearance Offices throughout the DOE complex relative to the clearance investigation process.

compression scheme

The process by which an image is compressed, or encoded. Image files tend to be much larger than other files, such as ASCII text. Compression schemes store information about the image so the file takes up less computer space. Frequently,

an image file is compressed to 1/10th of its size for storage, then decompressed for operations such as viewing, editing and printing. Two common and standard methods for compression are CCITT Group III and CCITT Group IV.

DOE Integrated Security System / Electronic Transfer

DISS/ET; The name of the overall project and system focused on automating the DOE clearance investigation process including interfaces with the applicant, Clearance Offices, DOE Operations Offices, Office of Personnel Management and DOE Headquarters, Washington, D.C.

DOE Operations Office User Interface

DUI; The system module of the DISS/ET system that focuses on satisfying the requirements of the DOE Operations Office environment relative to the clearance investigation process.

duplex

A scanner with the capability to scan both sides of a page in a single pass.

dynamic data exchange

DDE; Allows Microsoft Windows applications to communicate with one another. For example, a database application might request an imaging application to retrieve and display an image.

image-enable

Adding imaging functionality to an existing application,. For example, adding the ability to display invoices to an accounts payable system.

index

In an imaging system, indexing can mean assigning keys to documents for retrieval; the central repository of these keys for all documents in a system is also sometimes called its index.

interapplication communication facilities

Methods used to allow computer applications to communicate with one another. On the Microsoft Windows platform, common interapplication communication facilities include the Clipboard, DDE, and OLE.

Personnel Security File Folder

PSF; the name of the folder stored in a Personnel Security Vault which contains all information is filed for each individual who is the subject of a clearance investigation.

NFS

Network File System; Allows one computer to remotely mount disk resources physically connected to another computer and then access those resources as if they were directly attached to it.

NFS client

Network File System client is a software module that provides tools for using files physically located on media attached to remote machines.

object linking or embedding

OLE; One of Microsoft Windows interapplication communication facilities; objects from one application are placed inside an object from another application. For example, an image can be placed in a database record using OLE.

ODBC

Open Database Connectivity; a standard API or interface language to heterogeneous database systems. Provides a standard way to communicate with any database management system which is ODBC compliant.

optical character recognition

OCR; A technology that can analyze a bit-map of some text, determine what the characters in the bit-map are, and produce a text file containing those characters. Imaging systems that use full-text searching also use OCR hardware or software to convert page images to text, so that text can be searched.

resolution

The fineness or coarseness of a page image as it was digitized or as it is displayed; measured in dpi (dots per inch).

retrieval keys

A word, number, or other identifier assigned to a document, used to retrieve the document from storage later (for example, the document's title, author or document type). Typically, several keys are assigned to every stored document.

scanner speed ratings

Typically the rating describing the speed at which the scanner can read and digitize pages of information. The rating is usually stated in pages per minute (ppm).

SCEPTR

Security Clearance Electronic Processing Transfer and Recordkeeping System; The original name given to the DISS/ET project.

simplex

A scanner with the ability to scan a single side of a page per pass.

temporary holding queue

A storage area, typically magnetic disk, where scanned images are held until they are indexed and/or stored on longer-term storage media such as optical disk.

1.4 References

The following documents were reviewed or created during the analysis stage and used towards the creation of this document:

Adjudicative Security Investigation Support Tool DOE Integrated Security Systems / Electronic Transfer, Lychin Chang, August 8, 1994

Design Document, DOE Security Clearance Electronic Processing Transfer and Record keeping System (SCEPTR), dated March 11, 1994

Requirements Document Phase I, DOE Security Clearance Electronic Processing, Transfer and Record keeping System (SCEPTR), dated January 12, 1994

Service Agreement for Historical Document Archival and Retrieval Subsystem of DISS/ET, dated July 1, 1994

Site/User Survey Details, DOE Security Clearance Electronic Processing, Transfer and Record keeping System (SCEPTR), dated December 21, 1993

Standard Operating Procedures for the PSB Vault, DOE Albuquerque Operations Office, Personnel Security Branch (PSB), dated July 28, 1994

Standard Operational Procedures for the PSB Vault, DOE Oakland Operations Office, Personnel Security Branch (PSB), Version 2, dated July 20, 1992

Vault Analysis, DOE Albuquerque PSB Vault, H.B. Guyer/C. A. McChesney, July 27, 1994

Vault Analysis, DOE Oakland PSB Vault, H.B. Guyer/C. A. McChesney, August 26, 1994

1.5 Overview

This document is made up of the following sections:

- Section 1, Introduction, describes the purpose of the document, the objective and capabilities of the HDAR subsystem and provides a list of related documents. It also describes how this document is organized.
- Section 2, General Description, describes the general factors that affect the product and its requirements.
- Section 3, Specific Requirements, contains all the functional requirements to be considered for detailed software system design.
- Section 4, Recommendations, contains specific recommendations for implementation.
- Appendix A contains the results of the analysis for the DOE Oakland Operations Office Personnel Security Division vault.
- Appendix B contains the results of the analysis for the DOE Albuquerque Operations Office Personnel Security Division vault.
- Appendix C contains a description of the types of documents contained within Personnel Security File Folders.

Section 2: General Description

This section describes the HDAR subsystem functions and identifies user characteristics, general constraints, assumptions and dependencies. See Figure 2-1 for the Conceptual Model diagram.

2.1 Product Perspective

HDAR provides document image capture, retrieval and management functionality within the DISS/ET system. DISS/ET contains four major components which will require some or all of the functionality offered by HDAR:

- DISS/ET Database. Stores applicant data, contractor and DOE additions and notes, request and case status, scanned document and fingerprint images, secure E-mail messages of applicant's data, and database access audit trail.
- DOE Operations Office User Interface (DUI). Used by regional DOE Operations Offices to track, review, and process clearance cases, send investigation requests to OPM via secure E-Mail, and receive OPM investigation results via secure E-Mail.
- Adjudicative Security Investigation Support Tool (ASIST). Used by DOE analysts to determine clearance suitability and process clearance decisions.
- Clearance Office User Interface (CUI). Used by Contractor Clearance Offices to initiate, track, review, and process clearance requests, and receive DOE clearance decision notification.

The Applicant User Interface (AUI) is a major component of DISS/ET that will not directly require the functionality offered by HDAR and will not be discussed further within this document. The CUI component may require only printing capability. Applicants typically request a copy of their last QSP for use in completing the current QSP application. This request is often made to the Contractor Clearance Office. Consequently, either a change in this process is needed or the print function needs to be provided in the CUI component. For the purpose of this document, it is assumed that CUI will make use of the print function. The major beneficiaries of the HDAR functionality are the DUI and ASIST components. Also, HDAR will integrate with the DISS/ET Database.

The HDAR subsystem contains three primary modules: capture, retrieve and manage. They interface with the DISS/ET components as follows:

The capture module provides document image scanning, indexing, and storage functionality. The image retrieval indexes and supporting data will be stored in the DISS/ET database. The capture application could stand alone or be integrated into DISS/ET's DUI and CUI components.

The retrieve module provides image retrieval, viewing, and output functionality. The retrieve application interfaces with the DISS/ET database for query purposes and must be integrated into the DISS/ET DUI for use by the DOE Operations Offices.

The manage module provides management, control, and audit functionality for document images. It will interface with the DISS/ET database and be integrated with DUI, though only a subset of DUI users will have access to these functions.

Because users will access DUI and CUI from IBM PC's running Windows, the capture and retrieve modules must also support this platform. Users of both the capture and retrieve modules should be provided with a large-screen, high-resolution monitor; this is especially important for users of the capture module who must perform image-assisted data entry. Additionally, HDAR's capture module requires document image scanners and a long-term storage medium such as optical disk.

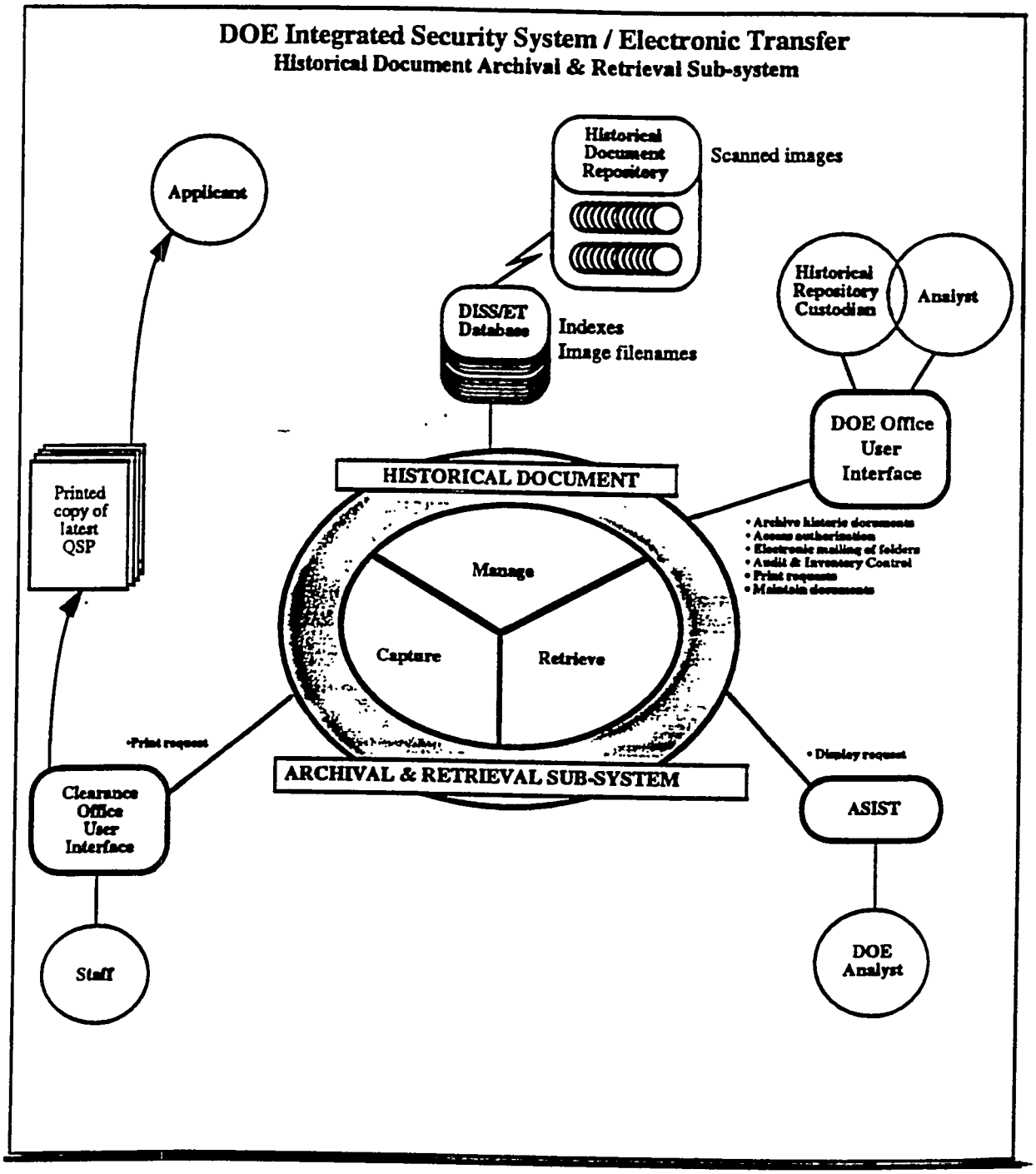


Figure 2-1

Conceptual Model: Historical Document Archival and Retrieval Subsystem

2.2 Product Functions

HDAR's functions are separated into three modules: capture, retrieve, and manage.

The capture module provides the following functions:

- scan: converts paper documents contained by personnel security folders in DOE Operations Offices' vaults to electronic form (document images);
- index: allocates retrieval keys provided by input operator to document images and stores these in DISS/ET Database;
- store: provides long-term storage for document images.

The retrieve module provides the following functions:

- browse: allows user to see index of document images associated with a given individual;
- view: allows user to view selected document images on screen;
- fax: allows user to transmit selected document images by facsimile;
- print: allows user to print selected document images;
- annotate: allows user to annotate a document image with an "electronic sticky note."

The manage module provides the following functions:

- modify: remove document images for a selected individual;
- control: provides access control management and audit trail reporting.

2.3 User Characteristics

Capture Module

Specific users of the capture module have not yet been identified. Generally, they will be staff within the DOE offices where Personnel Security Vaults are managed. These users must be able to locate retrieval keys within documents, operate the scanning equipment, and enter indexing data into the system. This role is crucial for effective retrieval of document images by other users.

Retrieve Module

Because this module will be integrated with the DUI, the user characteristics are the same as those described in the SCEPTR Requirements Document. These users generally use computers daily.

Manage Module

The user whose duties are analogous to the current vault custodian will be responsible for providing access authorization to document images for specific individuals.

2.4 General Constraints

Interfaces to other components.

