Reference Manual for PERPOS:
An Electronic Records Repository and
Archival Processing System
Version 3.1

PERPOS Technical Report ITTL/CSITD 06-02
September 2006

William E. Underwood
Sandra Laib
Marlit Hayslett

Computer Science and Information Technology Division
Information Technology and Telecommunications Laboratory
Georgia Tech Research Institute
Georgia Institute of Technology

The Army Research Laboratory (ARL) and the National Archives and Records Administration (NARA) sponsor this research under Army Research Office Cooperative Agreement DAAD19-03-2-0018. The findings in this paper should not be construed as an official ARL or NARA position unless so indicated by other authorized documentation.
Foreword

The Presidential Electronic Records Pilot System (PERPOS) is a research project sponsored by the ERA Program of the National Archives and Records Administration and lead by the Georgia Tech Research Institute. The initial objective of the PERPOS project was to support archivists in gaining intellectual and physical control over the personal computer records created and used during the administration of President George H. W. Bush.

PERPOS is also a software system that is being developed using a method known as evolutionary prototyping. An initial prototype was constructed to learn more about the problems of separating operating system and office application software files from user-created files and viewing personal computer (PC) files in legacy file formats. Once the prototype had been used in processing actual PC files from the White House Offices and the requisite knowledge gained, the prototype was adapted to satisfy the now better understood needs.

Archivists who used PERPOS learned that there were some files that could not be viewed. The files included password protected or encrypted files, damaged files, and files of an obsolete format for which there was no viewer. The prototype was extended to include the capabilities to recover passwords from protected or encrypted files, to use the recovered passwords to decrypt files, to repair damaged files, and to convert obsolete file formats to current or standard formats for which there was a viewer. Then the prototype was used again to process Bush PC e-records, more was learned, and the prototype was re-adapted based on archivist's recommendations.

Five years after the end of the Bush Presidential Administration, Archivists at the Bush Presidential Library began to respond to Freedom of Information Act (FOIA) requests, and that continues today. This requires the capability to index and search the repository of Presidential e-records for e-records relevant to the request. It also requires the capability support review of the relevant records including redaction of copies of the records. In addition, finding aids must be created for FOIA collections, and the collections must be made available to the requestor. The prototype was extended to include these capabilities and Archivists at the Bush Library and Archives II will soon be using these new features.

This process of prototype use, learning and re-adaptation repeats until the prototype system satisfies all needs and has thus evolved into a system. Research is ongoing to provide advanced technologies to support archivists in review of electronic records, automatically recognizing record types, metadata extraction, and automatic summarization of record series.
We want to acknowledge the contributions of Archivists at the Bush Presidential Library and the Virtual Laboratory at Archives II who learned from the use of the prototype and through their comments facilitated the evolution of the prototype. These archivists include Bill Harris, Sam McClure, Laura Spencer, Stephanie Oriabure, Debbie Carter, and Brooke Clement. We appreciate their continuing partnership and contributions.

William Underwood, Principal Investigator
Email: William.underwood@gtri.gatech.edu
TABLE OF CONTENTS

1. INTRODUCTION.................................................................................................................. 1

1.1 Describing the Arrangement of Holdings by Provenance............................................ 1

1.2 Accession...................................................................................................................... 4
  1.2.1 Associating Containers with an Accession and Saving Them in the Repository.......... 8
  1.2.2 Detecting Computer Viruses in Accessioned Containers........................................ 16

2. SYSTEMATIC PROCESSING ......................................................................................... 17

2.1 Systematic Processing Dataflow.................................................................................. 17

2.2 Systematic Case Management..................................................................................... 20
  2.2.1 Create a Systematic Case...................................................................................... 20
  2.2.2 Checkout a Container......................................................................................... 23

2.3 Exploring a File System.............................................................................................. 25
  2.3.1 Choosing a Processing Activity............................................................................ 26
  2.3.2 Opening a Container......................................................................................... 27
  2.3.3 View a Container's Properties.......................................................................... 29
  2.3.4 Viewing a File’s Properties............................................................................... 30
  2.3.5 Viewing a File with Quick View Plus................................................................. 33
    2.3.5.1 Quick View Plus Options............................................................................ 34
    2.3.5.2 Viewing Document Files ............................................................................ 35
    2.3.5.3 Viewing Archive Files............................................................................... 35
    2.3.5.4 Viewing Bitmap or Drawing Files............................................................... 36
    2.3.5.5 Viewing Spreadsheet, Presentation and Database Files............................ 37
    2.3.5.6 Viewing File Types that are Unknown to Quick View Plus......................... 37
    2.3.5.7 Exiting Quick View Plus............................................................................. 39
  2.3.6 Viewing the Manifest of a TAR Archive............................................................... 39
  2.3.7 Saving the Results of our Work........................................................................... 40
  2.3.8 Closing a Container............................................................................................. 40
  2.3.9 Help..................................................................................................................... 40
  2.3.10 Processing Status.............................................................................................. 49
  2.3.11 Exiting the APT................................................................................................. 52

2.4 Filtering the Files in a Container............................................................................... 52
  2.4.1 The Filtering Menu............................................................................................ 53
  2.4.2 Filter on OS/App Hash Code and File Type......................................................... 54
    2.4.2.1 Loading the Filter....................................................................................... 54
    2.4.2.2 View the OS/App Hash Code Filter............................................................ 56
    2.4.2.3 View a Defined File Type Filter.................................................................... 57
    2.4.2.4 Filter on OS-App Hash Code & File Type.................................................... 58
    2.4.2.5 Manually Filtering the Files Passed through the Filter............................... 59
    2.4.2.6 Extract Archived Files............................................................................... 60
    2.4.2.7 Transferring the Blocked Files.................................................................... 62
    2.4.2.8 Saving the User-Created Files................................................................. 72
    2.4.2.9 Exiting the APT from the Filtering Activity............................................... 74
  2.4.3 Filter Manually..................................................................................................... 75
    2.4.3.1 Creating an OS/App Hash Code Filter......................................................... 75
    2.4.3.2 Cancel......................................................................................................... 80
    2.4.3.3 Transferring the Blocked (OS/App) Files..................................................... 80
    2.4.3.4 Saving your work....................................................................................... 82
2.4.3.5 Saving a Filter ................................................................. 83
2.4.4 Filter on OS/App Hash Code ....................................................... 83
2.4.4.1 Loading a Filter ................................................................. 83
2.4.4.2 View OS/App Hash Code Filter .............................................. 84
2.4.4.2 Filtering on OS/App Hash Code ............................................ 85
2.4.4.3 Viewing Types of Blocked Files................................................ 86
2.4.4.4 Manually Filtering the Files Passed through the OS-APP Hash Code Filter ........................................... 87
2.4.5 Filter on File Type ................................................................ 87
2.4.5.1 Define a File Type Filter ..................................................... 92
2.4.5.2 Filter on File Type ......................................................... 94
2.4.5.3 Manually Filtering the Files Passed through the File Type Filter .................................................. 94
2.5 ARRANGING FILES IN A CONTAINER ......................................................... 95
2.5.1 The Arrangement Menu ....................................................... 95
2.5.2 Moving a File into a Folder .................................................... 96
2.5.3 Reordering the Files in a Directory ......................................... 98
2.5.4 Viewing the Properties of a Folder or File ................................ 99
2.5.5 Extending a Folder Title ....................................................... 100
2.5.6 Creating New Folders .......................................................... 101
2.6 PRESERVING ELECTRONIC RECORDS .................................................... 103
2.6.1 The Preservation Activity Menu ............................................ 104
2.6.2 Recovery of Passwords and Decryption of Password Encrypted Files ...................................................... 105
2.6.2.1 Recovery of Passwords .................................................... 105
2.6.2.2 Decryption of Files ......................................................... 107
2.6.3 Repair Corrupted Files ........................................................... 112
2.6.4 Convert Files in Obsolete Formats to Current or Standard Formats ...................................................... 116
2.6.5 Associate File Type with Application ........................................ 120
2.6.5.1 Associating a File Type with a Password Recovery Utility ........................................ 121
2.6.5.2 Associating a File Type with a Decrypter ................................ 124
2.6.5.3 Associating a File Type with a Repair Utility .......................... 127
2.6.5.4 Associating a File Type with a Conversion Utility .................... 130
2.7 REVIEWING FILES IN A CONTAINER ..................................................... 134
2.7.1 The Review Menu ................................................................. 134
2.7.2 Reviewing a Record ............................................................... 136
2.7.2.1 Opening a File to Public Access ......................................... 136
2.7.2.2 Withdrawing a Record due to Restrictions on Disclosure ......... 137
2.7.2.3 Redacting Terms or Passages with Restrictions on Disclosure .... 143
2.7.2.3.1 Redacting Document Images .......................................... 144
2.7.2.3.2 Redacting PDF Copies of Documents ............................. 149
2.7.2.3.3 Redacting MS Word Documents .................................. 151
2.7.3 Marking for Transfer ............................................................. 153
2.7.4 Mark a Record as a Personal Record Misfile ......................... 154
2.7.5 Reversing Review Actions by Unmarking ................................. 155
2.7.6 Changing a Review Action to Another Review Action ............... 156
2.7.7 Viewing File Properties ......................................................... 157
2.7.8 Saving a Reviewed Container .................................................. 158
2.7.9 Updating Previous Review Decisions ....................................... 159
2.7.9.1 Opening to Public Access a Previously Withdrawn Record .... 159
2.7.9.2 Redacting a Previously Withdrawn Record ............................ 159
2.7.9.3 Re-redacting a Previously Redacted Record ......................... 159
2.7.10 Edit Reasons Withdrawn ....................................................... 160
2.7.11 Edit Document Type List ..................................................... 162
2.7.12 Checking for Access Restrictions ........................................... 162
2.8 DESCRIBING A RECORD SERIES ......................................................... 163
2.8.1 Moving a Container to a Different Record Series ...................... 169
1. Introduction

PERPOS is a software system that supports accessioning e-records into an electronic records repository and systematic processing of these records. Systematic processing includes arrangement, preservation, review, and description of record series. In addition, PERPOS supports search for records related to a FOIA request, review of those records and creation of a FOIA collection and finding aid.

There are two major subsystems of PERPOS—the Archival Repository Tool (ART) and the Archival Processing Tool (APT). ART, the computer platform on which it operates, and accessioned holdings make up the electronic record repository. The APT and the computer platform on which it operates make up the archival processing subsystem.

ART and the APT are accessed through two icons placed on the archivist's desktop when PERPOS is installed.

There are certain activities that need to be performed before either Systematic Processing or FOIA Processing can be supported with these tools. These are described in the following sections. In section 2, the use of ART and the APT to support systematic processing of e-records is described. In section 3, the use of ART and the APT for FOIA processing is described.

1.1 Describing the Arrangement of Holdings by Provenance

Archival arrangement is "the organizing of materials with respect to their provenance and original order to protect their context and to achieve physical or intellectual control over the materials [Pearce-Moses 2005]. Archival description is the activity of providing information about the archived records, their origin and context, and any archival actions. Information about archived records includes information about their intellectual content, intellectual access, physical description, and physical access [Miller 1990]. Access here refers to any information the archivist or researcher needs to access a specific set of records."
This type of information includes a description of the arrangement of records. This section describes how ART supports description of the arrangement of holdings. Section 2.8, Describing a Record Series, discusses the support for describing the intellectual content and physical description of records.