The HDAR subsystem must be compatible and consistent with the following hardware and software environment used in the DISS/ET system:

DISS/ET Database

The DISS/ET Databases will be located at DOE Operations Offices. For example, the DOE Oakland Operations Office is on the LLNL site in Trailer 5980. Its database will be located there. The database machine is a SUN Sparc 10-40 running the Solaris 3.1 operating system. The relational database management system (RDBMS) is Oracle. The system is firewalled with a router, a bastion host, and a direct connected laser printer used for audit trail printing only.

DOE Operations Office User Interface (DUI)

The DUI will run on IBM PC or compatible with the Windows 3.1 operating system. These computers will have an associated laser printer. They will be behind the firewall with the DISS/ET DB computer. Macintosh implementation is not discussed in this report.

Adjudicative Security Investigation Support Tool (ASIST)

This subsystem is currently in the analysis stage. It is assumed that the requirements for ASIST will be consistent with the overall DISS/ET system and as a result, will be consistent with HDAR requirements. In general, the ASIST will run on IBM PC or compatible with the Windows 3.1 operating system and have access to a laser printer.

Clearance Office User Interface (CUI)

The CUI will run on IBM PC or compatible with the Windows 3.1 operating system. These computers will have an associated laser printer available for printing.

Regulations

DOE Order 5631.2C, "Personnel Security Program," prescribes the requirements for PSF's. Wording of this document and standard operating procedures may need to be modified to reflect the electronic capture, storage and control of historical personnel security folders' contents. DOE Orders 5631.2C and 1324.2A are examples of orders which currently articulate regulations for manual, paper-oriented systems. The electronic nature of DISS/ET requires, in some cases, different process descriptions to accomplish similar ends. Access control and auditing are typical of functions accomplished with different processes.

Audit and control

Audit and control requirements are detailed in Section 3.3, Manage Module Functional Requirements. The audit and control characteristics of the new HDAR must minimally meet the current level of audit and control required by the Standard Operating Procedures in effect.

Criticality

Timely and convenient access to historical personnel security folders is essential to the security investigation process. Applicants often request copies of previous QSP documents to assist them in completing current QSP forms. OPM investigators require access to Personnel Security File Folders during the investigation process. Analysts require access to entire Personnel Security File Folders during the adjudication process. The two DOE assurance processes, PSAP and PAP, require access to large numbers of Personnel Security File Folders. All of these process steps will be severely affected if access is not convenient and timely.

Section 3. Specific Requirements

For the purposes of requirements definition, HDAR is subdivided into three modules: capture, retrieve, and manage. The functional requirements for each module are discussed in Section 3.

3.1 Capture Module Functional Requirements

The capture module is subdivided into three primary functions: scan, index and store. Section 3.1.1 discusses requirements for the scan function and Section 3.1.2 discusses requirements for both the index and store functions. Section 3.1.3 contains general requirements for the capture module.

3.1.1 Capture Module: Scan Function Requirements

3.1.1.1 Introduction

The scan function converts paper documents contained within personnel security folders in DOE Operations Offices' vaults to electronic form (i.e., document images).

3.1.1.2 Inputs

Inputs by the user to the scan function include:

- selection of temporary holding queue for document images produced by scan function;
- provide a base name used to identify scanned documents until permanently indexed in the database;
- adjustment to scanner controls and options;
- selection of desired level of image quality assurance (QA) interaction;
- initiation of the scan operation;
- paper documents taken from a specific individual's personnel security folder.

Appendix C, Description of Document Types, provides details about the types of documents commonly found in personnel security folders.

In addition to providing the user input discussed above, the scanner operator must also feed pages into the scanner.

3.1.1.3 Processing

Processing tasks of the scan function include:

- verify existence and check validity of temporary holding queue and base name;
- validate scanner control settings;
- transmit scanner control settings to scanner;
- initiate physical scan process;
- display each image on screen for QA interaction with user;
- stop physical scan process;
- notify user of errors communicating with scanner;
- write document image files to temporary holding queue;
- allow user to view document images contained by temporary holding queue;
- allow user to add or remove images from the temporary holding queue.

3.1.1.4 Outputs

Outputs of the scan function include the temporary holding queue containing an image for each page inserted by the user to the scanner. The user-specified base name is used to name the document image files placed in the queue. Error messages are produced in response to invalid user input, scanner errors, and file creation errors.

3.1.2 Capture Module: Index and Store Function Requirements

3.1.2.1 Introduction

The index function assigns user-provided index and retrieval keys to the document images within a temporary holding queue and stores the keys in the DISS/ET database. The store function moves document images from the temporary holding queue to a long-term storage medium.

3.1.2.2 Inputs

- The user enters a key to link the images within a specific holding queue to the DISS/ET database record for the correct individual. The key used most often is the individual's social security number. However, other keys currently used are clearance investigation number, current last name and previous names.
- For each image within the queue, the user enters keys sufficient to uniquely describe the image (i.e., the document description, document type, document date, etc.).
- The user initiates a request to save the record and provides confirmation of same.

3.1.2.3 Processing

- Display the individual's name and other data from DISS/ET database to confirm that the user-specified key (e.g., social security number) is correct.
- Display images from the temporary holding queue for image-assisted data entry.
- Perform validity check of image descriptive keys against yet-to-be-established rules for naming documents.
- Commit data to DISS/ET database.
- Establish location index within the file system of long-term storage medium for each image file.
- Move the image files from the temporary holding queue to the long-term storage medium.
- Report errors to user if:
  - social security number not contained in DISS/ET database;
  - invalid image descriptive key provided by user;
  - communication, file transfer or database error.

3.1.2.4 Outputs

- The only output of the index function is whatever SQL commands are necessary to commit the user provided keys to the database. The expected and desired result is that the keys are correctly stored in the database.

- The store function:
  - generates indexes to the file system of the long-term storage medium for each image;
  - issues the necessary SQL commands to store the indexes;
  - creates the image files on the long-term storage medium.

3.1.3 Capture Module: General Requirements

3.1.3.1 Performance Requirements

- Static numerical requirements
  - Document image file size averages 50KB and varies widely depending on compression scheme, document type, and document quality.
- Dynamic numerical requirements
  - Based on industry experience, elapsed time for the scan function will be driven by user document preparation time, not by the scanner speed rating as might be expected. The daily workload of the scan, index, and store functions will depend on the selected backfile conversion strategy.

    Section 4, Recommendations, provides a detailed discussion of backfile conversion strategies, scanner speed ratings, and the relationship between them.

    Case Example: (this is not an estimate of the actual workload)

    If 10 folders, containing a total of 750 pages, are scanned on a given day, the expected workload for each capture module function is as follows:

    - the scan operator(s) will feed 750 pages into the scanner;
    - the scan function will create 750 image files within 10 temporary holding queues, requiring approximately 38.4 M of disk space (NOTE: this space can be freed after files are transferred to long-term storage);
    - the index operator(s) will enter keys for 750 images;
    - at least 750 records will be created within the database;

- 750 files will be transferred from temporary to long-term storage medium, where approximately 38.4 M of disk space will be required.

3.1.3.2 Design Constraints

- Standards compliance
  - The capture module user interface (supporting scan and index functions) must comply and be consistent with user interface guidelines for DISS/ET;
  - The temporary holding queue and base file naming conventions employed by the scan function must be consistent with yet-to-be established office procedures for scanning;
  - The database schema supporting the index and store functions should be consistent with the overall DISS/ET database schema; **Section 4, Recommendations**, discusses this issue in detail.
  - Because the index function updates the DISS/ET database, an audit trail scheme consistent with that employed by DISS/ET shall be maintained.
- Security:
  - The roles of scan and index operator are analogous to the role played by the individuals that file and retrieve documents within the vault, with the exception that the operator(s) must view every document within a personnel security folder in order to fully process it;
  - The capture module will employ the same login and security scheme used by DUI.

3.1.3.3 External Interfaces

User interface

- The scan function user interface, in a style consistent with DISS/ET, must provide controls to support:
  - selection of a temporary holding queue;
  - specification of a base file name;
  - access to scanner-specific controls (such as source, contrast, brightness, etc.);

- selection of desired QA level (pause on each page, display each page, no QA);
- initiation of scan process;
- canceling the scanning process;
- display images on screen for QA;
- provide QA controls (accept and stop, accept and continue, reject and stop, reject and continue);
- addition and removal of pages from holding queue;
- error reporting.

- The index function user interface, in a style consistent with DISS/ET, must provide controls to support:

  - display of temporary holding queues and their contents;
  - entry of database key to select appropriate individual's record from the database;
  - display data about individual from database for verification;
  - display images from temporary holding queues;
  - entry of database keys to uniquely identify each image;
  - initiation of request to save record with confirmation;
  - feedback on successful creation of database record and file transfer;
  - error reporting.

Hardware interfaces

- The scan function must interface with a scanner connected to the PC scan workstation. **Section 4, Recommendations**, discusses potential scanner interface hardware and software, as well as actual scanning hardware. There are two common modes of communication between a PC and scanner: SCSI or proprietary interface hardware

- The temporary holding queue employed by the scan function should reside on magnetic disk, either locally or mounted remotely via NFS.

- The store function requires the support of a long-term storage medium. **Section 4, Recommendations,** discusses alternatives for long-term storage equipment and potential architectures.

Software interfaces

- The capture module will interface with one or more software packages that support imaging functions such as scanning, compression, file creation, and display. **Section 4, Recommendations**, discusses this issue in more detail.

- The index and store functions will interface with the DISS/ET database for storage of image retrieval keys.

- Depending on the selected storage architecture, the store function may interface with a driver for the long-term storage medium.

Communication interfaces

- An NFS client might be used by the capture module to provide the ability to mount remote disks for the temporary holding queue or long-term storage. There are several commercially available products that provide this functionality.

- The index and store functions will require the support of the ODBC and database client components discussed in the SCEPTR Design Document for database communication.

3.1.3.4 Other Requirements

Database

- Although the temporary holding queue employed by the scan function is ephemeral, it may be desirable to store a queue index within the DISS/ET database for improved control and recovery.

- The index and store functions require database storage within the DISS/ET database for image retrieval keys and indexes.

- After an individual's personnel security folder is scanned and indexed, the database will contain a set of images associated with that individual.

- Personnel security folders must be maintained for all active clearances and must be kept for ten years after a clearance is terminated. The images and their indexes, therefore, must be maintained on-line or near-line during this time period.

3.2 Retrieve Module Functional Requirements

The retrieve module is divided into two primary functions for the purposes of requirements definition: browse and view. Secondary functions that support browse and view include: fax, print, and annotate. The functional requirements for browse and scan are discussed in Sections 3.2.1 and 3.2.2, respectively. Requirements for the secondary functions are discussed in both sections as appropriate. General requirements for the retrieve module are discussed in Section 3.2.3.

3.2.1 Retrieve Module: Browse Function Requirements

3.2.1.1 Introduction

The browse function allows the user to see a listing of the document images associated with a given individual, get information about specific images, print the entire listing of images, print selected images, and fax selected images.

3.2.1.2 Inputs

Inputs to the browse function include:

By DUI -
- specification of a key that identifies an individual's record within the database.

By the user -
- initiation of request to print listing of images;
- selection of image(s) from the listing;
- initiation of request to get information about the selected image;
- initiation of request to print selected image(s);
- initiation of request to fax the selected image(s).

3.2.1.3 Processing

Processing tasks of the browse function include:

- query database for images associated with specified individual;
- display listing of available images;
- if requested by user, print image listing;
- if requested by user, query database for more information about a specific image and display it;
- if requested by user, print selected image(s);
- if requested by user, fax selected image(s);
- report errors to user if database, file transfer, print, fax, or communication errors encountered.

3.2.1.4 Outputs

Outputs of the browse function include:

- on-screen display of the image listing;
- capability to produce hard copy of the image listing;
- capability to produce hard copy of selected image(s);
- capability to transmit selected image(s) by facsimile.

3.2.2 Retrieve Module: View Function Requirements

3.2.2.1 Introduction

The view function allows users to view images selected from the listing produced by the browse function. The view function user can also print, fax, or annotate the image.

3.2.2.2 Inputs

Inputs to the view function include:

By the browse function -
- specification of the image to display.

By the user -
- initiation of request to zoom the image;
- initiation of request to go to next or previous image;
- initiation of request to print the image;
- initiation of request to transmit the image by facsimile;
- initiation of request to annotate an image.

3.2.2.3 Processing

Processing tasks of the view function include:

- query database for long-term storage location index for specified image;
- transfer image from long-term storage to local cache;
- display the image;
- if requested by user, zoom image as specified;

- if requested by user, display the next or previous image from the image listing;
- if requested by user, print the image;
- if requested by user, fax the image;
- report errors if database, file transfer, print, fax, or communication errors encountered.

3.2.2.4 Outputs

Outputs of the view function include:

- on screen display of specified image;
- capability to produce hard copy of the image;
- capability to transmit the image by facsimile.

3.2.3 Retrieve Module: General Requirements

3.2.3.1 Performance Requirements

No specific performance requirements are identified for the retrieve module. This module is accessed from DUI. Consult the **SCEPTR Requirements Document** for relevant performance requirements.

3.2.3.2 Design Constraints

- Standards Compliance
  - Because it is accessed from DUI, the retrieve module user interface must comply with stylistic guidelines established for DUI;
  - The DUI audit trail mechanism should be adequate for the retrieve module. Audit tracing for the print, fax, and view functions must also be provided.
- Security
  - Only those DUI users specifically provided access to the image files associated with a given individual via the appropriate user of the manage module will be able to access retrieve module functions;
  - The DUI login will be used to establish the user's suitability to access retrieve module functions;

- Retrieve module output may only be sent to authorized print or fax devices.

3.2.3.3 External Interfaces

User Interface

- The browse function user interface, in a style consistent with DISS/ET DUI module, must provide controls to support:
  - display of image listing associated with the specified individual;
  - initiation of request to print the image listing;
  - selection of image(s) from the listing;
  - display information about selected image;
  - initiation of request to print selected image(s);
  - set print and fax controls;
  - initiation of request to fax selected image(s);
  - feedback on status of print and fax requests;
  - error reporting.

- The view function user interface, in a style consistent with DISS/ET DUI module, must provide controls to support:
  - display of specified image;
  - initiation of requests to zoom, go to next or previous, print, fax, or annotate the image;
  - set print and fax controls;
  - feedback on status of print and fax requests;
  - error reporting.

Hardware Interfaces

- The retrieve module requires access to a local or networked printer meeting the security requirements stated above;

- The retrieve module requires access to a local or networked PC to facsimile gateway, fax server, or fax modem meeting the security requirements stated above;

- The retrieve module requires access to the long-term storage medium supporting the capture module's store function.

Software Interfaces

- DUI will provide the browse function a specification of the database key associated with a given individual's database record (the record currently displayed within DUI);

- The browse function will provide the view function a specification of the image to be displayed;

- Depending on the selected storage architecture, the retrieve module may interface with a driver for the long-term storage medium;
- The retrieve module will interface with one or more software packages that support imaging functions such as display, print, and fax transmission;
- The fax function will interface with a software fax driver.

Communication Interfaces

- An NFS client might be used by the retrieve module to provide the ability to remotely mount the long-term storage medium;
- Like DUI, the browse and retrieve functions will require the support of ODBC and database client components for database communication.

3.2.3.4 Other Requirements

Database

- The retrieve module will utilize the database schema established for the capture module's index and store functions. The schema developed to support the capture module must be extended to include annotation support;
- The retrieve module will utilize the database schema established for DISS/ET audit tracing.

3.3 Manage Module Functional Requirements

The manage module is comprised of two primary functions: modify and control. Sections 3.3.1 and 3.3.2 detail the requirements for modify and control, respectively. Section 3.3.3 provides general requirements for the manage module.

3.3.1 Manage Module: Modify Function

3.3.1.1 Introduction

The modify function allows the user to remove selected document images from the database for a given individual. The modify function is simply an extension to the browse function discussed in Section 3.2.1. Instead of repeating those requirements here, this section only discusses added requirements.

3.3.1.2 Inputs

Inputs to the modify function are:

- initiation (by user) of request to delete images selected in the image listing and confirmation of same.

3.3.1.3 Processing

Processing tasks of the modify function include:

- delete the selected document image from the associated individual's database record;
- add entry to audit tracing facility;
- provide feedback on success or failure of operation;
- report any database, communication, or printing errors to user.

3.3.1.4 Outputs

There are no outputs of the modify function other than on-screen feedback, error reporting and the audit trace mechanism.

3.3.2 Manage Module: Control Function

3.3.2.1 Introduction

The control function provides access control management for document images and audit trail reporting. Document access will be controlled using a mechanism like access control lists. By default, only manage module users have access to document images. Manage module users specifically assign retrieve module users access to the document images for a given individual. User access to a given individual's record can be removed when no longer needed. Like the modify function, the control function is simply an extension to the retrieve module's browse function.

3.3.2.2 Inputs

Inputs to the control function include:

By the user -

- initiation of request to view the access list for the images associated with the currently selected individual;
- initiation of request to add access for a specified user;
- the name of the user to which access should be provided;
- initiation of request to remove access for a specified user;

- specification of the name of the user for which access should be removed;
- initiation of request to display or print audit trail for images associated with selected individual's database record;
- initiation of request to display or print audit trail for the specified user's access to the images associated with the selected individual's database record.

3.3.2.3 Processing

Processing tasks of the control function include:

- if requested by user, query database for the access list associated with the selected individual;
- if requested by user, add a user to the access list;
- if requested by user, remove a user from the access list;
- if requested by user, display or print the audit trail for the selected individual's database record;
- if requested by user, display or print the audit trail for the specified user's access to the individual's database record;
- log modifications to access list using audit trace mechanism;
- report database, communication, or printing errors to the user.

3.3.2.4 Outputs

Outputs of the control function include:

- capability to display the access list for the selected individual's database record;
- capability to display or print the audit trail for the selected individual;
- capability to display or print the audit trail for a specified user.

3.3.3 Manage Module: General Requirements

3.3.3.1 Performance Requirements.

No specific performance requirements are identified for the manage module. This module is accessed from DUI and, as such, consult the

SCEPTR Requirements Document for relevant performance requirements.

3.3.3.2 Design Constraints.

- Standards Compliance
  - The manage module user interface must comply with stylistic guidelines established for DUI;
  - The DUI audit trail mechanism will be utilized by the manage module.
- Security
  - Manage module users are privileged DUI users. General DUI users will not have access to manage module functions;
  - The DUI login will be used to establish the user's suitability to access manage module functions.

3.3.3.3 External Interfaces.

User Interface

- The modify function extends the retrieve module browse function's user interface with the following support:
  - initiation of request to delete the image selected in the image listing;
  - confirmation of delete operation;
  - feedback on success or failure of delete operation;
  - error reporting.
- The control function extends the retrieve module browse function's user interface with the following support:
  - initiation of request to view access list for selected individual;
  - on-screen display of access list;
  - initiation of request to add a specified user to access list;
  - initiation of request to remove a specified user from access list with confirmation of same;
  - initiation of request to display or print audit trail for selected individual;
  - initiation of request to display or print audit trail for specified user;
  - set print controls;
  - on-screen display of audit trail;
  - feedback on modifications made to access list;
  - error reporting.

Hardware Interfaces

- No specialized hardware is required to support the manage module.
- The DUI audit trail printer will be utilized.

Software Interfaces

- The manage module does not communicate with other software modules; it is an extension of the retrieve module's browse function.

Communication Interfaces

- Like DUI, the manage module requires the support of ODBC and database client components for database communication.

3.3.3.4 Other Requirements.

Database

- The manage module extends the database requirements for the retrieve module's browse function by adding an access control list schema.

Section 4: Recommendations

Section 4 contains eight parts. The first part focuses on backfile conversion options and estimates which discusses the task of converting paper documents held in vaults to long-term electronic storage media. The remaining parts of this section include descriptions and discussions of scanner requirements and alternatives, backfile conversion and capture subsystem recommendations, document imaging software alternatives, database schema, long-term storage subsystem alternatives and recommendations, document imaging peripherals alternatives and recommendations, and possible future directions.

4.1 Backfile Conversion Options and Estimates

Section 4.1 contains options and estimates to accomplish the backfile conversion process. Specific hardware selections and quantities, the number of people assigned and the amount time spent on the scanning task are factors which will effect the duration required for scanning all historical documents into the system.

4.1.1 Backfile Conversion Process Options:

The following three options are offered with the intent of providing each DOE Operations Office the opportunity to choose an option that best suits their circumstances. This information can be used to determine what resources to commit to the backfile conversion process. Each option is described, assumptions are stated, and estimated effort is derived. When all active historical PSF's are accessible in electronic form, the backfile conversion process will be complete.

Option A: Scan all active PSF's held in the vault.

This would involve scanning all of the active PSF's in the vault without regard to the order in which the they were scanned.

Option B: Scan active PSF's as they are requested for check-out or check-in.

This option would involve scanning PSF contents upon request for check-out or check-in. The choice of "in or out" scanning would be driven primarily by the timeliness required for access to a requested PSF and/or the amount of advance notice given for check out. The relative desire by users for electronic form vs. paper documents would also be a consideration for this option.

Option C: Scan active PSF's scheduled for upcoming fiscal year reinvestigation.

This option would be driven by the number of reinvestigations scheduled and would focus on those specific PSF's. The project would have intermediate milestone completions based on fiscal year requirements.

4.1.2 Backfile Conversion Option Estimates

Certain assumptions need to be made when deriving backfile conversion estimates. Some of these assumptions apply to all options and are stated below. Assumptions specific to a vault or option are also identified within the specific option narrative. Figure 4-1 contains the estimates for all three options.

Assumptions for all options:

- 250 work days per year; 8 work hours per day;
- Each PSF consists of multiple documents;
- A document is defined as one or more related pages;
- All pages are single-sided;
- All documents in the PSF will be scanned;
- Each document will require indexing;
- Each page will require an average of 30 seconds for scanning;

  (The 30 second estimate for scanning includes all the time necessary for retrieving the PSF from the vault, preparing the PSF for page scanning, reviewing and identifying indexes to be entered, entering the indexes for the document, replacing the scanned documents to the PSF, and returning the PSF to the appropriate vault shelf.)

  This is a key assumption and has the highest likelihood for variation.

- Average number of pages scanned per operator per day is 960;
  (28,800 seconds per day / 30 seconds per page)
- Assigned person(s) is/are dedicated to the task of backfile conversion
- Minimum scan station configuration:
  (See **Section 4.3, Backfile Conversion and Capture Subsystem Recommendations.**)

DOE Oakland Operations Office *Vault assumption:* *- Average PSF is 60 pages*	DOE Albuquerque Operations Office *Vault assumption:* *- Average PSF is 76 pages*
Option A: Scan all active PSFs held in the vault. *Assumptions:* *- Total documents to be scanned is 13,200* *- Total pages to be scanned is 792,000* Estimated effort: 3.3 person years	**Option A:** Scan all active PSFs held in the vault. *Assumptions:* *- Total documents to be scanned is 39,500* *- Total pages to be scanned is 2,962,500* Estimated effort: 12.3 person years
Option B: Scan active PSF's as they are requested for check-out or check-in. *Assumptions:* *- Average daily retrieval requests equals 50 PSF's;* *- Based on the average daily page volume of 3,000 pages, three scanning workstations would be required for this option.* Estimated daily effort: 3.1 person days	**Option B:** Scan active PSF's as they are requested for check-out or check-in. *Assumptions:* *- Average daily retrieval requests 200 PSF's* *- Based on the average daily page volume of 15,200 pages, sixteen scanning workstations would be required for this option.* Estimated daily effort: 15.8 person days
Option C: Scan active PSF's scheduled for upcoming fiscal year reinvestigation. *Assumptions:* *- FY 95 page volume: 107,460 pages* *- FY 96 page volume: 103,500 pages* *- FY 97 page volume: 98,700 pages* FY 95 Estimated Effort: 112 person days (.4 person years) FY 96 Estimated Effort: 108 person days (.4 person years) FY 97 Estimated Effort: 103 person days (.4 person years)	**Option C:** Scan active PSF's scheduled for upcoming fiscal year reinvestigation. *Assumptions:* *- FY 95 page volume: 227,587 pages;* *- FY 96 page volume: 834,487 pages* *- FY 97 page volume: 568,912 pages* FY 95 Estimated Effort: 237 person days (1 person year) FY 96 Estimated Effort: 869 person days (3.5 person years) FY 97 Estimated Effort: 593 person days (2.4 person years)

Figure 4-1

Backfile Conversion Option Estimates
for DOE Oakland and DOE Albuquerque Operations Offices

4.2 Scanner Requirements and Alternatives

The survey of common PSF contents conducted during HDAR analysis and documented in **Appendix C, Description of Document Types,** identified the following basic scanner capability requirements:

Document Types

Though a majority of documents found in the PSF's are letter and legal size, other non-standard types of documents were encountered such as envelopes, finger prints (8" x 8"), and 2 page open-out forms (total size 17" x 22").

Document Feeding

Because of the variety of document types and conditions encountered within PSF's, two types of document feeders are necessary. Flatbed capability allows documents to be placed directly on a glass panel for scanning (much like a paper copier's basic operation). This is appropriate for the capture of books, bound reports, fragile or otherwise unusual originals. An automatic document feeder (ADF) allows a stack of 50 to 500 originals, depending on the scanner, to be scanned in a rapid, continuous mode.

Resolution

Even very high-resolution monitors support a maximum resolution of 200 dots-per-inch (dpi); most monitors are considerably lower resolution. Therefore, 200 dpi is probably adequate resolution for this project. However, even though 300 dpi resolution requires more storage space per image file, this higher resolution is required for extended image processing applications such as optical character recognition (OCR), bar-code indexing, and forms processing and might be a choice for the longer term.