The Bush Presidential Library arranges its holdings by collection, office, series, folder, and item. To describe the arrangement of Library holdings, select the Archival Repository Tool Icon and select Description from the Activity drop-down menu (also called a pull-down menu).

Then select Add Collection Name from the Edit drop-down menu.

The Add collection dialog box will be displayed.

Whenever a dialog box for data entry appears, all required fields will have a red star beside them and the OK button will be disabled. As valid information is entered into each field, the star beside it will turn black. Whenever invalid
information is entered into any field whether it is required or not the contents of the field turn red and the OK button is disabled. If the field also has a star, the star will turn red as well. Whenever the cursor placed over a field whose contents are red and left there for a time, a message appears describing the error. The OK button is enabled after all required fields have been entered and contain valid data. At any time, the Cancel button can be selected.

The Bush Presidential Library has standardized names for its collections of Presidential and Vice Presidential records, for example,

- Bush Presidential Records: Staff and Office Files
- Quayle Vice Presidential Records: Staff Files

Enter a unique number for each collection (or Record Group), the Name of the collection, and then select the OK button. The collection name will appear in the left pane of the ART window and the entered information in the right pane.

Other collections can be added at any time. Collections can only be added in the Description Activity. With a collection name highlighted, one can edit the name to correct spelling errors or changes in the standard name by selecting properties from the Edit drop-down menu. One can also add the names of offices included in that collection by selecting Add Office Name from the Edit drop-down menu, while the collection name is highlighted.
This brings up the Add Office Name dialog. The Office Name must be unique within its collection. If the Office Name is not unique, the entry and its star will turn red and the OK button will be disabled. A box containing the message “Duplicate Name or Title” will appear when the cursor hovers over the entry.

The Library also has standardized the naming of Offices in the Executive Office of the President. When Ok is selected, the Office name will be added to the collection.

Office names can be added at any time. They may also be added during accession.

1.2 Accession

Accessioning is the activity that an archives performs to gain initial physical, intellectual and administrative control over newly acquired records. To accession electronic records, one uses the Archival Repository Tool and selects the Add Accession from the Activity drop-down menu.
The Accession Number entered is the 4-digit current year and the 4-digit next sequential number, separated by a period. Accretion number is a 2-digit number assigned when a subsequent accession should actually be part of a previous accession.

The Date of Receipt is the date the records were received by NARA or the Bush Presidential Library. The Date Logged In is the date the archivist is filling out the accession register entry, and Logged by is the name of the archivist. The name can be selected from a list of archivist's names or entered.
The next section, *Donor Information*, is for Gifts of Personal Records donated to the Presidential Library. The description of these fields will be skipped because we are currently concerned with Bush PC Files. Some of Bush PC files may be determined later to be personal records that will be offered back to their creator. However, if the person who created the personal electronic records donates them to the Library, they would be accessioned and this donor information would be filled in. For one reason or another, the Bush Library Accession Form requires an entry in the Donor field. For Presidential or Federal records, one can enter N.A. for Not Applicable.

The next section, *Accession Information*, has as its first field, *Identification*. This is a tentative title of the Collection being accessioned, e.g., Bush Presidential Electronic Records, or Quayle Vice Presidential Electronic Records. The *Brief Description* is a description of the record series, for example, a description of the responsibilities or activities of the Office or Person who created, received, or used the records and the document types, e.g., Staff and Office Personal Computer Files. The content of these two fields is determined from a Records Transmittal Form or other documents associated with the transfer.

The next step is to indicate the beginning and ending dates of the files contained in the accession. Next, one indicates the approximate volume. For paper records, this is expressed in linear feet of paper. For the Bush hard drives, it might be expressed as the number of hard drives.
If one wants to enter to determine the *Inclusive Dates* of files in the file system and the *Approx. Volume*, of electronic records in all the containers, minimize the Archival Repository Tool Window and select the Archival Processing Tool icon. Open the CD containing the containers and select the Container Properties option from the View drop-down menu. You will see a display similar to the following.

The approximate volume of paper records is measured in linear feet. For electronic records, volume is measured in files or bytes (or kilobytes, megabytes, gigabytes). So, one might enter as the volume in the accession 63 files. However, one could enter the volume in bytes. It is recommended that one enter the abbreviation of the unit (ft, files, or bytes, KB, MB, GB). While the exact number of bytes in all the containers is 507134, it is recommended that bytes only be used if the number of bytes is less than 1 kilobyte, and that the volume be approximate, not exact.

1 Kilobyte = 1024 bytes  
1 Megabyte = 1024 Kbytes = 1048576 bytes  
1 Gigabyte = 1024 Mbytes = 1073741824 bytes

Therefore, in this example, what would be entered is 495.2 MB, since

\[507134 \text{ bytes} \approx 495.2 \text{ Kb}\]

In the Restrictions section, the archivist checks whether the records are likely to be subject to Presidential Record Act (PRA) restrictions and/or Freedom of Information Act (FOIA) exemptions. This is dependent on who created the records. For instance, records of the department of Treasury are Federal Records, which are only subject to FOIA exemptions. Records of the Staff
Members and Offices of the Executive Office of the President are Presidential Records and subject to both PRA restrictions and FOIA exemptions. *Deed of Gift* restrictions are dependent on restrictions placed on access to the records determined by the Donor.

The only other section of the form that is relevant to accession of the Bush PC files is the Additional Information section. In it, one might enter information about the reason for accessioning these records, since they were not originally scheduled for transfer to the National Archives.

One may print the form by selecting the *Print* button at the bottom of the form. When the OK button at the bottom of the form is selected, the information is added to the Accession Register, the accession number appears in the left pane of the Archival Repository Tool and the Accession information appears in the right pane.

If one needs to edit any information in the accession, you highlight the accession number, and select Properties from the Edit drop-down menu. The form containing the information entered is displayed where it can be edited and saved by selecting OK.

1.2.1 Associating Containers with an Accession and Saving Them in the Repository

Electronic Records are transmitted in some type of container, for example, a CD, a DVD, a directory of a CD, an archive file on a CD, a tape, or diskette and some type of identifier for the container. To associate containers with an
accession, highlight the accession number and select add container from Edit drop-down menu. A dialog box will be displayed. The Administrator sets the default source of accessions using Storage Management on the Options drop-down menu of the APT. The dialog box below shows that the Administrator has set the default to the CD drive.

If that is not the source, the archivist can browse to identify the actual source.
When the CD drive is selected, one finds the folders (containers) that contain the actual records. They are labeled 0001-0014.

When the National Archives received the Bush hard drives, they contracted to have them transferred to removable media. The contractor placed the contents of each hard drive in a directory. Hence, ART and the APT were designed to accept the contents of folders (directories) as containers. They also accept tar files as containers, and the content of a device such as a floppy diskette or CD as a container.

A records transmittal or Inventory list should indicate the provenance of the records in the container. The figure below is an example of a records inventory for sample records that will be used in explaining associating containers with an accession.

<table>
<thead>
<tr>
<th>Office Name</th>
<th>Person’s Name</th>
<th>Directory</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Lady’s Office</td>
<td>Barbara Bush</td>
<td>0001</td>
</tr>
<tr>
<td>Counsel’s Office</td>
<td>Lee Liberman</td>
<td>0002</td>
</tr>
<tr>
<td>Office of the Press Secretary</td>
<td>Unknown</td>
<td>0003</td>
</tr>
<tr>
<td>Office of the President</td>
<td>George H. W. Bush</td>
<td>0004</td>
</tr>
<tr>
<td>Office of Appointments and Scheduling</td>
<td>Unknown</td>
<td>0006</td>
</tr>
<tr>
<td>Council of Economic Advisors</td>
<td>Michael J. Boskin</td>
<td>0007</td>
</tr>
<tr>
<td>Counsel’s Office</td>
<td>C. Boyden Gray</td>
<td>0008</td>
</tr>
<tr>
<td>Office of Policy Development</td>
<td>Roger B. Porter</td>
<td>0009</td>
</tr>
<tr>
<td>Press Office</td>
<td>Unknown</td>
<td>0011</td>
</tr>
<tr>
<td>Public Liaison</td>
<td>Doug Wead</td>
<td>0012</td>
</tr>
<tr>
<td>VP Office</td>
<td>William Kristol</td>
<td>0013</td>
</tr>
<tr>
<td>National Security Council</td>
<td>Unknown</td>
<td>0014</td>
</tr>
</tbody>
</table>

Suppose that we select container 0001 by highlighting it and clicking on OK. The following dialog box would be displayed. The archivist must enter an OAID number, select or enter a new office name that is not yet in the Catalog of Holdings, and select a series title that is in the Catalog of Holdings, and indicate whether this records in this container are likely to have any types of Restrictions.
The Record Transmittal aids in determining this information. In addition, the contents of the container are loaded into the Explore Activity of the Archival Processing Tool so that the Archivist can browse the files if necessary. The stars to the right of some of the fields indicate required fields. Red stars indicate information that is still missing. The OK button will remain disabled or grayed out until all required fields have a valid value.

The figure above shows the folder and filenames in the directory (container) 0001. If one double clicks a filename, a viewer will display the contents of the file. For instance, the figure below shows the contents of filename 001.txt. From viewing this file and the other files, and from the abbreviated label of the folder containing these files the archivist might conclude that the series title that this container belongs to is "First Lady's Correspondence."
The Record Transmittal and the APT Window can be tiled or cascaded so that they can be seen at the same time or so that it is easy to go from one to the other. One does this by placing the mouse pointer on the status bar at the bottom of the desktop, pressing the right mouse button, and selecting one of the options: Cascade Windows, Tile Windows Horizontally, or Tile Windows Vertically.

You can also switch between open windows by clicking on the window's button on the task bar shown at the bottom of the screen. You can also switch to the last open window by pressing Alt Tab.

The figure below shows the New Container dialog box in which the OAID number has been entered. The OAID number field is a five-digit ID number assigned by Bush Library staff.

The Record Transmittal indicates that this directory contains files from the First Lady's Office. This is an office in the Executive Office of the President, so these
records belong in the collection of "Bush Presidential Records: Staff and Office Files." Selecting the downward pointing arrow at the left of the Office Name field displays Office names that have been entered into the Catalog of Holdings. Selecting the entry, "First Lady, Office of, Correspondence," results in that office name being entered into the Office Name field. If the office had not been in the list, the archivist could have selected the entry "New Office," entered the name of the office and the new office name would be entered into the office names of the Bush Presidential Records in the Catalog of Holdings.