The backfile conversion estimates discussed in Section 4.1, result in an additional scanner capability requirement:

Speed

The backfile conversion estimate of 960 pages per day, if literally translated into a scanner requirement, requires a scanner that can scan 2 pages per minute (ppm). The scan operator will likely process folders in batches, however, meaning that higher throughput is actually necessary. A scanner capable of at least 25 ppm should satisfy the backfile conversion requirements, as document preparation and feeding will be tasks the scan operator performs.

Other requirements resulting from general DISS/ET and HDAR requirements include:

Platform Support

Because the PC workstation of choice in the DOE Operations Offices is the IBM PC or compatible, the scanner must support this platform.

Interface

Two common approaches are used to interface with the scanner, a combination of RS232C for control and proprietary local for video signal, or SCSI-II for both control and signal. Both approaches are compatible with the IBM PC or compatible platform. Selection of the appropriate interface should be driven by the interfaces supported by the chosen document imaging implementation. This issue is discussed in more detail in **Section 4.5, Document Imaging Software Alternatives.**

Most scanner alternatives meeting these requirements fall into one of two categories, the high-end or low-end. These categories are based on the primary features of speed, durability and price.

High-End Scanners

Product entries in this category are capable of scanning at over 40 ppm and are rugged industrial strength scanners. Prices range from $12,000 to $30,000. The AIS Imaging Team supports three high-end Bell and Howell scanners in other imaging applications. They run at 42 ppm are available in two models: the 3338 is a simplex scanner that costs roughly $17,000; the 6338 is a duplex scanner that costs roughly $25,000. One limitation of the Bell and Howell and most scanners in this category, is that documents must be fed using the automatic document feeder-- they do not have flatbed capability. Other manufacturers with entries in this category are Kodak, Ricoh, DC, and Vision Shape.

Low-End Scanners

Most product entries in this category are capable of scanning at under 40 ppm and are intended for lower daily scanning volumes (i.e., under 2,000 pages per day). Prices for these products range from $3,500 to $11,000. The AIS Imaging Team presently supports one model in this category, the Fujitsu M3097E. It is rated at 39 ppm, has a 100 page ADF, a flatbed, and costs $11,000. Other manufacturers with entries in this category include Bell and Howell, Ideal, Microseal, Panasonic, and Ricoh.

4.3 Backfile Conversion and Capture Subsystem Recommendations

Backfile conversion options B and C are both common approaches to document acquisition for projects with a large backfile. The decision on which approach is most appropriate will likely be based on a number of factors, such as the resources available (people and dollars) to be applied to backfile conversion and the organization's relative desire to realize the potential benefits of electronic document management. (**Section 3, Specific Requirements,** describes these benefits.)

To implement backfile conversion option B, three dedicated scan operators are required for DOE Oakland and sixteen for DOE Albuquerque. Because of economies of scale, the number of scan operators needed does not translate directly into the number of scanning workstations required. By separating the tasks of retrieving folders, preparing documents, identifying pages for indexing, scanning, and indexing the documents, fewer scanning stations would likely be needed. Two scanning stations for Oakland and eight to ten stations for Albuquerque is a conservative estimate.

To implement backfile conversion option C, one scan station and one operator will suffice for either Oakland or Albuquerque for FY95. (See **Section 4.1, Backfile Conversion Options and Estimates,** for estimated levels in out-years.)

For either option B or C, it appears that scanners from the low-end category are probably most appropriate. The high-end scanners are very good for applications that have a document source with consistent size and quality, however, the low-end scanners handle unusual document types more flexibly and have an adequate speed rating for this application.

Section 4.8, Possible Future Directions, discusses bar code document identification, a technique that can considerably increase the overall throughput of document conversion.

4.4 Document Imaging Software Alternatives

There are alternative methods of implementing the imaging functionality described in **Section 3, Specific Requirements,** each of which falls into one of two general categories: commercially available image-enabling products or imaging application development toolkits. This section provides an architectural overview as well as the pros and cons for each approach.

4.4.1 Commercially Available Image-enabling Products

This method utilizes Windows multitasking and interapplication communication facilities, dynamic data exchange (DDE) and object linking or embedding (OLE) to image-enable existing Windows applications. Using this method, an existing application, typically a database, uses DDE or OLE to pass index information to the imaging system (an applet), which retrieves and displays the image. Figure 4-1 shows a high-level architectural model of HDAR using the image-enabling approach.

Under the image-enabling approach, a minor modification to the existing application, namely the addition of a field to hold the image index, is necessary. Likewise, a communication script must be developed for both the existing application and the imaging system.

Pros of the image-enabling approach include:

Development effort
- Minor modification to existing application is required, however, the existing application must have some means for locating and indexing documents at the page level;
- Most vendors provide all imaging functionality in a single package (scanning, storage, viewing, printing, etc.).

Cons of the image-enabling approach include:

Flexibility
- Little or no control over interface style and layout;
- Security is typically implemented within the imaging system, if at all, and is generally limited (i.e., members are assigned to scan, print, or display groups);
- Most vendors support a limited number of network protocols and optical disk interfaces;
- The imaging system is an independent application, loosely integrated with the original application.

Data Storage
- Many image-enabling products use proprietary databases for index and image storage. What if it is decided to change imaging systems? How can data be exported? What if you need to see the imaging data outside the imaging application?

Product Direction.
- A vendor has complete control over the direction for its product.

The following is a description of some image-enabling products and their approximate costs. Costs are provided for budgetary purposes only and should be verified before making any purchase decisions.

PaperClip Imaging Software PaperClip.

Uses folder concept to index images associated with a record in the application's database. Indexes images from ASCII text file created by an external application. $4,995 for five concurrent users.

Soda Creek SCT*Image

Uses DDE method of image-enabling. Includes scanning support. Database independent. Cost is $995 per user.

Genesys Image Extender

Uses DDE method of image-enabling. Requires essentially no programming. Requires OS/2 server with proprietary database. Cost is approximately $16,000 for ten simultaneous users, which includes scan station support.

Optical Technology Group Application Extender

Uses DDE method of image-enabling. ODBC compliant. Criticized for poor user interface. Uses query by example (QBE) form to enter query. Extensive document-level security. Server licenses run from 5 to 300 clients and cost from $8,995 to $59,995.

The AIS Imaging Team is not aware of any installations of these products, or products like them, at LLNL. From the information available to us, PaperClip seems to be the closest fit with HDAR requirements and should be evaluated in more depth if the image-enabling approach is preferred.

4.4.2 Imaging Application Development Toolkits

This method brings imaging functionality to new or existing applications through an application programming interface (API). The application is programmed to call the toolkit using the API; the toolkit provides imaging services directly to and under the control of the application. Figure 4-2 shows a high-level architectural model of HDAR using the application development toolkit approach.

Under the application development toolkit approach, the developer creates a graphical user interface (GUI) and data structures to hold image data using a traditional programming language such as C, C++, SmallTalk or Visual Basic. The GUI is then connected to the imaging toolkit by making calls to its API.

Pros of the application development toolkit approach include:

Flexibility
- The developer has complete control over interface look and feel;
- Data can be stored using the most appropriate method for the application;
- Imaging code can be reused in other places, including other modules and other applications;
- Security can be implemented at the appropriate level for the application.

Cons of the application development toolkit approach include:

Development Effort
- The imaging application is completely developed from scratch.

Imaging application development toolkit products are typically licensed as two components: the developer's toolkit and run-time licenses for distribution. Some products and their approximate costs:

TMS View Director

C library that provides software image compression and decompression for a variety of industry standards and a broad range of image viewing and printing functions. Provides an image manager to remove much of the data manipulation burden from the developer. Includes sample user interface with source code. Delivered as a DLL with an import library. Supports Windows, DOS, Sun SPARC, and Macintosh. Cost is $5,000 which includes developer tools and 20 run-time licenses.

Pixel Translations PixTools/File & PixTools/Scan

C library that allows application programs to work with many image file formats and scanners using a single, device-independent API. Supports nearly every SCSI scanner in existence. Provides many varieties of compression including TIFF and its variants. Creates image files on disk. Scanner interface is supported on DOS, Windows, NT, and Macintosh. Development tools cost approximately $2,500; run-time licenses $200 for Fujitsu and $500 for Bell and Howell.

Pixel Translations PixTools/View

C library that provides an image management library allowing applications to support common imaging display and print functions. Includes sample source code. Development tools cost approximately $2,000; run-time licenses are $110.

Kofax KIPP Developer's Toolkit

C library that provides general imaging functionality including scan, display, print, compression, and decompression. Requires proprietary hardware controller for high-speed scanning and printing, although it drives devices at near-rated speed. Includes sample source code. Costs approximately $1,495 per developer. There is no run-time charge, however, scan and print stations require a hardware or software controller costing from $500 to $4,000 each.

Diamond Head Software Image Basic

Adds the imaging functionalities of scanning, OCR, display, image processing, and bar code recognition to Visual Basic by packaging other vendors imaging products in Visual Basic controls. Approximate costs: $1,750 per developer; $50 per display or scanner run-time license.

There is some overlap between these products in the sense that they provide similar capabilities. The AIS Imaging Team selected TMS View Director to provide image display and printing capabilities for another project because of its multi-platform capabilities and high-level API. Pixel Translations PixTools/Scan was also purchased to provide scanning capability because of its Macintosh support and ability to drive SCSI scanners from a high-level API without requiring proprietary hardware. Such configuration is supported for maximum performance, if required.

4.5 Database Schema for Generic Document Repository

An intended future direction for DISS/ET is to support the storage of arbitrary, clearance-related documents falling outside the scope of cradle-to-grave processing provided by DISS/ET. Though HDAR requirements focus on the storage of scanned images, the database schema developed to support these requirements should form the foundation for a generic DISS/ET document repository, capable of storing documents of any type. **Section 3, Specific Requirements**, describes the database requirements for the capture, retrieve, and manage modules in sections 3.1.3.4, 3.2.3.4, and 3.3.3.4, respectively.

Figure 4-3 shows a high-level entity relationship (E-R) diagram of the database schema for the document repository. The E-R diagram is not adequate for implementation-- it is intended only to show the basic entities and their relationships. The "DISS/ET Entities" are assumed to exist within the existing database schema, though the actual names of these entities may be different. "Subject" is an individual for which a clearance is being processed. "Audit Tracing" is the DISS/ET audit mechanism. "Analyst" is a DUI user. "Queue" is a temporary holding queue as discussed in the requirements for the capture module's scan function. "Document" is a document image relative to the requirements put forth herein; this entity is intended to support documents of other types. "Authorized Access" implements the access control functionality discussed in the requirements for manage module's control function.

4.6 Long-Term Storage Subsystem: Alternatives and Recommendations

Optical disk is the most commonly used medium for the long-term storage of document images. A long-term storage subsystem, therefore, includes optical disk hardware, optical disk management software, and a server platform. The following sections discuss alternatives for optical disk hardware and management software respectively. The discussion also includes server issues.

4.6.1 Optical Disk Hardware

There are many manufacturers and systems integrators from which optical disk hardware is available. Drives support one of four standard optical disk sizes: 3.5 inch, 5.25 inch, 12 inch and 14 inch. Capacities range from just under 1 GB to 10 GB per disk. Drives are available as stand-alone units, but are typically packaged into optical disk jukeboxes for large imaging applications. The 5.25 inch jukebox has become a commodity and can be purchased with capacities from five to one thousand disks and one to twenty drives. Storage capacities range from 6.5 GB to 1.3 TB. These products are mostly indistinguishable, though some offer intelligent fetching algorithms and caching. Selection of an optical disk jukebox should be based on its capacity, performance, adherence to standards, and manufacturer reputation.

The AIS Imaging Team presently supports Hewlett Packard (HP) jukeboxes in two installations. Their products will be discussed here as being representative of the optical jukebox marketplace in general. Relevant features and approximate costs of the various HP jukebox models are illustrated below. Each model uses 5.25 inch drives with 1.3 GB capacity that conform to all accepted industry standards and are capable of supporting both rewritable and write-once operation. The jukeboxes have average disk exchange times of 8 seconds, spin-up and spin-down times under 2.5 seconds, and average seek times of less than 24 ms.

Model	HP20T	HP40T	HP120T	HP200T
Disk Capacity	16 disks	32 disks	88 disks	144 disks
Drives	1	2	4	4
Total Capacity	20.8 GB	41.6GB	114.4 GB	182.2 GB
Approximate Cost	$6,995	$16,400	$41,000	$54,000

The cost of HP 1.3 GB optical disk media, WORM or rewritable, is approximately $83 per disk.

4.6.2 Optical Disk Management Software

Optical disk management software can be placed into one of two categories based on features. Low-end managers function as device drivers allowing the host computer to see the optical disk jukebox's file system essentially as a very large hard disk. They typically treat each side of each disk as a sub-directory within a single sub-directory that represents the entire jukebox. High-end managers provide more intelligent capabilities such as:

- file attribute and data caching;
- optimized disk fetching algorithms;
- file system spanning (across disks);
- on-line access to file system meta-data;
- indexing of shelved volumes.

For budgetary and planning purposes, this section discusses Optisys System 2000 and Advanced Archival Products (AAP) AMASS because they are representative of the features and prices of the low-end and high-end markets, respectively. In addition they are recognized market leaders.

Optisys System 2000

Reads media formatted from other optical disk managers. Provides API. Auto-detection of WORM and erasable media. Supports DOS platform. I/O caching. Approximate cost (not including host computer): $2,300 to $5,500 depending on jukebox size.

AAP AMASS

Supports several Unix platforms. Single mount point, single file system interface. Unlimited file system size. Transparent to applications and network protocols (VFS compliant). Supports logical partitioning of capacity. Block oriented I/O. Full Unix command compatibility. Automatic back up of on-line index to jukebox. Supports daisy-chained jukeboxes. Approximate cost (not including low-end Unix host): $8,800 to $25,000 depending on jukebox size.

4.7 Document Imaging Peripherals: Alternatives and Recommendations

This section discusses the issues related to imaging peripherals (namely displays, printers, and fax) and their relevance to the HDAR subsystem.

Display

The importance of a high-quality display for document imaging applications cannot be overstated. A high-resolution (1600 x 1280, 120 dpi), 19-inch or larger, monitor is required to display a full-page image on screen. The most popular image display subsystems for the IBM PC Windows platform are made by Cornerstone. Their DualPage 120i gray-scale display with adapter costs approximately $2,000.

Print

Most of the imaging software products discussed above support image printing to the Windows printer. This requires decompression of the image to about 1 MB before sending it to the printer. As a result, performance is degraded considerably. For increased printing performance, image-capable printers are available which have built-in decompression capabilities. The AIS Imaging Team supports the Talaris 1794 ImageStation printer in several applications. It is capable of printing images at near its rated speed of 17 pages per minute. This printer connects directly to the network and costs approximately $9,300.

Fax

Any Windows-compatible fax mechanism will support the imaging system. Because document images are stored using a format similar to that used by fax transmission, some fax hardware can transmit images directly without first decompressing the file. This can improve performance considerably. Windows compatible fax hardware ranges from $200 to $5,000 depending on features and the number of users supported.

4.8 Possible Future Directions

This section discusses possible future enhancements to the HDAR subsystem, including optical character recognition and full text indexing, bar code recognition, and batch image processing.

OCR and Full Text Indexing

Optical character recognition is a process that converts document image files to machine readable text. This technique is often used to index the full text of documents, so that users can retrieve documents based on their content. In the context of DISS/ET, this capability is probably most relevant to the

adjudication process which will be supported by the ASIST system. OCR and full text indexing would allow adjudicators to search for words within the documents related to a given individual (e.g., find all occurrences of "alcohol" or "drug"). OCR processors range from $500 to $20,000 or more, depending on performance and speed.

Bar Code Recognition

PSF's contain a wide variety of documents, many of which are standard items. Commonly occurring documents can be identified using bar code, either printed directly on the form or affixed to it before scanning, so that a bar code recognition process can automatically index the document without human intervention. This can have a significant impact on the overall throughput of the indexing process. Prices for this capability are in the same range as OCR products. They are often included in the same package.

Batch Image Processing

An image processor can greatly enhance the utility of scanned images for viewing, OCR, or forms processing applications by assuring clean, consistent, and precisely positioned images. Some image processing functions often performed after scanning, in batch mode, include:

- skew detection and correction;
- document registration (positioning);
- dot shading removal and de-specking;
- line removal with character repair;
- inverse type detection and correction;
- image rotation;
- image zoning (breaking an image into sub-images).

Sequoia Data's ScanFix provides these capabilities, either via a Windows developer's toolkit or as a standalone application. The costs for these products are $795 and $495, respectively.