From viewing the records, the archivist concluded that these records were correspondence. Since "First Lady’s Correspondence" is not yet a series within that office, the archivist can enter a new series title and it will be entered in the list of series titles associated with the “First Lady, Office of, Correspondence”. It is preferred that Series Titles end in the word Files only when they are associated with a person. “Lee Liberman’s Files” is an example of a Series Title associated Lee Liberman as the person. This is an inference that ART make automatically. Whenever the Series Title does not end in the word “Files” ART infers that the person associated with the Series is “Staff Person Unknown.” In the case of FOIA Processing, electronic records are sorted by the name of the person with whom they are associated.
As shown above, the OK button is enabled as soon as all required fields have been entered. When the OK button is selected, the container is saved as a tar file in the Archival Repository, the Archival Processing Tool used for viewing the container is closed, and one is returned to the Archival Repository Tool where one sees that the Container with OAID 18000 has been associated with the Accession.
The information shown in the right pane of the window is the information entered in the New Container dialog box, its status is Unprocessed, the Media type HD (Hard Disk), and the volume on which it is stored.

The Archivist goes through a similar process to add the remaining containers associated with the accession. At any time, one can open a container associated with an accession as read only. For instance, if the container name is highlighted, one selects Open from the File drop-down menu. The Archival Processing Tool appears and the contents of the container are displayed.

The container has a manifest associated with it that shows the file path to each file, a SHA1 hash code used to check the integrity of the file, and a description of the file type. The manifest can be displayed by selecting Manifest from the View drop-down menu of APT. The figure below shows the manifest for container 15001.

```
Manifest Version: 5.0
Accession: 2005.0001
OAIM: 15001
Original Name: C001
Container Name: 15001.tar
Collection: Bush Presidential Records: Staff and Office Files
Office: First Lady, Office of, Correspondence
Series: First Lady's Correspondence
Processing Status:

Name: C001\itemID: 1
FileType: Directory

Name: C001\txt itemID: 2
SHA-Digest: 5FA4FA3AE9468073EC1DA9C1D68FA592C2052A96
FileType: ASCII 7-bit Text

Name: C001\txt itemID: 3
SHA-Digest: B0B32A10DF30C2AD63C08A4BA02362CD58AC71
FileType: ASCII 7-bit Text

Name: C001\txt itemID: 4
SHA-Digest: A4B65A33D10BAA06FE5923C75FAD13F29F961
FileType: ASCII 7-bit Text

Name: C001\txt itemID: 5
SHA-Digest: F6861F9AB39E267C860DD4C3A5C21CC639B59
FileType: ASCII 7-bit Text
```
1.2.2 Detecting Computer Viruses in Accessioned Containers

Records from PC file systems, E-mail systems and Records Management Applications should have been checked for computer viruses before the Presidential Library acquires them or before they are transferred to the National Archives. However, electronic records accessioned must be again checked for viruses to verify that the electronic records do not include viruses. Failing to do so can result in infection of other files in the Archives, and might result in reintroduction of the virus into the wild later when files are distributed to researchers.

Consequently, any record series read from floppy or CD-ROM is automatically checked for computer viruses. Furthermore, any containers of electronic records that are transferred over the Internet should be stored on a separate hard disk or other medium different from those containing accessioned, in process, or processed electronic record series.

When the APT and ART are installed, a copy of McAfee VirusScan is also installed. On-access scanning should be enabled. This means that the scanner will check for viruses in files that you access, that is, open, copy, save or otherwise modify and files that you read from or write to floppy disks, CD-ROMS, DVDs or network drives. It can render many of the viruses harmless, or remove and quarantine the virus-infected file.

To ensure that On-Access scanning is enabled, move the mouse pointer over the McAfee VirusScan Icon shown at the right side of the taskbar. If VirusScan On-Access Scanning is enabled, a message will appear stating so. If not, with the mouse button on the VirusScan icon, click and right mouse button and select “Enable on-access Scan.”
2. Systematic Processing

Systematic processing is the arrangement, preservation, review, and description of all the records in a record series. Systematic processing begins when the Supervisory Archivist assigns the processing of a record series to an Archivist. The archivist receiving the assignment starts work by creating a Systematic Case. Once a Systematic Case has been created, the archivist uses ART to access the Catalog and find the containers associated with the record series they are to process. They then add each container to the Systematic Case. To process one of the containers in a Systematic case, the archivist checks out the container. Once the container has been checked out, the archivist extracts any files that are part of an archive file, does any necessary filtering, arrangement, or preservation. The archivist then reviews each record, checking it for FOIA or PRA restrictions. They then check the container back into holdings and move on to the next container. When they have finished processing all containers in the record series, the archivist describes the record series, and then they transfer any Personal Record Misfiles (PRMs) in the containers to the PRM directory. In addition, they transfer to the Library any other records in a container that are marked for transfer. Finally, they create a reference copy of each container, which are saved in the Reference directory.

2.1 Systematic Processing Dataflow

The diagram below illustrates the dataflow of Systematic Processing activities that are supported by the PERPOS tools. The numbered, labeled circles represent systematic processing activities supported by the tools. The labeled parallel lines represent the kinds of information that are created, stored by the activities. The labels on directed edges represent the kinds of information that are inputs or outputs of an activity or stored as a result of activities and subsequently retrieved for use by other activities. The rectangles represent entities external to the PERPOS system. Stepping through the diagram in the numerical sequence of the activities, one sees the dataflow.
Archivists at the Presidential Library must first accession the electronic records (Activity 1). This activity, Accession Records, is described in the first section of this Manual. It is not a part of Systematic Processing. Transferred files are loaded from storage devices, such as floppy and compact disk drives or file transfer areas. An entry is made in an accession register, and files that are associated with the accession are stored placed into containers and stored in the Archival Repository (Holdings). Information about the accessioned containers, such as the name of the collection, office name, and record series title is also entered into the Catalog of Holdings. During accession, the archivist can open a container and browse its contents to determine this information.

The first activity in the Systematic Processing of accessioned electronic record series is for the Supervisory Archivist to schedule archival work. This involves loading the Accession register and or Catalog of Holdings to determine which containers remain to be processed, and assigning work to archivists.

Archivists start work by using ART to create a Systematic Case (Activity 2). Once a Systematic Case has been created, the archivist uses ART to access the Catalog and find the containers associated with the record series they are to process. They then add each container to the Systematic Case they just created (Activity 3). To process one of the containers in a Systematic case the archivist Checks Out the container (Activity 4). This leaves the original container
in the holdings area as a backup and places a copy of the container in the archivist’s Work Area. Information about the type of process being performed and who has the container checked out is stored, and is displayed as properties of the container.

A peculiarity of the Bush PC record series is that they include the entire file system of a personal computer—operating system and application files as well as user-created files. The APT supports filtering file systems by blocking operating system and applications files and passing through user-created files (Activity 5). This activity could be considered a preservation function as it involves separating records and non-records. After each step of work, the archivist saves their work back to the work area.

Archivists should attempt to maintain the original order of files in a file system, but some files may not have been stored in the proper directory (folder). For instance, some word processing files that should have been stored in a Correspondence directory may have been stored in the root directory or in the directory including the word processing application. Archivists may need to perfect the arrangement by moving misplaced files into the proper directory. This activity is called archival arrangement (Activity 6).

Some PC files may be in obsolete or proprietary file formats that can no longer be viewed. Other files may be corrupted due to media deterioration or file transmission errors. Other files may be encrypted, so that there is a need to recover a password and decrypt the file so that it can be viewed. These activities are referred to as archival preservation (Activity 7).

Next, PC records must be reviewed for Freedom of Information Act (FOIA) exemptions on their disclosure to the public. They must also be reviewed for Presidential Record Act (PRA) restrictions on their disclosure (Activity 8). During this activity, Presidential records are opened, closed, or redacted.

During review, an archivist may discover non-records such as software application documentation or sample files for application that were not removed during filtering. These can be marked for transfer to the Library. They may also discover personal records that were misfiled with the Presidential records. These can be marked as Personal Record Misfiles (PRMs) that can later be removed from the containers.

When the contents of a container have been filtered, arranged, preserved, and reviewed, the container of processed records is checked back into Holdings (Activity 9). This results in the processed container replacing the original container in the holdings area.
When archivists have completed the preceding activities, they must describe the record series (Activity 10). This involves loading the containers in the record series and then viewing and describing their contents, and determining the extent (number of files, number of bytes, or number of pages) of the processed container.

Non-records that are in a container and that were marked for transfer to the Library can be removed at this time (Activity 11). Personal records that were marked as PRMs should also be removed from containers. Support for this activity has not yet been implemented.

Since the master copy that is stored in archival storage may contain records whose access is restricted in whole or in part, it is necessary to create a Reference Copy that includes just those records that are open to the public or redacted versions of records. To do this, an archivist creates a Reference Copy of the series containers that can be made available to the Public Access System (Activity 12).

Archivists can access the containers of processed record series to re-review those closed files or originals of redacted documents when access restrictions have expired. They can also access containers in Holdings for preservation actions, such as converting to new file formats when current file formats become obsolete.

### 2.2 Systematic Case Management

The first activity in systematic processing of accessioned electronic record series is for the Supervisory Archivist to determine which series and which containers need processing. The Supervisory Archivist then assigns them to an archivist for processing. To do this the Supervisory Archivist does the following:

1. Select ART and select the Description Activity. The left pane will show the collections, offices within collections, record series within offices, and containers within record series.
2. Identify the series of records to be systematically processed and the containers within the series.
3. Assign one or more containers to an archivist.

#### 2.2.1 Create a Systematic Case

The next activity in systematic processing is to create a systematic case and add the containers to the systematic case.
1. The archivist selects *Systematic Case Mgmt* from the ART Activity drop-down menu.

![Diagram of Archival Repository Tool with Systematic Case Mgmt selected]

2. The archivist then selects *Add Systematic Case* from the *Edit* menu.

![Diagram of Archival Repository Tool with Add Systematic Case selected]

3. A dialog box will be displayed in which one enters the next systematic case processing number. The user id of the archivist who will be assigned this case will be automatically entered from the logon user id. Select the OK button.

![Add Systematic Case dialog box]
4. The Systematic Case Management window will show a new systematic case number in the left pane and the Archivist user id in the right pane.

![Systematic Case Management Window](image)

5. Select Add container from the Edit drop-down menu while a Systematic Case is highlighted. A dialog box for entering the container id will be displayed. Enter the container id and select the OK button.

![Add Container Dialog Box](image)

The container id will be shown as a branch of the systematic case number in the left pane and associated information such as collection name, office name, and series name will be displayed in the right pane.
6. Repeat this process for the other containers to be systematically processed.

2.2.2 Checkout a Container

Before any processing can be performed on a container, a copy of it must be checked out of Holdings. Containers can only be checked out for processing under either the Systematic Case Mngmt Activity or the FOIA Case Mngmt Activity. Since the archivist is performing systematic processing, the archivist selects the Systematic Case Mngmt Activity of ART. Once in the Systematic Case Mngmt Activity, the archivist selects a container to process.

1. Highlight the filename of the container to be processed. Select Checkout Container from the File drop-down menu.
2. Indicate the location of the work area to place the checked out container. This places a copy of the container in the indicated directory while leaving the original container as a backup. The copy of the container in the work area contains information about the type of processing being done, the archivist who checked it out, and the case no under which it was checked out. This information is saved in the manifest and is used to prevent any archivist other than the administrator or the archivist who checked it out from processing the container. This information is also stored in the repository database.

When the archivist selects the Save button, the archivist is returned to Systematic Case Mngmt screen and the containers properties have changed to
show the Processing Archivist, the Processing Type, and the Case No that the under which the container was checked out. Once a container has been checked out, its Status is changed to in process. This allows other archivists to know that the container is no longer available for processing until the processing archivist or the administrator checks the container back into holdings.

The archivist is now ready to process the contents of the container. This consists of arrangement, preservation, review, and description. In addition, they may need to filter operating system and software applications from the file system in the container. After each step of work, the archivist saves their work back to the work area.

### 2.3 Exploring a File System

To perform each of the systematic processing activities, the archivist must have first opened the container in the Archival Processing Tool (APT). There are two ways to open the Archival Processing Tool. It can be opened as a standalone application or it can be opened from within the Archival Repository Tool. If the container is checked out, the archivist that checked it out or the administrator can open it in the work area with all processing activities enabled. If the container has not been checked out or the current user is not the archivist that checked the container out, the container that is in the Holdings area will be open as read only with only the Explorer Activity enabled.

When working with the Archival Processing Tool, there are a few operations that you will perform regardless of whether you are filtering, arranging, preserving, or reviewing records. Here are the operations you will perform frequently.
You will choose an **APT activity**. These activities are Filtering, Arrangement, Preservation, and Review.

You will **open a container** to perform every APT activity. Though containers can be directories or TAR files, all accessioned containers are tar files.

You will also **view an individual file's properties**. File properties give you information about the file size, the message digest (SHA1 hash code) of the file (this is a unique number calculated from the contents of a file for purpose of checking data integrity. If the contents change, it changes), the date the file was last modified and the file format type (WordPerfect 4.2, Excel 2.0 Spreadsheet, etc.). While you may not need to examine every record's properties, it will be important to know how to do so.

You will frequently **open and view files**.

You will save the results of your work by **saving a container** as a TAR file with manifest.

You will **close a container** in order to open another container.

### 2.3.1 Choosing a Processing Activity

There are two ways to choose a processing activity. First, from the main drop-down menu, select Activity. The five activities of the APT appear on the menu. Select the activity you want to perform.

Secondly, with one mouse click, select the activity you want to perform from the Activity Toolbar. Below the main drop-down menu, you will notice a row of buttons with various images, or icons. This is referred to as the toolbar. The functions available on the toolbar are also available on the drop-down menu. As the toolbar requires just one click, it is slightly quicker than selecting an item on the menu bar and then selecting an item from the drop-down menu. The
challenge is learning which button corresponds with which activity. From left to right, the activities are: Explore, Filtering, Arrangement, Preservation, and Review. You can always confirm the activity as it is written in text to the right of the toolbar.

The container path is in the Title bar and the Systematic Case No and the processing activity are shown in the Toolbar. If there is not a container currently open, then just the Activity is listed in the Toolbar.

2.3.2 Opening a Container

After checking out a container, one selects the container to be processed by opening it. There are two ways to open a container. You can open a container from inside ART by selecting the container and then selecting Open Container from the File drop-down menu.

You can open a container from inside the APT by selecting Open from the File drop-down menu.
The following dialog box appears.

![Archival Processing Tool](image)

The default directory is whatever the Supervisory Archivist set as the “Working Storage” location. The archivist selects their personal directory in this location, and then selects the container they have checked out.

In the following example, the accessioned and checked out container is used to test the accuracy of the file type identifier. Once the container has been opened, it becomes obvious that it contains operating system files in the “DOS” directory and some application files in the “WP” and “DBASE” directories. This container also has at least one archive file. This file is in the root directory and it and any other archive files should be extracted before filtering. This is because it could contain other files in need of filtering.
If you highlight a filename, the file’s properties will be displayed in the right windowpane. Some of these properties are shown in the figure above. If you highlight a folder title, the names of files and folders contained in that folder will be shown in the right windowpane. In addition, the date last modified and length of the files and folders will be shown.

### 2.3.3 View a Container’s Properties

When accessioning a file system, a container is created. Viewing the container’s properties can be used to determine the size of the original file system. To get the true number of folders, files and bytes it is necessary to extract all archived files. Extracting archived files will be covered later in section 2.4.5.4. To view the properties of a container, perform the following steps.

1. **Step 1:** Select the View Drop-down Menu
2. **Step 2:** Select Container Properties.

A dialog box will appear that shows the **Accession No**, and the **Accretion** when applicable, under which that the container was accessioned. The **Earliest File Date**, the **Latest File Date**, the **Total Number of Folders**, the **Total Number of
Files, the Total Number of Bytes, and the Approximate Number of Bytes. Approximately shows the number of bytes rounded off to the number of kilobytes, megabytes, or gigabytes, whichever is the greatest that has whole number value.

1024 bytes = 1 kilobyte (Kb)
1024 kilobytes = 1 megabyte (Mb)
1024 megabytes = 1 gigabyte (Gb)

2.3.4 Viewing a File’s Properties
A file’s properties include its name, its current path in the container, its length, the date it was last modified, its file type, and its SHA-1 message digest. There are additional properties depending on the kind of processing that has been preformed. In the case of the Rearrangement, the file is given an original path property. In the case of the Preservation Activity, a file may have a set of preservation actions taken. In the case of the Review Activity, a file has an access property, and may have additional properties that are related to the kind of access granted. To view a file’s properties:

Step 1: Using your left mouse button, click the plus signs to open folders until you reach the file you wish to open.
Step 2: Highlight the file name and click it once. In the right plane, you will notice that the file’s properties are listed. You can confirm the file’s path, the date it was last modified, the file type and the SHA-1 message digest. This SHA-1 is a unique identifier calculated for the file by the APT.

There are other ways to view a file’s properties. Highlight the file’s name, and click the right mouse button. A menu will appear. Select Properties.
With the filename highlighted, select the properties icon from the toolbar or select Properties from the View drop-down menu.

However the archivist selects properties, a File Property dialog box similar to the following will appear.
The File Type Field of the File Properties box can be edited. This should only be done if the File Type Identifier has indicated that the File Type is *Unknown* and the user knows the File Type from another source, such as the list of File Extensions and Descriptions provided in the on-line Help.

### 2.3.5 Viewing a File with Quick View Plus

Due to the rapid life cycle of computer technology, hardware and software used to create documents and databases eventually become obsolete. The challenge this poses to archivists is how to view these legacy files without access to the software or hardware on which they were created. The APT uses Quick View Plus, a suite of over 225 software viewers to meet most of this requirement. However, there are some legacy file formats that it does not recognize and display. Hence, additional viewers are required.

The Quick View Plus User’s Guide discusses the features and options of Quick View Plus [JASC 2000]. Some of these features are described in this manual, especially those that relate to archival processing and interface to the APT.

To view a file:

Step 1: Using your left mouse button, click the plus signs to open folders until you reach the file you wish to open.

Step 2: To view the record, double click the file name. This action opens Quick View and allows you to view the record.
The document displayed is from the Bush Public papers. It is displayed with a file viewer, not the original application that was used to create it. The name of the original application and the file type is shown in the lower left-hand corner.

A second way to view a file is to highlight the file name and press the right mouse button. A menu will appear that contains an option Quick View. If you select this option, the file will be displayed with the Quick View Plus file viewer.

A third way to view a file is to highlight the file name and select Quick View from the View drop-down menu on the menu bar. The file will be displayed with the Quick View Plus file viewer.

2.3.5.1 Quick View Plus Options

The Quick View Plus Options should be as follows: In Quick View Plus select View > Options > General

1. Display Quick View Plus logo on startup should not be checked
2. Auto Should not be checked
3. New Views have a File Navigation Pane should not be checked.
4. New Views have a status bar should be checked.

In Quick View Plus select View > Options > Display

1. View Unknown files as should have either Text or Hexadecimal checked.
2. Character Set for Text Files should have either ASCII 7-bit or DOS (Latin) checked.

2.3.5.2 Viewing Document Files
Quick View Plus provides different options for different classes of file types. For document (word processing) files, Quick View Plus provides the capability to View a document in three modes—Draft, Normal, or Preview. The default is preview, which corresponds to what you would expect to be printed. It includes line wrapping, columns, and page margins; all fonts, character formatting, paragraph alignment, spacing, tabs and embedded objects; headers, footers and annotation.

Dear Mr. Speaker: (Dear Mr. Chairman)

Pursuant to Section 708 of Public Law 95-425 (1 U.S.C. 112b(b)), I transmit herewith a report prepared by the Department of State concerning international agreements.

Sincerely,
George Bush

Note: Identical letters were sent to Jim Wright, Speaker of the House of Representatives, and Claiborne Pell, chairman of the Senate Foreign Relations Committee.

The page size options are Full Size, Fit to Window, and Fit to Window Width. The default is Full Size. Either Full Size or Fit to Window Width is best for reading the text and for creating a document image when redaction is needed.

2.3.5.3 Viewing Archive Files
Quick View Plus can display the file structure, display the files and extract the files from a number archive files.
You can view the files by highlighting the file name and double clicking the left mouse button. It is recommended that you do not extract the files from these archives using Quick View Plus, as the APT recognizes a much larger set of archive file types and provides the capability to extract files from these archives using the Extract option on the Files drop-down menu.

2.3.5.4 Viewing Bitmap or Drawing Files
After viewing a bitmap file (BMP, GIF, TIP, JPEG) or drawing file (DRW, CHT, CGM, PIC), you can adjust the display to fill the entire screen, change the display size, or zoom in and out.
2.3.5.5 Viewing Spreadsheet, Presentation and Database Files

After viewing a spreadsheet (WKS, WB1, WQ1, XLS), presentation (SHW) or database files (DBF), you can enlarge or reduce the on-screen font size, show or hide the gridlines, and display any sheet in a multiple-sheet spreadsheet. The figure below shows Quick View Plus's view of a Dbase IV or V database (DBF).

Note that Quick View Plus does not display the values of memo fields that are associated stored in an associated memo file (FOXUSER.FPT). Consequently, the APT uses a different viewer for database files that have associated memo files that also displays the values of memo fields.

2.3.5.6 Viewing File Types that are Unknown to Quick View Plus

While Quick View Plus can view more than 225 file types, there are many files whose file type it cannot identify and for which it does not have an appropriate viewer. Such file types are said to be Unknown file types. Assuming that Quick View Plus has been configured to view Unknown file types as Text, the file ASHLAND.PLN from the PC Files Test Suite would be displayed as follows.
It is displayed in this form because Quick View Plus was configured to view file types that are Unknown as Text. The abbreviation WPC at the beginning indicates that the file was produced by a product of WordPerfect Corporation. However, by selecting View > View as > Hexadecimal, you can change the view to Hexadecimal.

The figure below shows the ASHLAND.PLN file represented in Hexadecimal notation.
The column at the right shows the text corresponding to the hexadecimal representation of the contents of the file. By scrolling through the rest of the file, one can see that the file seems to contain a 1988 Advertising and Sales Report for Ashton Tours and Travel New York Office.

There may be user-created files that cannot be recognized by Quick View Plus and that it cannot display properly. However, it has two other views that can give you some information about them.

The APT recognizes many file types that are not recognized by Quick View Plus, and other viewers can be provided for those file types. This feature will be discussed in a following section.