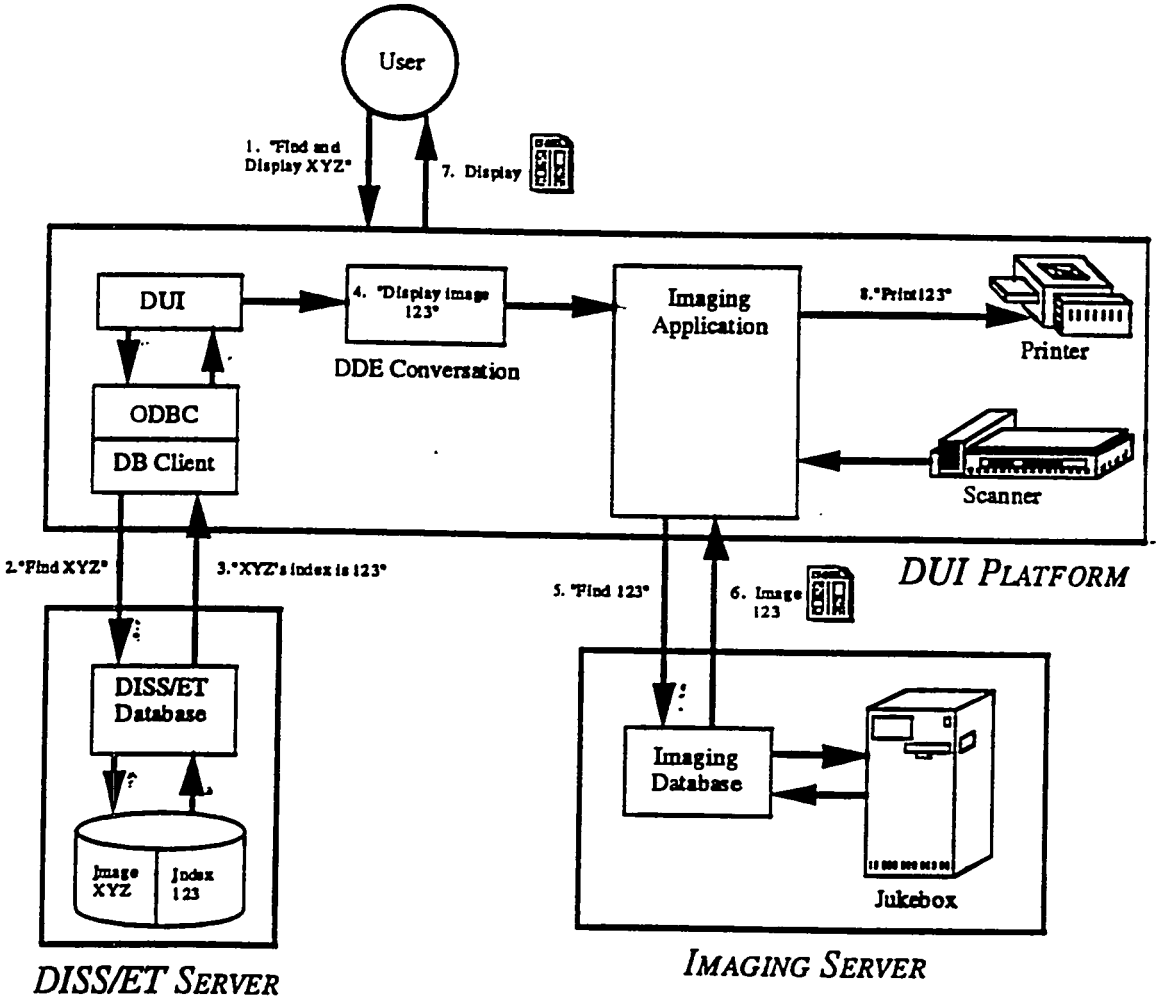


Figure 4-1

HDAR Architecture Using Image-enabling Approach

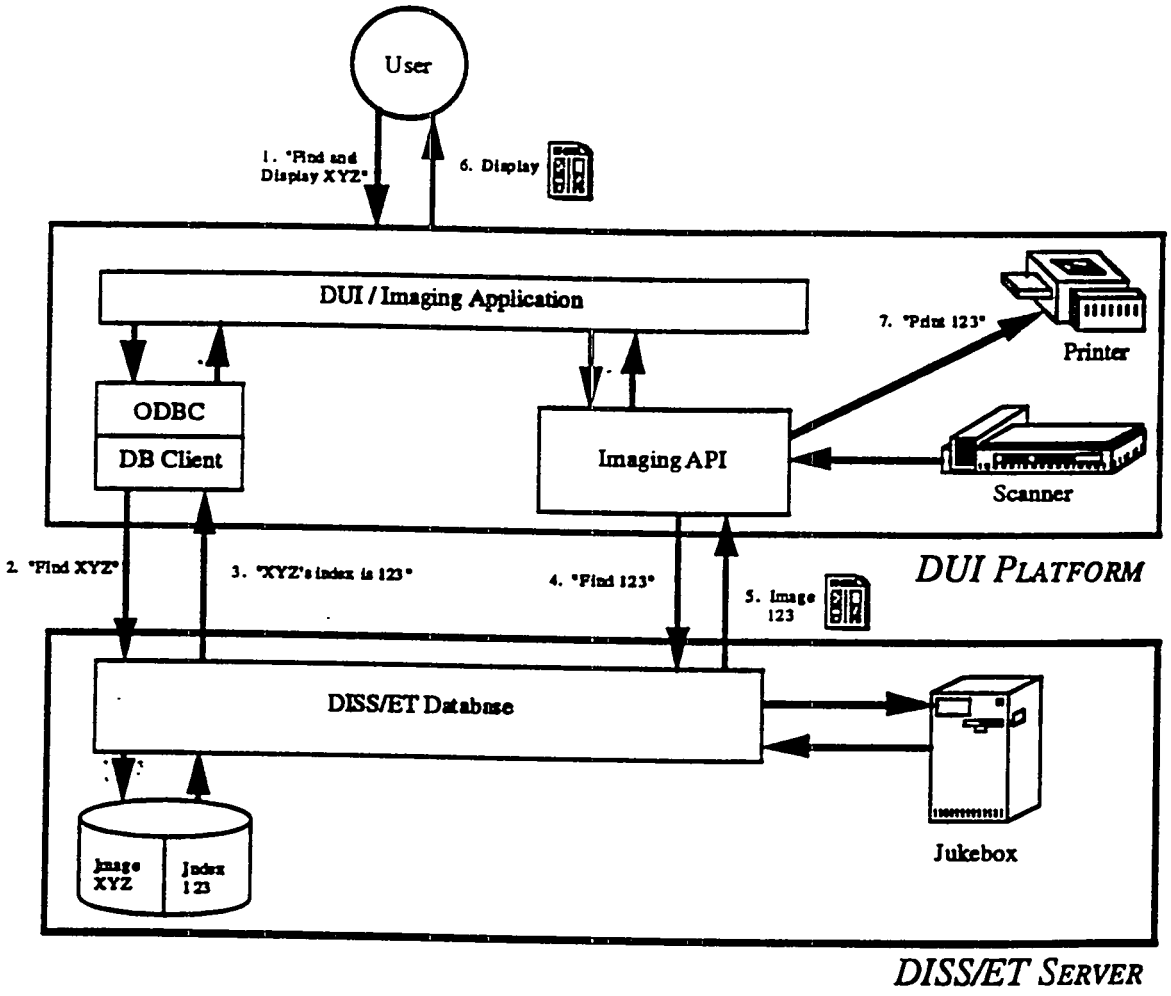


Figure 4-2

HDAR Architecture Using Imaging Application Development Toolkit Approach

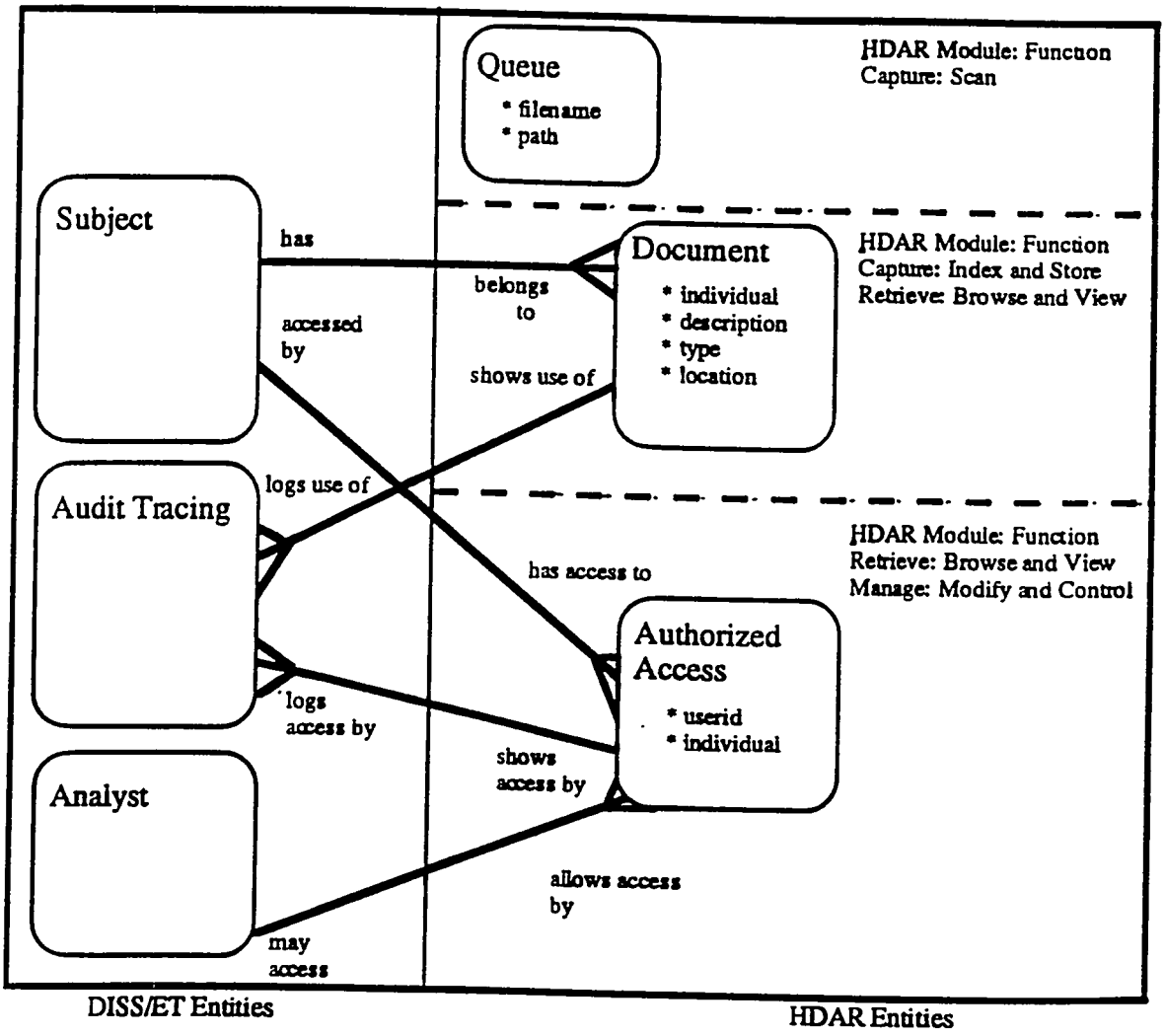


Figure 4-3

HDAR High-level Document Repository Entity Relationship Model

Vault Analysis

DOE Oakland Operations Office
Personnel Security Division

Table of Contents

Background 49
Reference Materials Received or Referenced 49
Physical Attributes of the Vault 49
Personnel Security File Folder (PSF) 50
Access Control & Audit 51
Changes to the Contents 51
Filing Pages in Personnel File Folders 51
Removing Pages from Personnel File Folders 52
Changing Data in Personnel File Folders 52
Filing Folders in the Vault 52
Retrieving Folders from the Vault 52
Retention Rules & Exceptions 53
Document Volume Estimates 53
Folder Estimates: 53
Page Estimates: 54
Reinvestigation Estimates 54
Questions asked / Suggestions for imaging features: 55
Other comments: 545

Background

During July and August, 1994, numerous discussions and interviews were held with the following DOE Oakland Operations, Personnel Security Division staff:

Sally Murphree, DOE supervisor;
Tess Stamatis, Vault operator, Lee Associates;
Brian Bouldin, Project Manager, Lee Associates.

The purpose of these meetings was to conduct an analysis of the Personnel Security Division's Secure Vault used to store all Personnel Security File Folders under the auspices of the DOE Oakland Operations Office.

Reference Materials Received or Referenced

"Personnel Security Program", DOE Order 5631.2C;

"Records Disposition", DOE Order 1324.2A;

"Standard Operating Procedure, Personnel Security Vault",
DOE Oakland Operations Office, Personnel Security Division (PSD).

Physical Attributes of the Vault

The Personnel Security Vault managed by the DOE Oakland Operations Office Personnel Security Division is located in Trailer 5979, Room 133. The size of the vault is approximately 20 feet by 22 feet. Standard metal filing cabinets are placed around the perimeter of the room. Two rows of double-sided metal filing cabinets are in the center of the room. Each filing cabinet has seven shelves, each of which is 46" wide and deep enough to file one row of personnel security folders. In the vault, the majority of space consists of Active Security Clearance Personnel Files. A smaller section is where the Pending Security Clearance Personnel Files are filed. Immediately inside the door to the left are various sections of shelves assigned to specific "Screeners." This is where the files that are being adjudicated rest. See Figure VO-1 of this analysis for the vault layout.

Personnel Security File Folder (PSF)

Statistics and Characteristics

Volume

The vault contains approximately 13,200 active personnel security file folders. In addition to active PSF's, a filing cabinet containing Secret files is also in the vault. The files in this cabinet are outside the scope of this analysis. Records which are classified above the normal classification for personnel security investigation files will not be included as a part of the historical document archival and retrieval system. Terminated folders are kept in this vault until they are transferred to longer term storage at the Federal Records Center in San Bruno, California.

Folder Makeup

DOE Order 5631.2C is used as a guide for physical file folder creation. The standard PSF consists of a blue manila folder with a single index tab containing the individual's name, social security number and clearance investigation number. Most folders observed are approximately one-half to one inch thick. A very small number of folders are large enough to require "rubber banding" to keep several folders or individual documents together on the shelf.

File Retention

DOE Order 1324.2A is used to determine file retention periods. Personnel files are kept for all active clearance holders. Some of the folders date back to the 1950's. After an employee terminates, the folder is kept for approximately 6 months. The files are then transferred to the Federal Records Center. The files are kept there for 10 years and then destroyed by shredding and/or burning.

** Special Note **
As of the date of this analysis, there is a DOE-wide moratorium in effect preventing the destruction of any personnel records. This moratorium has been ordered by the Secretary of Energy and is related to current issues and investigations revolving around past radiation research.

Document Condition

The older personnel files have more handwriting on the documents contained within. This is due to previous methods of filling out the forms as well as the methods used by analysts in making notations within the individual PSF's. The documents in the folders vary in color, consistency, size and condition. The form of the data on the documents varies considerably from handwritten to typed and includes some graphics, such as lines for field separators, fingerprints, and logos. Highlighting and underlining appear throughout the documents.

Note: This information was gained by reviewing, from a distance, the contents of a randomly selected folder. Subsequent to this review, a more thorough review was authorized and conducted. See Appendix A, Description of Document Types in Personnel Security Folders for more detailed information.

Access Control & Audit

Vault access procedures are based on DOE 5631.2C. An audit trail is created for every PSF moved from the confines of the vault. Normally, all PSF's are stored in the vault until the employee terminates. At that time, they are then scheduled to be physically moved to longer term storage at the Federal Records Center. Access to PSF's is only allowed on a "need-to-know" basis. Requests for PSF check-out are routinely received from analysts within the DOE Oakland Operations Office Personnel Security Division. The analysts have duties which include clearance adjudication, Personnel Assurance Program (PAP) administration, and Personnel Security Assurance Program (PSAP) administration.

In addition, requests for PSF's are received from personnel in the Office of Personnel Management (OPS) as OPM conducts the overwhelming majority of DOE clearance investigations. A few requests are made by personnel from the Federal Bureau of Investigation (FBI) commensurate with their involvement in a small number of clearance investigations. On occasion, the DOE Office of the Inspector General (OIG) is involved in a clearance investigation and will request access to Personnel Security File Folders. The OIG becomes involved only in cases of waste, fraud or abuse of government property.

Requests to check out specific PSF's are documented on the PSF Checkout form signed by the requester and the vault custodian. In addition, a checkout card, normally placed inside of the PSF folder, is removed by the vault custodian who places it in a checkout file until the folder is returned. At that time the vault custodian reinserts the card into the PSF.

Changes to the Contents

Occasionally, vault personnel will be asked to file new pages, remove pages, or change data on a page or pages within the PSF.

Filing Pages in Personnel File Folders

An example of the need to file new pages in a PSF is when Internal Security prepares an Infraction or Incident Report involving an individual. Security sends a copy to the Personnel Security Division who initiates the process to have the report filed in the individual's PSF.

Removing Pages from Personnel File Folders

It is sometimes necessary to remove single or multiple pages from an individual's Personnel File Folder. Removing pages may be due to an error in original filing, information on the page(s) is no longer relevant or accurate, or in the case of an Infraction or Incident Report, the issue is resolved and no permanent record is desired.

Changing Data in Personnel File Folders

One example of the need to change data on specific pages within the Personnel File Folder is in the case of a marriage and the resulting name change. Any legal name change must be reflected in the PSF. Another example, less frequent but further illustrating the need, is the case of an individual who has undergone a sex change operation.

Filing Folders in the Vault

Request Types (Analysts, Processors, PSAP, PAP)

Filing requests come from several sources. Routine requests for single or small batch filing comes from the processors and analysts. These requests, on average, range from one to ten PSF's. In addition, two DOE programs involve regular and routine review of PSF's for accuracy and completeness. These programs are the Personnel Security Assurance Program (PSAP) and the Personnel Assurance Program (PAP). Filing requests from these programs can range as high as 20-30 PSF's per day.

The average PSF filing requests into the vault is approximately 50 per day.

Process Used

PSF's are brought to the vault custodian's desk and are placed in boxes on an adjacent table for batch filing. Occasionally, the PSF is filed in the vault immediately upon receipt by the vault custodian. The primary filing "key" for PSF's is alphabetically by last name.

Retrieving Folders from the Vault

Retrieval requests for PSF's are essentially the inverse of the description in the above section on Filing Personnel Security File Folders with the exception of the creation of new folders.

In the event that an individual will undergo a psychiatric evaluation, the entire contents of the PSF is duplicated (and labeled "Duplicate") and checked out to the evaluator. Upon return or when the duplicate is no longer needed, it is destroyed.

Retention Rules & Exceptions

Per DOE Order 1324.2A, Personnel Security File Folders are required to be maintained for all personnel with an active security clearance. When a person is terminated, a "ten-year clock" begins. Upon completion of the ten-year period, the individual's PSF records are destroyed. The only exception to this rule is when a terminated employee dies; their records can be destroyed immediately.

** Special Note **
As of the date of this analysis, there is a DOE-wide moratorium in effect preventing the destruction of any personnel records. This moratorium has been ordered by the Secretary of Energy and is related to current issues and investigations revolving around past radiation research.

Document Volume Estimates

Folder Estimates:

Based on Staff knowledge: 13,200 folders

Estimates ranged from 10,000 to 16,000 folder files in the vault. Sally Murphree investigated further and stated that as of June 7, 1994, the vault contained 13,200 files. Discussion around estimating the number of pages per file led to no conclusive ideas. Files range from less than 1/2" thick to some described as 8 to 10" thick. (My visual observation was a range of 1/2" to about 3 or 4 " with ~90% of the folders less than 1" thick.)

Sample folder count method: 14,700 to 17,220 folders

Based on a sample count of folders on four shelf compartments, the total estimated folder capacity in the vault is in the range of 14,700 to 17,220 folders. (Folders counted per shelf were 70,73,71 & 82.) The range was derived using the low (70) and high (82) numbers multiplied by the number of shelf compartments available (210).

Page Estimates:

Based on staff estimate of number of folders:

Assumptions: 1" = 75 pages
All pages single-sided content
Staff estimated folders: 13,200

Folder size:	Number folders this size:		% of total
1/2"	11,880	(5940")	90%
1"	990	(990")	7.5%
4"	330	(1320")	2.5%
Total inches:		8,250"	

Estimated volume: 618,750 pages

Based on sample folder count:

Assumptions: 1" = 75 pages
All pages single-sided content
Estimated folders: 17,220 (high estimate)

Folder size:	Number folders this size:		% of total
1/2"	15,490	(7,749")	90%
1"	1291	(1,291")	7.5%
4"	430	(1,720)	2.5%
Total inches:		10,760"	

Estimated volume: 807,000 pages

Based on estimated linear feet of shelf storage (805 feet):

Assumptions: 1" = 75 pages
All pages single-sided content
10% shelf space is unused

Estimated volume: 724,500 pages

Reinvestigation Estimates

The following Clearance Reinvestigation projections were obtained from the CPCI System (Central Personnel Clearance Index) These estimates are for clearance investigations to be processed through DOE Oakland Operations Office in the fiscal years 1995, 1996 and 1997. Two types of reinvestigations are conducted: Periodic Reinvestigation - Special Background Investigation (PRIS) and National Agency Check with Credit (NACC). The numbers for each type are listed separately by fiscal year.

Several backfile conversion strategies will be evaluated and considered as a follow-on to this analysis. The information below will be used for estimating the effort required for backfile conversion of the personnel security folders which are involved in reinvestigations.

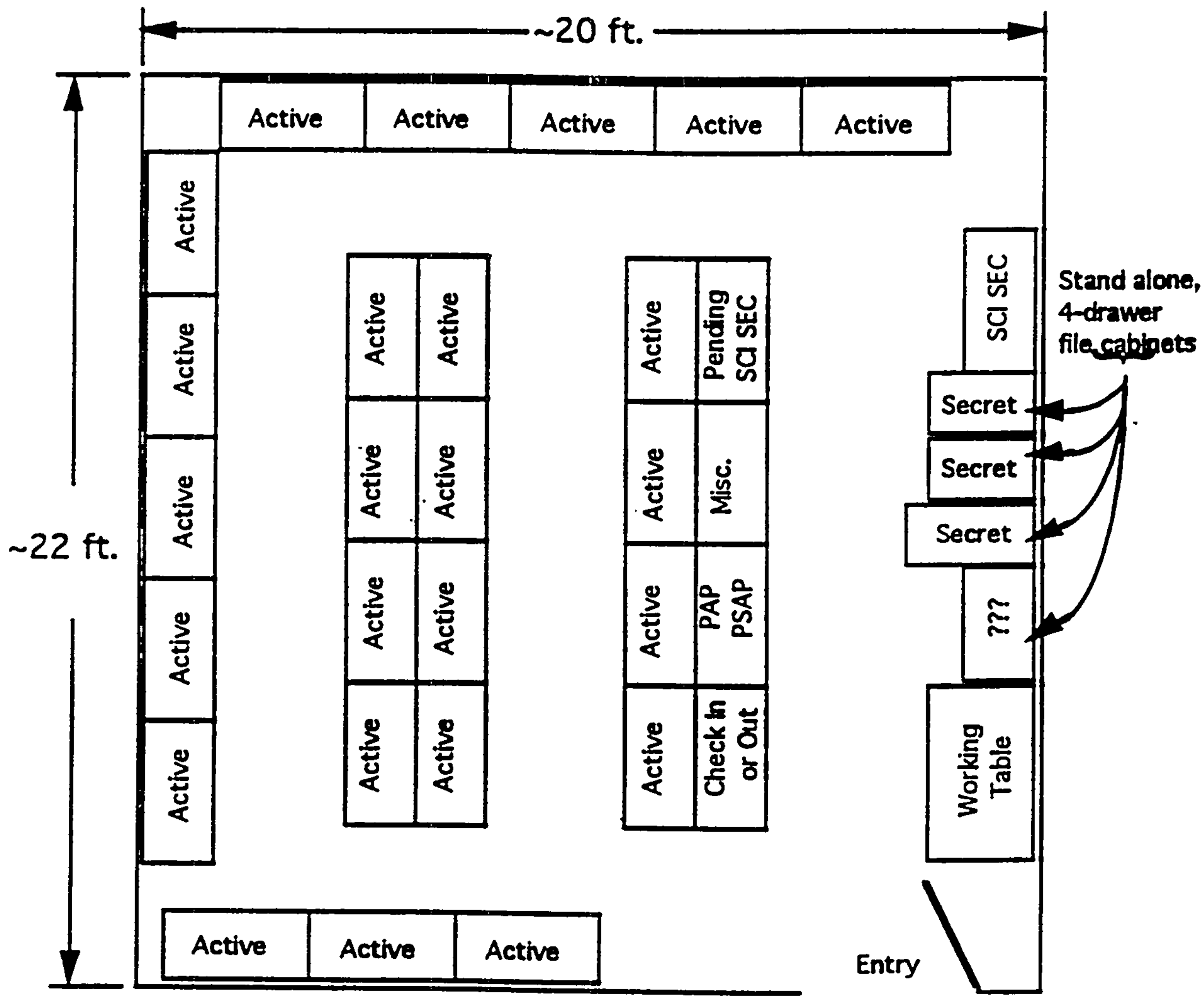
Fiscal Year 1995 -	645	(PRIS)
	1146	(NACC)
Fiscal Year 1996 -	970	(PRIS)
	755	(NACC)
Fiscal Year 1997 -	884	(PRIS)
	761	(NACC)

Questions asked / Suggestions for imaging features:

Will we (DOE Oak) be able to electronically transmit files to other offices?

Other comments:

We need to begin to determine the security access methods to replace the manual check-out process. Sally Murphree stated she is or will be working on this.



Key:

Description	Represents a shelf compartment; each compartment is stacked 7 high and is 46" wide. Each compartment holds personnel security folders in one of the following categories:

- Active
- Pending
- PAP
- PSAP
- Check In or Out
- SEC SCI

Figure VO-1

DOE Oakland Operations Office Vault Layout

Vault Analysis

DOE Albuquerque Operations Office
Personnel Security Division

Table of Contents

Background 59
Reference Materials Received or Referenced 59
Physical Attributes of the Vault 59
Personnel Security File Folder (PSF) 60
Contents 60
Access Control & Audit 61
Changes to the Contents: 61
Filing Pages in Personnel Security File Folders 61
Removing Pages from Personnel Security File Folders 62
Changing Data in Personnel Security File Folders. 62
Filing Folders in the Vault. 62
Retrieving Folders from the Vault. 63
Retention Rules & Exceptions. 63
New System Requirements 63
Bar Code Project 64
Reinvestigation Estimates 64
Document Volume Estimates 64

Background

The information contained in this document was obtained during an interview conducted on July 27, 1994 with Karen Kerby and Manuel Cabral of the DOE Albuquerque Operations Office Personnel Security Division. The focus of the interview was on the nature, operations and characteristics of the secured vault used to store all active Personnel Security File Folders (PSF) under the auspices of the DOE Albuquerque Operations Office. The goal of this meeting was to obtain a detailed understanding of the requirements which will lead to the specification of a document archival and retrieval system integral to the DOE Integrated Security System Electronic Transfer (DISS/ET) solution being created at this time.

Reference Materials Received or Referenced

"Personnel Security Program", DOE Order 5631.2C;

"Records Disposition", DOE Order 1324.2A;

"Standard Operating Procedures for the PSB Vault",
DOE Albuquerque Operations Office, Personnel Security Branch (PSB),
July 28, 1994.

Physical Attributes of the Vault

The Personnel Security Vault is located in Building 331 at the DOE Albuquerque Operations Office complex adjacent to Sandia National Laboratory, both of which are on Kirtland Air Force Base in Albuquerque, New Mexico. See Figure VA-1 for the vault layout.

The entry to the vault is on the southeast corner of the room. Just inside and to the west of the single entry door, office cubicles take up approximately 200 square feet of floor space. A small work area of approximately 60 square feet used for hole punching and assembling copies is located in the northwest corner of the vault. The side perimeter walls of the room are mostly covered by shelving (eight shelves high). The total linear storage space on the perimeter shelf units is approximately 520 feet. Two sets of movable shelf units occupy the majority of floor space. Each of the movable storage shelves are 18 1/2 feet long and have 5 movable shelf units each of which has 8 shelves per side. The total linear storage space on both movable shelf units is approximately 2,960 feet. The total vault linear storage space for PSF file retention is approximately 3,480 feet.

Personnel Security File Folder (PSF)

Statistics and Characteristics.

Volume

Approximately 40,000 active PSF's are filed within the vault. In addition, approximately 20,000 PSF's for terminated employees are held there. The total PSF's in this vault is approximately 60,000. Terminated PSF's are regularly transported to another on-site location for longer term storage until their ultimate destruction. Approximately 25,000 PSF's for terminated employees are stored at that location.

Folder Makeup

DOE Order 5631.2C is used as the basis for physical file folder creation. The standard PSF consists of a colored manila folder with a single index tab. The color of the folder is determined from a scheme matching the first character of the first name with a specific color. On the index tab is the person's name and their clearance number (format: aannnnnn). The contents of some employees' files have exceeded the size limitations of the manilla folder and consist of multiple folders held together with rubber bands. Some PSF's are large enough to require multiple boxes to hold the individual's complete set of records. The larger files are often due to the inclusion of transcripts from an administrative review process. The number of PSF's this size are few and are typically those of terminated employees.

Contents

Due to controlled access regulations, review of the contents of the Personnel Security File Folders was not allowed. In the event an in-depth review is deemed necessary, a formal written request will need to be sent to the appropriate DOE management personnel for authorization. For the purpose of this analysis, visual observation was allowed from a distance sufficient to prevent content compromise.

Generally, the contents of PSF's include a variety of paper types, colors and conditions. Some of the folders and contents are over 30 years old. The data is recorded in a variety of forms including handwritten, typed, computer printouts, graphic, and table layout. Handwritten annotation and highlighting was observed on various pages throughout the PSF. Single and double-sided forms are present. Paper size is mostly standard 8 1/2 x 11 inches, however, a variety of paper sizes from small note pad to legal size and larger was noted.

Physical thickness of the PSF's ranged from a newly created PSF with a few pages content to one PSF contained within four "xerox" boxes. For the purpose of estimating the volume of pages within the entire vault, the following estimates represent active PSF's only.

Folder Size	Approximate number of PSF's this size
1/2" or less	500 - 1000
1-2"	38,000 (Majority about 1 ")
2-4"	200-300

Note: Some larger files were observed. For the purpose of noting their existence, estimates are included below. However, vault personnel were fairly certain that most of the larger PSF's belonged to terminated employees and will therefore not be included in the scanning process. The remaining 20,000 terminated PSF's contained within the B. 331 vault average 1 - 2 " thick.

4-12"	200-300
Over 12"	50 - 75 (up to 8 feet)

Access Control & Audit

Vault access procedures are based on DOE 5631.2C. An audit trail is created for every PSF moved from the vault. Normally, all PSF's are returned to the vault for storage until they become classified as "terminated." At that time, they are then scheduled to be physically moved for longer term storage. Access to PSF's is only allowed on a "need to know" basis. Requests for PSF check-out are routinely received from analysts and processors within the DOE Albuquerque Operations Office Personnel Security Division. The analysts have duties which include clearance adjudication, Personnel Assurance Program (PAP) administration, and Personnel Security Assurance Program (PSAP) administration.

In addition, requests are received from personnel of the Office of Personnel Management (OPS) as OPM conducts the majority of DOE clearance investigations. A few requests are made by personnel from the Federal Bureau of Investigation (FBI) commensurate with their involvement in a much smaller number of clearance investigations. On occasion, the DOE Office of the Inspector General (OIG) is involved in a clearance investigation and requests access to Personnel Security File Folders. The OIG becomes involved only in cases of waste, fraud or abuse of government property.

Changes to the Contents

Occasionally, vault personnel will be asked to file new pages, remove pages or change data on a page or pages within the PSF.

Filing Pages in Personnel Security File Folders

An example of the need to file new pages in a PSF is when Internal Security prepares an Infraction or Incident Report involving an individual. Security sends a copy to the Personnel Security Division who initiates the process to have the report filed in the individual's PSF.

Removing Pages from Personnel Security File Folders

It is sometimes necessary to remove single or multiple pages from an individual's Personnel Security File Folder. Removing pages may be due to an error in originally filing it in the PSF, information on the page(s) no longer relevant or accurate, or in the case of an Infraction or Incident Report, the issue is resolved and no permanent record is desired.

Changing Data in Personnel Security File Folders.

One example of the need to change data on specific pages within the Personnel Security File Folder is the case of a marriage and the resulting name change. Any legal name change must be reflected in the PSF. Another example, less frequent but further illustrating the need, is the case of an individual who has undergone a sex change operation.

Filing Folders in the Vault.

Request Types (Analysts, Processors, PSAP, PAP).

Filing requests come from several sources. Routine requests for single or small batch filing comes from the Processors and Analysts. These requests, on average, range from one to ten PSF's. In addition, two DOE programs involve regular and routine review of PSF's for accuracy and completeness. These programs are the Personnel Security Assurance Program (PSAP) and the Personnel Assurance Program (PAP). Filing requests from these programs can range as high as 300-400 PSF's per day.

The average PSF filing requests into the vault is approximately 200 per day.

Process Used.

PSF's are brought into the vault and placed just inside the door on the Pending Filing shelves. A file cabinet with several shelves is reserved specifically for incoming PSF's. The PSF's are left on these shelves by analysts, processors and vault personnel. Vault personnel then take the staged PSF's and place them in their appropriate location in the proper file locations on the shelves in the vault. The primary filing "key" for PSF's is alphabetically by last name. Based on the beginning character of the first name keyed to a color, the secondary filing "key" is the person's first name.

Routine filing occurs when PSF's are left in a "staging area" just inside the vault door.

Retrieving Folders from the Vault.

Retrieval requests for PSF's are essentially the inverse of the description in the above section on Filing Personnel Security File Folders with the exception of the creation of new folders.

In the event that an individual will undergo a psychiatric evaluation, a request is received to duplicate the entire contents of the PSF. The copy is labeled "Duplicate" and checked out to the evaluator. Upon return or when the duplicate is no longer needed, it is destroyed.

Retention Rules & Exceptions.

DOE Order 1324.2A requires Personnel Security File Folders to be maintained for all personnel with an active security clearance. When a person is terminated, a "ten-year clock" begins. Upon completion of the ten-year period subsequent to the termination, the individual's PSF records are destroyed. The only exception to this rule is when a terminated employee dies; their records can be destroyed immediately.

** Special Note **
As of the date of this analysis, there is a DOE-wide moratorium in effect preventing the destruction of any personnel records. This moratorium has been ordered by the Secretary of Energy and is related to current issues and investigations revolving around past radiation research.