### 2.3.5.7 Exiting Quick View Plus

In the Apt, you should always exit Quick View Plus by either

- Selecting "Exit Quick View Plus" from the File drop-down menu, or
- Selecting the Quick View Plus icon (a magnifying glass) from the lower right part of the Task bar, right clicking the mouse, and selecting "Exit Quick View Plus."
Selecting "Close this View" from the File drop-down menu, or closing the view window using the X at the top right of the window, will leave Quick View Plus running in the background. This can cause an error message when you leave the APT and the APT is trying to erase temporary files that were provided to Quick View Plus.

### 2.3.3 How the APT Identifies File Types

MS-DOS and Windows file names use a 1-to-3-character file name extension to distinguish different file types, i.e., file format. However, file extensions alone are often not enough to discriminate file types. For instance, file extensions such as DOC are ambiguous, having two or more applications that create files with that extension but with different formats. Furthermore, there are WordPerfect document files that do not have the .DOC or .WPD extension recommended by the WordPerfect manual. Instead, the document creator avails himself of the 8-character filename plus 3-character filename extension to create a longer mnemonic filename. These extended names sometimes result in an extension used for another file type. For instance, SPEECH.COM a user-created WordPerfect document file from that Bush hard drives that contains an address (speech) to the Commonwealth Club. However, the .COM extension is also used to represent a MSDOS compressed executable file. The PRS file extension is customarily used to represent a WordPerfect printer resource file. A file named INVSALES.PRS from the Bush hard drives is a user-created WordPerfect file concerning the ratio of inventories to sales in a previous recession. ARC is a file extension used to indicate an archived file. However, there is a WordPerfect document file with the name CLEAR.ARC, which contains a memo concerning clearing archivists to enter the White House Offices to aid in transferring records to the National Archives.

Under the Unix operating system, the operating system and application programs distinguish between types of files by looking for a magic number or a fixed pattern in the file header. Microsoft and developers of software applications for DOS and Windows included magic numbers in the headers of some files that are used to distinguish file types.

Macintosh file types are also distinguished by having a signature encoded in the file. This consists of two four-letter codes that signify the application that created it and the format of the file. These codes are registered with Apple. Currently, there is not a central registry for file name extensions or magic numbers.

Some file extensions do not determine the file format but the documentary form or function of the file. For instance DOS batch command text files (file extension .BAT) are in IBM extended ASCII file format, but what makes it a batch command file is the commands to the DOS operating system that it contains.
Similarly, WordPerfect Notebook 3.0 files (file extension .NB) are in WordPerfect 5.1 file format. It is the form of the document that makes it a Notebook 3.0 document.

Some binary files do not have a magic number or fixed file pattern that can be used to determine its file type, e.g., files with filename extension SND or AU are raw unsigned Pulse Code Modulation sound files, which can be played through most sound devices without further manipulation. However, there is no magic number or file header, so one must resort to using the file name extension and context of the file to determine its file type.

Some software manufacturers do not publish the specifications for the file formats of their software applications. For example, IBM did not publish the file format specifications for IBM's DisplayWrite 4 documents. The format is proprietary.

A File Type Identifier has been developed to identify file types. It performs five sets of tests: (1) file property tests, (2) magic number, Macintosh file type code, and file pattern tests, (3) character set tests, and (4) tests of the content and layout of a text file, and (5) tests of file extensions and file context. The file type identifier first checks whether the file is empty, or if it is a directory file for a shortcut (LNK) file.

The magic number test checks the file to see if the data is in particular fixed formats. For example, DOS 16-bit executable files are binary files have the magic number 4D5Ah (MZ) stored in the first two bytes of the file that tells DOS operating system that this is a binary executable file. The following figure shows the first 176 bytes of a DOS 16-bit executable file.

```
000: 4D 5A 28 01 1F 00 17 00 - 20 00 00 00 FF FF A6 03 MZ<.........
010: 08 00 0C 64 00 00 37 01 - 1E 00 00 00 01 00 00 04 ..d........
020: 00 00 A8 10 00 00 DF 10 - 00 00 16 12 00 00 21 13 ...........
030: 00 00 2A 13 00 00 31 13 - 00 00 41 13 00 00 55 13 ..........A...
040: 00 00 D5 03 68 03 15 01 - 37 01 A7 01 37 01 F5 01 ........h...
050: 37 01 15 02 37 01 35 02 - 37 01 57 02 37 01 76 02 ?...?.?...v.
060: 37 01 44 04 37 01 EA 05 - 37 01 76 06 37 01 70 08 ?.D.?...?.z.?.v.
070: 37 01 07 09 37 01 37 09 - 37 01 00 00 00 00 00 00 ?...?.?...?
080: 00 00 00 00 00 00 00 00 - 00 00 00 00 00 00 00 00 .............
090: 00 00 00 00 00 00 00 00 - 00 00 00 00 00 00 00 00 .............
```

Additional tests are performed to determine the type of executable, e.g., types of self-extracting archive files.

The character set tests determine whether a file consists of US-ASCII, IBM (DOS) extended ASCII, or Windows ANSI 1251 character sets. If they do, the
file is a text file. Additional tests are performed to determine the kinds of terms occurring in the text file. These tests determine such file types as DOS Batch Command Text File, Quick Basic Program Text, or dBase Program Text.

Having identified as many files as possible by file properties, magic numbers, file patterns, character sets, content and form of text files, one finally checks the file extension and context of the file to determine the file type. If a file type cannot be determined by any of the above tests, its file type is said to be "Unknown."

2.3.4 Associating a File Type with a Viewer

APT uses file viewers or helper applications to interpret and display different types of computer files. To keep track of which helper application should be used with a file type, APT maintains a mapping from file types to helper applications. When an archivist double clicks on a file name, APT's File Type Identifier attempts to identify the file type, and if successful uses the mapping to find the appropriate viewer or helper application capable of handling that file type.

File viewers such as Quick View Plus are needed to view word processing documents, pictures, spreadsheets, and databases. Media players are needed to listen to sounds or play video. Archive file extractors such as WinZip are needed to extract files from archive files.

MIME (Multipurpose Internet Mail Extensions) is a standardized method for naming file formats. A User Agent sending email via the Internet (e.g., Eudora) uses a file's MIME type to communicate to a recipient User Agent the file type of attached files. The receiving User Agent maintains a mapping of MIME types to applications that can interpret and display a file of that MIME type.

When a web browser retrieves a file from a server, the server provides the MIME type of the file. The web browser uses the MIME type to determine whether the file type can be interpreted and displayed by the browser's built in capabilities or, if not, whether a helper application is available to interpret and display or play the file.

In the Windows 2000 operating system, file name extensions are associated with file types and file types with programs that can open the file. For instance, files that have the .txt or .log extension are of the file type Text Document and can be opened with any text editor, e.g., Notepad. The mapping can be seen
under the File Types tab of the Folders Options of the Control Panel (or the Tools option on the menu bar of Windows Explorer).

To see the current file types recognized by the APT and the file extensions and MIME Type associated with each file type, and the viewer or helper application associated with that file type, select the select Viewer under the Assoc. File Type with Application option under the Options drop-down menu.

The following dialog box will appear. This dialog box enables the archivist to add file types and edit information about file types that the APT does not recognize. There is a Remove button. This button is only enabled when a file type is highlighted that the user has added. This way the user cannot accidentally remove a file type that the APT recognizes. Under the File type details group in the dialog box there is a property called Defined by. This property shows whether the APT or the user added the file type to the list. Most of the properties in the details group can be edited using the Edit button. The only property that cannot be edited is the file type itself. It is an APT defined file type. All other properties are editable. If the file type is user defined, then even the file type property can be edited.
To add a new file type and associate it with a viewer helper application:
Step 1: Click the Add button to display an empty dialog

Step 2: Type the Group if you know it. Some Group names are Document, Graphics, Archive, and Executable Program.
Step 3: Type the name of the file type in the File Type name field.
Step 4: Type the common file extensions used by the File Extension field.
Step 5: Type the MIME type in the MIME Type field.
Step 6: Type (or use the Browse Button to select) the location of the application capable of interpreting files of this type in the Viewer field.
Step 6: Click OK to store the new information.

To edit an existing association of File Type and viewer application:

Step 1: Select a File Type from the List.
Step 2: Click the Edit button to display a dialog for changing the information associated with the selected File Type. The dialog displays the Group, file extensions, MIME Type and helper application for the selected item.
Step 3: Edit the Group, MIME Type and/or File extensions.
Step 4: Select the Viewer property of the dialog box. You can designate:
- Qvp32.exe (Quick View Plus)
- Another application: Type (or use the Browse Button to select) the location of the application capable of viewing files of this type in the Viewer field.
Step 5: Click OK to store the new information.

Note: You cannot change the file type name if the APT defined it, because this is the name associated with a file by the File Type Identifier. Changing the file type name of a user-defined file type adds a new file type. The file type with the previous spelling will need to be removed.

**2.3.5 Associate Archive File Type with Extractor**

You can add Archive Extraction Utilities to the APT by choosing *Extractor* from the *Assoc. Archive File Type with Application* option on the *Options* drop-down menu.
Archive File types, e.g., Zip, Zip self-extracting, ARC, must be associated with a program that can extract the contents of an archive of each type that the APT file type identifier recognizes. From the Options drop-down menu select Assoc. File Type with extractor. The following dialog box will be displayed.

The dialog box displays the archive file types that are recognized by the APT. This is about 13 different archive types plus their self-extracting versions.

This dialog box shows the name of the file type, possible filename extensions, a MIME-Type, the path to the program file that can perform the extraction, and any parameters that might be needed, e.g., -x indicating that the program should extract files from the archive.

2.3.6 Viewing the Manifest of a TAR Archive

If an archivist has used the APT to save a file system, the file system will be saved in a TAR archive file and it will have a manifest file. At a minimum, the manifest contains the names of directories in the file system, the names of the files, the file type of each file, and a message digest for each file. The message digest is computed from the file using the Secure Hash Algorithm (SHA-1). The
APT uses it to check that the file has not been inadvertently or purposely modified since an archivist last saved it.

If the file system has been rearranged, there may be an attribute Original Path. If the file system has been reviewed and saved, each file that has been reviewed will have additional attributes: Access (Open, Closed, Redacted), and if Closed, Reason for Withdrawal, Date of Withdrawal, Subject, Author, and Archivist initials. If the file system has been described, the entire file system may have attributes: Organization, Office, Person, and Series Title.

The manifest can be viewed by selecting the View>Manifest from the menu bar. The following figure shows the initial entries in the PC File Test Suite that has not been filtered, arranged, preserved, reviewed, or described but has been saved as a TAR file with Manifest.

2.3.7 Saving the Results of our Work
After exploring, filtering, arrangement, preservation, or review of the contents of a container, you will save the results of your work.

Step 1: From the main menu of the activity window, select File > Save As. You will see a directory with Archivist names or initials including yours. Open the directory with your name.

Step 2: In the **File name** field, enter the name you would like to call the file system. Click **Save**. You will use this container in future activities. Be sure to note where you store it.
If the file name you choose has already been used (perhaps because you are saving your changes to the same filename that you loaded, the following message will be displayed.

All containers are saved in a single file of file type TAR Archive. In addition, a METAINF directory is added to the file system that contains a file named manifest.mf. (See Viewing the Manifest of a File System). Sometimes the only changes to the content of the container are changes to the manifest file.

2.3.8 Closing a Container
To close a Container:
Step 1: From the main drop-down menu, select File>Close. The file container you had open will close.
Step 2: Exit the APT or open a different container.
2.3.9 Help

The Help drop-down menu provides you with an on-line version of this manual. In addition, it provides the capability to search for topics or terms in the on-line Help. It also indicates the version of the APT installed. Finally, it has a copy of Filex, a compiled Windows Help file containing a database of 4700 file extensions and descriptions. One can search for the description of a file extension, by selecting opening the Descriptions folder, selecting the beginning character of the file extension, and then scrolling down until the file's extension is found.

2.3.10 Processing Status

At the completion of each step of archival processing, the archivist should indicate the processing status of a container. In the APT with a container open, select Processing Status from the View drop-down menu. The Processing Status dialog box for this container will be displayed.

1. If the container has been filtered, or does not need to be filtered, Filtered should be checked by pointing at the Filtering check box with the mouse pointer and left clicking the mouse.

2. If the contents of the container have been arranged (e.g., extending folder titles), or does not need to be arranged, the Arranged check box should be checked.
3. If the contents of the container have been preserved (e.g., converted to a different format, file passwords recovered and files decrypted, files repaired), or preservation actions are not needed, the Preserved should be checked.

4. The Reviewed checkbox cannot be manually checked. It is automatically checked when all files in a container have been reviewed. After review is completed, the container should be checked back in, this returns it to Holdings. Return to ART, choose the Systematic Case Mngmt activity from the Activity drop-down menu, highlight the name of the container, and select Checkin Container from the File drop-down menu.

A container can be checked back in at anytime. If the tar file in the work area should become corrupted or damaged in any way, the archivist can select Undo Checkout from the File drop-down menu. The corrupted or damaged file can be deleted and the copy of the container that stayed in the holdings area can be checked out again.
5. Record Series are described in the Description Activity of ART. After the contents of containers in a record series are described, the Described checkbox for each container can be checked by highlighting the container and selecting Container Properties from the Edit drop-down menu. A dialog box similar to the following will be displayed.

![Edit Container dialog box]

There is a Processing Status button at the lower left. Selecting this button will display the Processing Status dialog box where Described can be checked.

![Processing Status dialog box]
The Described checkbox can only be checked after the other four checkboxes have been checked.

If a container has been copies to working storage by an archivist, the processing state that it was in when it was in the holdings area is viewable in ART, but not editable in ART except by the archivist who moved it to working storage or by the Administrator.

If a container has not been moved to Working Storage, the Processing status is viewable in the APT, but not editable in the APT. The status is editable in ART.

2.3.11 Exiting the APT

You can exit the Archival Processing Tool by selecting File > Exit or selecting the X on the title bar. If you have performed any of the APT activities, information about the files in the file system may have changed. If you have not saved the file system and its metadata, you may be asked if you want to do so.

2.4 Filtering the Files in a Container

When the personal computer files from the White House offices of the Bush administration were accessioned by NARA, it was discovered that they included operating system and office application program files as well as copies of Presidential e-records. The purpose of the filtering activity is to support archivists decisions in separating user-created files from the files of DOS and Windows operating systems and office application program files. A file filter is a pattern that is matched against files. A file blocking filter is a filter in which files that match the pattern are blocked from passing through the filter. The files that do not match the filter pass through the filter.

A coffee filter is analogous to a file filter. A coffee filter blocks the grounds from passing through the filter and allows the coffee infusion to pass through resulting in an enjoyable cup of coffee. It blocks what you do not want (the coffee grounds), and allows to pass what you do want (the coffee brew).
After you filter the contents of a container, you should keep the user-created files that passed through the filter for further processing. You should transfer the operating system and software application files that were blocked to the Library. In the Library, researchers can see for themselves that any files that were removed from the original file system are not records. Those blocked files may also be of historical interest as examples of the technology uses to create the Presidential records.

A record series (or file system of records) transferred to a central archives should contain only records created in the process of carrying out primary business activities. There are exceptional circumstances in which file systems are transferred that include operating system and/or application software files. If you are reasonably sure that the record series transferred contains only records, there is no need to perform the filtering activity.

2.4.1 The Filtering Menu

There are six drop-down menus available from the main Filtering menu.

The File, Activity, View, and Help drop-down menus are the same as explained in the section *Exploring a File System*.

**Filter:** This new menu provides the filtering functions of this activity. One can create a new filter, load a previously created filter, and save a filter that has been updated. There are four ways to filter: Filter Manually; Filter on File Type; Filter on OS/App Hash Code; Filter on OS/App Hash Code and File Type.
Archivists need only use *Filter on OS/APP Hash Code and FileType*, which is described in section 2.4.2. Furthermore, they need only read that section regarding filtering. The other filters were developed during the process of developing a technology to reliably solve the problem of filtering user-created files for non-user-created files. What we learned is that Computer Technicians should develop the OS/APP Hash Code and File Type Filter and the archivists should use it.

**Options:** This menu includes two new options: view a specific hash code filter, and define a file type filter.

The option *View OS/App Hash Code Filter* allows the archivist to see the hash codes, file names, and the reason that these files are included in the filter. The option to *Define File Type Filter* shows the entire list of file types than can be automatically identified. Those that are blocked have checks in the checkbox next to the name of the file type. Those that should have check marks are also listed in Appendix B of this manual.

### 2.4.2 Filter on OS/App Hash Code and File Type

As a result of experiments with the different kinds of filter, it is concluded that this is the best method of filtering to separate operating system and application software files from user-created files. As one filters additional file systems, the OS-App hash code part of the filter increases in size until almost all of the passed files are only user-created files.

#### 2.4.2.1 Loading the Filter

With the container to be filtered opened in the APT, select *Filter* from the *Activity* drop-down menu.
From the Filter drop-down menu, select Load.

You will see the filenames of several filters with the filename extension .flt.
Highlight the file name **OS-App Hash & File type.flt** and select **Open**.

This filter contains the hash codes and descriptions of several thousand previously identified by computer technicians as operating system and software application files. It also includes a score or more file types that have been previously identified by a computer technician working with archivists as the file types of files that will never be user-created records.

### 2.4.2.2 View the OS/APP Hash Code Filter

To view the loaded OS/App Hash Code Filter, select **View OS/APP Hash Code Filter** from the Options drop-down menu. You will see a display similar to the following.

This filter contains the SHA1 hash codes of files that were moved from the right to the left pane and which were added to the OS/APP Hash Code Filter. It also contains the file name, file length, File Type, and the comments entered. One can edit the comments.
View a Defined File Type Filter

To view the loaded File Type Filter, select Define File Type Filter for the Options drop-down menu. You will see a list of the names of categories of File types with check boxes beside them. The boxes with check marks in them indicate that the filter will block all file types in that category. For instance, in the following figure, the empty file, as well as all Help files, Operating System files and Word Processing Application files will be blocked.

Grayed checked boxes indicate that some but not all file types in that category have been checked and thus will be blocked. For instance, the figure below shows that there are only two Graphic File Types checked. Windows uses these types of files to display cursors and icons.
2.4.2.4 Filter on OS-App Hash Code & File Type

Select Filter > Filter On > OS/App Hash Codes and File Types.

If the hash code of a file in the opened container matches one of the hash codes in the OS/APP hash code filter, or the file has a file type that is in the list of checked file types, then the file is blocked and its path and filename is displayed in the left windowpane. Otherwise, the file is passed through the filter and its path and filename is displayed in the right windowpane.
2.4.2.5 Manually Filtering the Files Passed through the Filter

The Passed files, whose filenames are shown in the right windowpane, should be viewed to determine whether any OS and application files not defined in the filter have passed through the filter. Operating system and application software files should be moved to the left windowpane. This process should be continued until all that remains in the right windowpane are files that users created, used, or received in carrying out their primary business responsibilities. The blocked files will be transferred out of the current file system and save them in a container in the Library directory. The passed files will replace the contents of the container before filtering.

The field at the top of the left pane shows the number of files that were blocked by the filter and the field at the top of the right pane shows the number of files that passed through the filter. In the left windowpane, if one selects the down-arrow to the right of the field All File Types, one sees a panel showing the file types blocked by the filter. One can scroll down the list to see additional file types that were blocked.

If one selects one of these file types, the right windowpane will be repainted to show only the files of that type that occur in this file system. For instance, if one selects ASCII 7-bit Text, the left windowpane shows the six files of that type.
One can highlight the name of a file, such as GORILLA.BAS, and right click the mouse button. If the file is included in the OS/APP Filter, you can select OS/APP Properties.

You will see the Hash Code for the File and the reason that it is an Operating System or software application file. The hash code of GORILLA.BAS was the same as a hash code in the OS/APP Hash Code and File Type Filter, so it was blocked.