New System Requirements

The following items are intended to translate into requirements but are not necessarily in that form at this time. The comments below represent the perspective of vault management personnel interviewed.

1) Need an inventory listing of all files stored.

2) Need the system to be consistent with or interface with the Central Personnel Clearance Index system located at DOE HQ and used by the Operations Offices to assist in the management of Personnel Security File Folders.

3) Retrieval methods need to include the capability to search on last name as well as to be able to phonetically search on last name (e.g., Smith vs. Smyth).

4) Need to be able to conduct historical name searches. (Names that an individual had prior to a legal name change.) Currently, name is overwritten on new paperwork and in the available tracking and support systems.

5) Need to be able to remove duplicate forms and pages from within the PSF.

6) Need to make sure that when records are scanned, critical retrieval information is verified.

Bar Code Project

A project is currently underway to add bar code labels to the index tabs of the individual Personnel Security File Folders. The intent of the Bar Code project is to facilitate the filing and retrieving processes relative to the data entry related part of those processes. Estimated completion date for this project is February 1995.

Reinvestigation Estimates

The following estimated Clearance Reinvestigation projections were obtained from the Central Personnel Clearance Index (CPCI) System. These estimates represent the approximate number of clearance investigations to be processed through DOE Albuquerque Operations Office.

Several backfile conversion strategies will be evaluated and considered as a follow-on to this analysis. The information below is provided for estimating the effort required for backfile conversion of the personnel security folders which are involved in reinvestigations:

Fiscal Year 1995 -	3,000 reinvestigations;
Fiscal Year 1996 -	11,000 reinvestigations;
Fiscal Year 1997 -	7,500 reinvestigations.

Document Volume Estimates (Active PSF's only)

Entire Vault:

Based on available shelf space

Assumptions: 1" = 75 pages
Shelf space 90% full
All pages single-sided content
Available shelf space is 3480 linear feet

Estimated volume: 3,132,000 pages

Based on estimated count of folders and folder sizes

Assumptions:
1" = 75 pages
All pages single-sided content
Estimates below are valid:

Folder size:	Number folders this size:		
1/2"	1000	(500")	2.5%
1"	38,000	(38,000")	96.7%
4"	300	(1200")	.8%

Estimated volume: 2,977,500 pages

Based on Fiscal Year Projection for Re-investigations: (from CPCI System)

Fiscal Year 1995: Estimated 3,000 reinvestigations

Assumptions: Folder sizes will be statistically representative:

1/2 "	=	75 folders	(37.5")
1"	=	2,901 folders	(2901")
4"	=	24 folders	(96")

Total folder "inches" for 3,000 reinvestigations = 3,034.5"
1" = 75 pages
All pages single-sided content

FY 1995 Estimated volume: 227,587 pages

Fiscal Year 1996: Estimated 11,000 reinvestigations

Assumptions: Folder sizes will be statistically representative:

1/2 "	=	275 folders	(137.5")
1"	=	10,637	(10,637")
4"	=	88 folders	(352")

Total folder "inches" for 11,000 reinvestigations = 11,126.5"
1" = 75 pages
All pages single-sided content

FY 1996 Estimated volume: 834,487 pages

Fiscal Year 1997: Estimated 7,500 reinvestigations

Assumptions: Folder sizes will be statistically representative:

1/2 "	=	187 folders	(93.5")
1"	=	7,252	(7,252")
4"	=	60 folders	(240")

Total folder "inches" for 11,000 reinvestigations = 7,585.5"

1" = 75 pages

All pages single-sided content

FY 1997 Estimated volume: 568,912 pages

DOE Albuquerque Operations Office
Personnel Security Vault

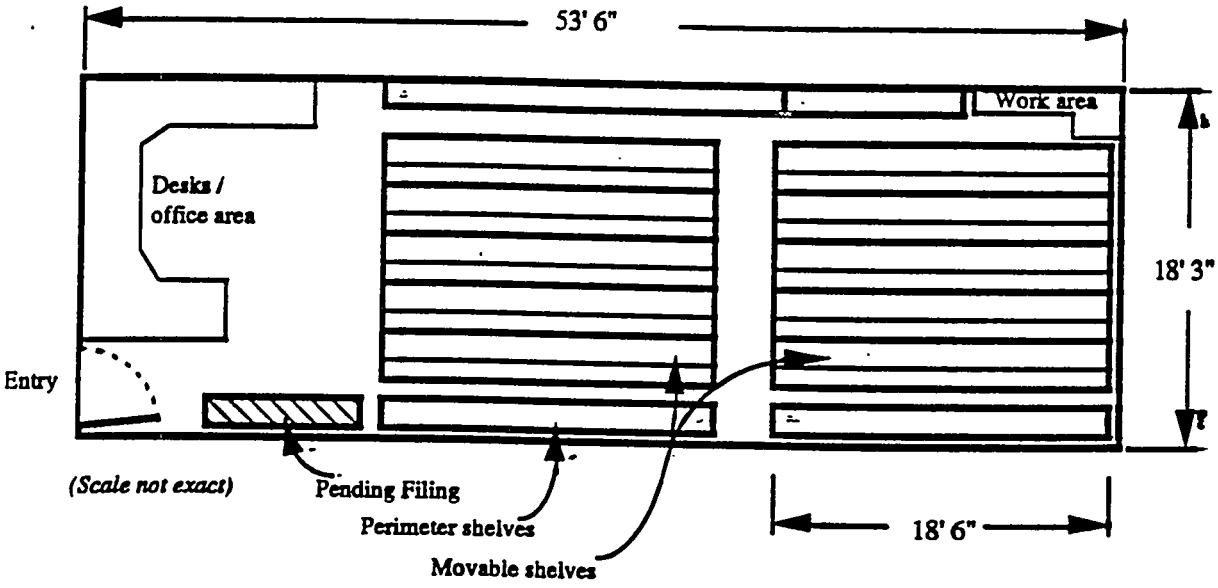


Figure VA-1

DOE Albuquerque Operations Office Vault Layout

Description of Document Types

The following pages contain a list of documents identified during a review of a random sample of Personnel Security File Folders. This list does not include every possible type of documents found in PSF's. It is intended to provide a representative sampling only.

Document Name or Description	Single/ Double Sided	Size / Color	Document Characteristics
Authority for Release of Information	S	8-1/2x11 White	Black (FAX copy) ink and preprinted text; black signature and date.
Background Investigations Case Transmittal	S	8-1/2x11 White	Preprinted black, line separators; blue ink handwriting; some typed entries.
Case Closing Transmittal	S	8-1/2x11 White	Impact printer; time stamp, black ink.
Case Evaluation for Security Clearance	S	8-1/2x11 Turquoise	Typed with line separators; black and blue signatures and dates; purple handwriting; black handwriting.
Case evaluation	S	8-1/2x11 Blue	Typed black ink; black ink handwriting.
Certification of Amended Investigation Form	S	8-1/2x11 White	Black ink; preprinted black ink/lines; black ink handwriting.
Certification of Investigation	S	8-1/2x11 Off-White	Impact print (black).
Check (Xerox copy)	S	8-1/2x11 White	Typical bank check w/handwriting.
Clearance Activity Sheet	S	8-1/2x5 Beige	"Flimsy" paper; carbon printing.
Clearance Activity Sheet	S	8-1/2x5 Green	Carbon printing; carbon handwriting; blue ink date.
Clearance Justification and Certification	S	8-1/2x11 White	Preprinted; blue ink handwriting; black ink handwriting.
Conduct of Personnel Interview	S	8-1/x11 White	Appears to be a xerox copy; logo stamped over text; black ink signature
Contact Log	S	8-1/2x11 White	Black column separators and column typed names; black ink handwriting.

Credit Inquiry Panel	D	8-1/2x11 White	Impact print; blue pencil mark bottom of page; Preprinted on back.
Credit Summary	S	8-1/2x11 Cream	Blue carbon printing; impact printing; thin.
Data Report on Spouse DOE F5631.34 (10-86)	D	8-1/2x11 White	Preprinted form with field delineators and descriptors; black ink handwriting; blue ink signature and date; red stamp "Mailed to...."; black ink notations (text circled); time stamp purple.
DOE F5631.10 U.S. DOE Waiver	S	8-1/2x11 White	Preprinted; blue ink signature and date.
DOE Form DP-1 (Part 1) P.S.Q..	S Page 2 is "D"	8-1/2 x 22 White	Preprinted black.; typed black; stamped red; black felt handwritten notations; black ink signature; thin paper (Tissue like)
Drug Certification	S	8-1/2x11 Yellow	Black preprinted w/logo and text. Black ink printing and signature.
Employee's Initial Medical Examination Record	S	8-1/2x11 White	Xerox or fax copy; medical form with check boxes and handwritten comments and information.
Envelope from M. D.	n/a	3"x8" White	Black ink handwritten ; postmark w/postage stamp.
Envelope with Standard Cassette tapes inside	n/a	3"x8" White	Return address preprinted; black ink writing; pencil writing.
Facsimile Transmission Sheet and Medical History Forms	S	8-1/2x11 White	Black preprinted; black ink handwriting.
FBI Identification Record	S	8-1/2x11 White	Typed black ink.
File Review Log	S	8-1/2x11 White	Typed with black ink handwriting.; black xerox copy with original black ink signature.

ID 258 - Finger Print Form	D	8-1/2x8-1/2 White	Blue preprinted field separates and descriptors; fingerprints (black ink); blue and black signatures. Red stamp (LLNL reinvestigation); green time stamp; various red/burgundy time stamps.
Medical Statement for New Employees	S	8-1/2x11 White	Black preprinted; black type; black pen handwriting.
Memo from DOE PSB	S	8-1/2x11 Yellow	Typed; routing stamp black ink, stamped "Original signed by".
Memo-affiliation change	S	8-1/2x11 White	Black typed; black time stamp.; blue pen annotation; red highlight.
Memo-SNL	S	8-1/2x11 White	Typed; back ink signature; black time stamp.
Notice of Commencement Under Chapter 7 of Bankruptcy code	S	8-1/2x11 White	Preprinted black ink, black typed.
Official Transcript Informal Interview/Security Clearance	S	8-12x11 White	"Legal Format"; typed with line numbers in left margin; red stamp (OUO); some smudging throughout (72 page transcript).
OP Form 366 (Rev. 10/81)	S	8-1/2x11 White	Preprinted with delineated fields (lines) and field descriptors; typewriter print; blue pencil mark on each page.; red pencil parentheses around some text; handwritten notations black ink; paper clip at bottom of numerous pages in this document.
Personal Financial Statement	S	8-1/2x11 White	Typed; tabular form.
Personnel Investigations Case Data	S	8-1/2x11 White	Preprinted black; matrix used for checks in various categories. Highly graphic. Handwriting black ink.

QSP SF 86	D	8-1/2x11 White	Xerox copy - all black - graphics/lines/printing/ typed.
Questionnaire for Sensitive Positions	S	8-1/2x11 White	Black typed.
Reinvestigation Program	S	8-1/2x11 White	Purple pen writing.
Report of Agency Adjudicative Action on OPM Personnel Investigations	S	8-1/2x11 White	Blue preprinted; impact print (black); "OCR" mark the oval boxes.
Report of Birth		8-1/2x11 White	Preprinted black, typed content, black ink signature.
Report of Investigation	S	8-1/2x11 Off-White	Black preprinted with typed content.
Report of National Agency Check	S	8-1/2x11 White	Preprinted black ink; impact printing; blue pencil check marks; red ink box stamped on bottom with black ink handwriting; thin paper.
Sealed Envelope with dictaphone cassette tape inside	n/a	3x8-1/2 White	Red "OUO" stamp; black ink handwriting.
Security Acknowledgment	S	8-1/2x11 White	Preprinted text.; black ink dates and signature; red ink LLNL stamp.
Security Analysis of Proposed Travel to Sensitive Countries	S	8-1/2x11 White	Preprinted black, typed, handwriting
Specific Release	S	8-1/2x11 Light Blue	Preprinted; black ink; blue ink handwriting
Standard Form 86 (86-108) Questionnaire for Sensitive Positions	D	8-1/2x11 Off-White	Preprinted brown ink with field separators and field identifiers (A, B, C, etc.); typewriter print; handwriting black ink; time stamp in red ink; red notations throughout; blue pencil line on bottom of each page; yellow highlights; signatures in black ink.

Summary of Security Suitability Data and Recommendation	S	8-1/2x11 White	Xerox: typed. black handwriting.
Summary Sheet	S	8-1/2x11 White	Preprinted; typed; handwriting; all black ink.
Summary Sheet including standard lined notebook paper (blue lines on white)	S	8-1/2x11 White	Black pre-printed sections; black ink handwriting.
U.S. Individual Income Tax Form (incl.. all attachments W-2, etc.)	S	8-1/2x11 White	Xerox copy, black.
Witness Identification Sheet	S	8-1/2x11 Orange	Preprinted form; black/lines; typed content.