You can also highlight the name of a file, say GORILLA.BAS, and double click the left mouse button to view the file.
Exiting Quick View Plus and returning to the left windowpane, you can select other file types that were blocked, for instance, DOS Batch Command Text.
One can view the properties of these files by highlighting their file name and right clicking the mouse button. Notice that the above menu does not show OS/APP Properties. This is because the hash code of this file in not in the OS/APP Hash code Filter. It was blocked because its file type, DOS Batch Command Text, was in the File Type filter. Select Properties.
Selecting OK or Cancel, you are returned to filtering screen. With the filename of INSTALL.BAT highlighted, double clicking the left mouse button displays the contents of the file.

```plaintext
DRAFT

DRAFT

Selecting OK or Cancel, you are returned to filtering screen. With the filename of INSTALL.BAT highlighted, double clicking the left mouse button displays the contents of the file.

One can see that this is not a user-created file, but a DOS Batch file for installing dBase III Plus.

Viewing the files and the properties of the files in the left windowpane should increase an archivist’s confidence in the reliability of the filter in blocking OS and office application files.

In the right windowpane, if one selects the down-arrow to the right of the field All File Types, one sees a panel showing the file types blocked by the filter.
```
Highlighting a filename and double clicking the left mouse button, one can view a file. For instance, the contents of filename 107.txt in the CORR folder is:

```
The White House

February 4, 1991

Dear Fred Kleinknecht,

What a wonderful idea it is to dedicate a special issue of the Scottish Rite Journal to America’s children with learning disabilities. Scottish Rite Childhood Centers have helped so many who might otherwise go unaided, and your magazine is another fine way to offer support.

I remember with great fondness my visit with the children and staff of your wonderful facility in our nation’s capital. The enthusiasm and responsiveness of everyone I met touched me deeply, and it makes me so glad to know that this is only one of 74 clinics serving our children across the country.

Thank you for caring, and best wishes for continued success,

Warmly,

Barbara Bush

Mr. Fred Kleinknecht
Ancient Accepted Scottish Rite
Valley of Cleveland, Ohio```
This is a copy of a user-created e-record. However, the file COUNTRY.TXT in the DOS Directory is documentation about the operating system.

You should move the file and its path from the file system shown in to the right pane to the operating system and software application files shown in the left pane. To accomplish this:

Step 1: Highlight the text "Supplemental Information on MS-DOS 6.22." Select Copy from the Edit drop-down menu of Quick View Plus. Exit Quick View Plus.

Step 2: Highlight the filename.

Step 3: Click on the arrow between the windowpanes that points to the left windowpane. A dialog box will pop up that shows the properties of the file.
The dialog box shows some properties of this file: the SHA-1 hash code of the selected file, its file name, length and file type.

Step 4: One should enter in the comment field other information that can be inferred about the file: the manufacturer, operating system, product name, product version, and/or file function. In this case, we paste in the text copied from the document, by placing the mouse pointer in the Comment field, right clicking the mouse button and selecting paste.

Step 5: If one selects the Yes button, then the hash code of that file and the information in the dialog box will be added to the OS/App Hash Code filter. If one selects No, the file will be moved to the left pane, but will not be added to the OS/App Hash Code filter. One would select No, if one could not determine the manufacturer, operating system, product name, product version, and/or file function, but was confident that it was not a user created file. In this case, Yes is selected.
The file and its path is moved from the right windowpane to the left window pane. Notice that the arrowhead between the two panes points to the right because a filename in the left pane is highlighted. This enables one to move a file back to the right windowpane showing Passed, user-created files, if they decide that they made a mistake.

You can scroll through the remaining file types, viewing files and deciding whether they are user-created files or OS/Application files, and separating those that are OS/Application files from the user-created files. In any case, always return to view All File Types in the Passed Files to ensure that those files that remain are only user-created files. In this case, DOSHELP.HLP and GRAPHICS.PRO are also OS/APP files and moved to the left window pane of Blocked non-records.

The passed files include an archive file, ARC600.EXE. Archive files include files created with PKZIP, ARC, ARJ, and other archiving applications. Some archive files have an “exe” file name extension because they extract their own contents. These files are called self-extracting archive files. Even self-extracting files can be viewed to show their contents. By highlighting the filename and double clicking the left mouse button, the contents of the archive file can be viewed.
All archive files in a container should have their contents extracted in order to determine the file type of the files in the container. An Archive file can have its contents extracted during any activity, though it is best to extract files during filtering, because some of those files might be operating system or software application files. Archive files can be extracted by exiting the filtering process and returning to the main filtering window. Exit the filtering process, by selecting **Cancel**. You will be asked:

![Exit Filter](image1.png)

Answer Yes. You will be asked:

![OS/App Hash Code Filter](image2.png)
Answer Yes. The hash codes of any files that you moved from the Passed files to the Blocked Files will be saved in the OS/App Has Code Filter. You will be returned to the main Filtering screen.

### 2.4.2.6 Extract Archived Files

Highlight the name of the archive file.

![Image of Archival Processing Tool - C:\Work Area\sl\15001]

Select Extract from the File drop-down menu. If the Extract is grayed out, Extract is not enabled. This would be the case if the selected file's file type was not an archive.

![Image of Archival Processing Tool - C:\Work Area\sl\15001.tar]

If no extractor had been associated with the file type of the selected file, the following message box will be displayed.
If an extractor has been associated with archive files of that type, but the APT cannot find it in the given location, the following message box will be displayed. This can happen if the extractor has been moved or its name has been changed for some reason.

![Extraction Error]

A dialog box will be displayed that shows the filenames and directories that would be extracted. This is done in case the archive is damaged. This allows you to see whether all the files and folders will be extracted correctly.

![Replace Archive?]

You will be prompted as to whether you want to replace the archive file. If so, a directory with the same name as the archive file and on the same path as the archive file is created. The files in the archive are extracted into that directory. The files are checked for computer viruses. The file types of the files are identified.
The Files in the container need to be filtered again. Select Filter on OS/APP Hash Code Filter. The following Window is displayed. Only the user-created files are passed through the filter. The files previously included in the OS/APP Hash Code Filter were blocked and the files extracted from the archive were already included in the OS/App Hash Code Filter.
2.4.2.7 Transferring the Blocked Files

When all that remains in the right windowpane are user-created files, you should select the Transfer button below the left windowpane. A Transfer Information dialog box will appear.

The archivist need only describe the kinds of files being transferred. The other information is pulled from the manifest of the container and predetermined locations. When OK is selected, the blocked files are transferred. If you are prompted as to whether you want to save the OS>APP Hash Code Filter, answer Yes. The APT window will appear with the file system that was last displayed in the right windowpane displayed in the left panel.
The information in the dialog box is used to create a transfer sheet file named "!TransferSheet.txt" that is included in the package of blocked operating system and application software files and with the user-created files. The file name has the "!" in front to prevent the chances of a duplicate file name in the root directory. If one views the file "!TransferSheet.txt", this is what it will look like.

2.4.2.8 Saving the User-Created Files

After filtering the container, and transferring files that are blocked or manually filtered to the Library, you should save the remaining user-created files back to the container checked out from Holdings. From the main menu of the filtering activity window, select File > Save.
Though currently there is a Save As… option under the File drop-down menu, this option is not recommended and may be removed in future versions of the application. The Save As… option allows the archivist to save a copy of the container with a different name. This could cause confusion when there are multiple copies of the same container in the archivist’s work area with different filenames. When it was time to check the container back in the archivist would have to decide which one to check back in and manually delete all remaining copies.

2.4.2.9 Exiting the APT from the Filtering Activity

One can change to another Activity by selecting that activity form the Activity pull-down menu. The container that you have filtered will remain open in the APT. However, if you exit the APT from the Filtering Activity by selecting File > Exit, or selecting the Close button, X, at the upper Right of the Window, you may be prompted...
If you are not sure that you saved any work that you have performed on the currently open container, select Yes. You will see the same screen as described in the previous section,

### 2.4.3 Filter Manually

A file system needs to be filtered manually only when no filter has previously been created. After filtering a single file system, a filter should have been created, so one may filter manually only once. However, automatic filtering often does not remove all operating system and software applications, so the filtering ends up being semi-automatic, with some files being filtered manually.

In this mode, the archivist will act as the filter, viewing each file or instance of a file type, deciding whether it is a user-created file (record) or whether it is an operating system or software application that might have been used to create the records. The APT will support this process with a user-interface for separating the files in the file system into two file systems, one containing operating system and software application files and the other containing just the user-created files. Since the archivist is unlikely to be able to recognize legacy operating system and software application files, the APT's file type identifier will perform that function. The results of the archivist's decisions as to what are operating system or software applications can be saved as an OS/App filter. Hence, the second file system can be filtered using that filter, and the archivist will not have to make the same decisions twice.

#### 2.4.3.1 Creating an OS/App Hash Code Filter

If one selects *Filter > Filter Manually*, the file types of the files in the file system will be identified and a window will be displayed with two panes.
The field at the top of the right pane shows the number of files in the file system. If one selects the down-arrow to the right of the field All File Types, one sees a panel showing the file types identified in the file system.

One can scroll down the list to see additional file types that are identified in this file system. The entire list of file types that can be identified is shown in an appendix to this Reference Manual.

If one selects one of these file types, the right windowpane will be repainted to show only the files of that type that occur in this file system. For instance, if one selects DOS 16-bit Compact Executable, the right windowpane shows the 16 Files of that type.
Note that these 16 files are of the same file type, but have three different file extensions: COM, GRB, and SG3. One can double click on a filename to view a file. Suppose that we view DBC.COM.
Even though the APT properly identified this file type, the lower left-hand corner of this Quick View Plus display indicates that Quick View Plus did not recognize this file type. Quick View Plus is configured to display Unknown File Types as Text Files. The occurrence of some box drawing characters in the display indicates that Quick View Plus displayed this file using the DOS Extended ASCII character set. The text indicating that this file dBCODE (2.06) copyrighted by Ashton-Tate, is also indicative that this is not a user-created file. If one scrolls through the rest of the file one finds the following text.

```

  -e ASCII copyright header file
  -i generate information file(s) (dbg)
  -o destination directory
  -r response file
  -s source directory

file... dBASE III source code file(s) (.src)

Enter a source filename (.src) or press ENTER to end.
WHILEKEY

*** unexpected input out of memory onParse error and of input
onRead error onWrite error unable to open respondSource -
output filename conflict -Unrecognized command illegal DO
statement illegal RunTime+ SET option illegal command oacro
(6) inIllegal condition in ON statement illegal action in ON
statement in file 012345678ABCDEF Completed.

** expected END

*** not found!
```

This file is a dBASE III Runtime Psuedo-compiler, not a user-created record. Close the viewer. You should move the file and its path from the current file system to the operating system and software application files shown in the left pane.

Step 1: Highlight the filename.

Step 2: Left click the mouse button on the arrow between the windowpanes that points to the left windowpane. A dialog box will pop up that shows the properties of the file.
The dialog box shows some properties of this file: the SHA-1 hash code of the selected file, its file name, length and file type.

Step 3: One should enter in the comment field other information that can be inferred about the file: the manufacturer (Ashton-Tate), operating system (DOS, because it's a DOS compact executable), product name (dBase, because its in a dBase directory with other dBase executable programs), product version (III+), and file function (Runtime Pseudo-Compiler).

Step 4: If one selects the Yes button, then the hash code of that file and the information in the dialog box will be added to the OS/App Hash Code filter. If one selects No, the file will be moved to the left pane, but it will not be added to the OS/App Hash Code filter.
The remaining DOS Compact Executable Files should be viewed, and if non-records, moved to the file system in the left windowpane, described and included in the OS/APP Hash Code filter.

You can scroll through the remaining file types and view them one at a time, viewing them and deciding whether they are user-created files or OS/Application files, and separating those that are OS/Application files from the user created files. In any case always return to view All File Types in the Passed Files to ensure that those files that remain are only user-created files.

2.4.3.2 Cancel

If at any time you select the Cancel button in the lower right-hand corner of the Window, you will be asked whether you want to save the OS/App hash codes added during this session. If you say yes, any OS/APPs you added to the filter will be removed, and you will be returned to the main filtering activity window with the original file system loaded.

Suppose you are working with a large container. You could be filtering using any of the methods. Suppose you have only partially reviewed the passed files, and need to stop. Should you save the blocked files and the passed files? There is no need to. Just be sure to save the Filter. When you come back, use the filter to which you saved your work. Filter on either OS/App Hash code or OS/App Hash Code and File type and you will be put into the same state as when you had to stop your work previously.

2.4.3.3 Transferring the Blocked (OS/App) Files

When all that remains in the right windowpane are user-created files, you should select the Transfer button below the left windowpane. A Transfer sheet dialog screen will appear.
The information in this dialog box will be used to create a transfer sheet file named "!TransferSheet.txt" that is included in the package of blocked operating system and application software files and with the user-created files. The archivist need only describe the kinds of files being transferred. The file name has the "!" in front to prevent the chances of a duplicate file name in the root directory.

You need to do the same for any records that are marked for transfer during review. If one subsequently views the file, this is what it will look like.
The archivist will be returned to the Main APT window with the file system that was last displayed in the right windowpane displayed in the left panel.

### 2.4.3.4 Saving your work

After filtering the container, and transferring files that are blocked or manually filtered to the Library, you should save the remaining user-created files back to the container checked out from Holdings. From the main menu of the filtering activity window, select *File > Save*. Though currently there is a *Save As...* option under the *File* drop-down menu, this option is not recommended and may be removed in future versions of the application. The *Save As...* option allows the archivist to save a copy of the container with a different name. This could cause confusion when there are multiple copies of the same container in the archivist’s work area with different filenames. When it was time to check the container back in the archivist would have to decide which one to check back in and manually delete all remaining copies.
2.4.3.5 Saving a Filter

If you have added OS/App Codes to the Filter, you need to save the filter. Select Filter > Save As.

In the next section, you will learn how filtering a filing system using the OS/App Hash Codes Filter can reduce the manual effort required to filter operating system and application software from file systems.

2.4.4 Filter on OS/App Hash Code

Suppose you have opened a file system that needs to be filtered and have selected the Filtering Activity. You can load a previously created OS/App Hash Code filter and select the Filter on OS/App Hash Codes option. Using this action, the hash code of each file in the loaded file system will be computed and compared to the hash codes in the filter. Those that match the filter will be automatically blocked and those that do not match will be passed.

2.4.4.1 Loading a Filter

From the Filter drop-down menu, select Load. You will see the filenames of several filters with the filename extension .flt.

The filter OS-AppTutorial Filter.flt is the one created during the process of understanding the Manual Filtering process.
The Filter OS-APP Hash Code.flt contains the hash codes and descriptions of several thousand operating system and software application files that have been identified during experiments with the Bush PC Hard Drives.

A filter constructed from the National Software Reference Library (NSRL) Reference Data Set (RDS) version 1.1 is included with the APT. It contains SHA-1 hash codes of 3,001,846 files from versions of approximately 1450 operating systems or products. This filter could be used instead of the OS-APP Hash Code.flt. The NSRL continues to add legacy and current operating and application files to the RDS, so it can evolve into a valuable file system filtering resource. Unfortunately, it does not contain the hash codes of many of the operating system and software applications from the period of the Bush Administration (1989-1993).

2.4.4.2 View OS/App Hash Code Filter

To view the loaded OS/App Hash Code Filter, select View OS/APP Hash Code Filter from the Options drop-down menu. You will see a display similar to the following.
This filter contains the SHA1 hash codes of files that were moved from the right to the left pane and which were added to the OS/APP Hash Code Filter. It also contains the file name, file length, File Type, and the comments entered. One can edit the comments.

2.4.4.2 Filtering on OS/App Hash Code

Select **Filter On > OS/App Hash Code** from the **Filter** drop-down menu. The file system will then be filtered using the loaded OS/App Hash Code filter. The blocked and passed files will appear.

The “Blocked files” field is found above the left pane. These files have been filtered out of the original filing system because their hash codes matched hash codes in the OS-App Hash Code Filter. The number of blocked files is also given.

The “Passed files” field is found above the right pane. These files have passed through the filter. Note the number of passed files. “Time to Filter” is in the lower right-hand corner of the screen displays the length of time in seconds necessary to filter the entire file system.
2.4.4.3 Viewing Types of Blocked Files

The types of blocked files can be viewed by selecting the down arrow next to the filed All File Types below Blocked Files. After acquiring some experience with different file types, you will recognize that all of these file types, with the possible exception of ASCII 7-bit Text, are operating system or software application files.

You can select ASCII 7-bit Text and will see a display similar to the following.
You can double click on each of these files to view their contents. The CONFIG.DB files are text files used to configure the DBASEIII+ database management system. README.TXT files are commonly installation information or information that was not included in the User Manuals. The files with .BAS extensions are BASIC programs, often games supplied with the operating system.

2.4.4.4 Manually Filtering the Files Passed through the OS-APP Hash Code Filter

The Passed files should be examined manually as described in the section Filter Manually:

- View the file types of passed files.
- View the individual files.
- Move operating system and application software files to left windowpane.
- Add them to the OS-App Filter.
- Continue this process until all that remains in the right windowpane are files that users created, used, or received in carrying out their primary business responsibilities.
- Transfer the blocked files out of the current file system and save them in a TAR file in the Library directory.
- Save the passed files in the Archivist's Work Area directory.
- Save the updated OS-App Hash Code Filter.

2.4.5 Filter on File Type

There are some file types that will never be created by a user in carrying out their primary business responsibilities, unless they are a programmer. These are files that are used by an operating system or software application. Examples are font files, built in dictionaries, and icon files. Other file types will differ from PC to PC depending on what applications are loaded, e.g., system configuration files. It would certainly be a simpler matter, if we could filter on file type rather than having to identify individual files to create a filter of operating system and software application files.

If one selects Define File Type Filter from the Options drop-down menu, one sees the dialog box shown in below.
With the exception of the empty file, each of the terms represents a category of file types. You can see the file types included in a category by clicking on the plus icon in preceding each category. Each of these categories will be described. Then it will be explained how to define a file type filter.

**Archive**

An archive file is a file that groups several related files into a single easily managed file. Often the files in an archive file are compressed. This compression is performed to make transmission or copying of these files faster. For instance, downloading an archive file only requires one file transfer operation and the file transfer time is minimized if the files in the archive are compressed.

An archive file may contain operating system files, application software files, user-created files, other archive files, or a combination of these. One needs to open an archive and view its contents to know whether they are dealing with OS files, application files, or user-created files. Hence, one should not filter on Archive file types.

While most executable files are operating or software application files that are not created by users of the software applications, the users of the applications may have archived some of the documents they created in a self-extracting archive file. A **self-extracting archive file** is an executable program file that includes both an archive file and software to extract the contents of the archive.
file. The contents of a self-extracting archive file can be extracted by simply executing the file. When Quick View Plus or WinZip is used to open a self-extracting archive, they do not automatically extract the files, but simply display the file names of the files. Each of the files can then be selected and viewed with Quick View Plus.

Calendar Files

PC Calendar software for scheduling appointment and meetings became very popular in the late 80's. Version 2.04 of the APT only recognizes the format of WordPerfect Office Calendar version 3.0, but calendar files created by other calendar software are likely to occur. One should not filter on calendar file types.

Database Files

PC database management software such as Ashton Tate's dBASE III Plus and dBASE IV became very popular in the late 1980's, and were often used to develop databases to support business activities. The APT currently recognizes only the database formats of the dBASE and related products. However, there are other database applications and file formats that are likely to occur such as Paradox, Advanced Revelation and Reflex.

Empty Files

An empty file is a file of length zero or a file containing only an end-of-file character. It has no content. This category should always be filtered on because it is a file with no content.

Executable, Overlay and DLL

An executable file is a program file in a format that a computer can directly execute. While humans can read source program files, they cannot read executable files. To transform a source program file into an executable file, the source program must be passed through a compiler or assembler. An overlay file is a file containing additional parts of a program that is too large to fit into memory along with the main executable file.

A dynamic link library (DLL) is a collection of small programs, which can be called when needed by the executable program (exe) that is running. The advantage of DLL files is that, because they do not get loaded into random access memory (RAM) together with the main program, space is saved in RAM.
When and if a DLL file is called, then it is loaded. For example, if you are editing a Microsoft Word document, the printer DLL file does not need to be loaded into RAM. If you decide to print the document, then the printer DLL file is loaded and run. A DLL is an executable file that cannot run on its own, it can only run from inside an executable file.

Lz.exe is an executable file compression utility for MS-DOS. It adds a minimal header to the executable to decompress it when it is executed. PKLITE is an executable file compression utility for MS-DOS from PKWARE, Inc. that compresses the body of the executable and adds a small, fast decompress routine in the header.

Executable, Overlay and DLL files are usually operating or software application files. However, there are important exceptions. For instance, some software applications create executable report files that may be records created in the course of their business activities. Furthermore, self-extracting archive files, which are executable files, may contain user-created documents.

The APT is able to distinguish some self-extracting archives from other executable files, but it cannot identify all kinds of self-extracting executables. Furthermore, some self-extracting archives are compact executables. Hence cannot reliably filter on executable or compact executable file types.

**Graphic Files**

A graphics file is a file that contains only graphic images such as line drawings, paint program files, scanned images, photographs and other halftones, or type designs within a graphics file format. A vector graphics file is one that uses geometrical formulas to represent images. The other method for representing graphical images is through bit maps, in which the image is composed of a pattern of dots. This type of file is called a raster graphics file. Vector-oriented images are more flexible than bit maps because they can be resized and stretched. Fonts represented as vectors are called scalable fonts or outline fonts. Most output devices, including dot matrix printers, laser printers and display monitors, are raster devices (plotters are the notable exception). Thus, most graphics objects must be translated into bit maps before being output.

Some of the graphics files in a file system, such as icon and cursor resource files, are system files. Other graphics files in a file system may be user-created files. However, it is likely that sample files provided with software applications will need to be included in OS-App Hash Code filter. Hence, should not filter on Graphic File types.
Help Files

Operating systems such as DOS and Windows and software applications such as WordPerfect provide online documentation through help systems. These files are not user-created documents and can be included in a file type filter.

Operating System Files

Some operating system files contain device drivers, system data, hardware configuration, and system installation information. These are not user-created files. One can include these in a file type filter.

Some files are created by users for communicating with the operating system or software applications. These files should not be preserved as records. Furthermore, these files will differ from file system to file system depending on the configuration of application software in the filing system. The following shows the contents of one of these files, AUTOEXEC.BAT, a DOS batch command file.

```
C:\WINDOWS\net start
@ECHO OFF
SET COMSPEC=C:\DOS\COMMAND.COM
SET PATH=C:\WINDOWS;C:\DOS;C:\wbivb4;C:\DOS\BATS
SET QBACKUP=C:\DITTO\DOS
PATH=C:\DITTO\DOS;%PATH%
LH /L:1,16944 SHARE /L:500 /F:5100
set ROOTDIR=c:\MK3
set TMPDIR=c:\temp
SET TMP=C:\TMP
SET TEMP=C:\TMP
SET BW=C:\BW
SET WINPMT=\s\e\s\e\7m\e[K\{Windows Shell}\$p\e[m\e[u\n\g
SCANDISK C: /NOSUMMARY
LH C:\DOS\SMARTDRV.EXE /L /X 2048 128
PROMPT $P$g
LH /L:1,36544 C:\WINDOWS\MSCDEX.EXE /S /d:mcd001 /m:10 /n /e
C:\mouse\mouse.COM
LH /L:1,6384 C:\DOS\DOSKEY.COM
win
```

Each filing system filtered will likely have a different AUTOEXEC.BAT file. Other examples of files created to communicate with the operating system are CONFIG.SYS, a DOS hardware configuration file, Windows Program Initialization Files (.ini), and Windows 3.1x, Windows 95, and Windows NT Installation Information Text Files (.inf).
Spreadsheet Files

Spreadsheet applications, such as Lotus 1-2-3, Excel, and Quattro Professional, store spreadsheet labels, cells, and formulas in files of specific format. These applications support graphic features that enable users to produce charts and graphs from the data. Unless they are sample or tutorial files, the spreadsheet files, chart and graph files are user-created and should not be included in a file type filter.

Application Resource Data

Software applications for word processing, spreadsheets, and presentations use a number of data files to support users. These include keyboard macros, template or style files, dictionaries, and fonts. Most of these are supplied with the software application, but some may be user-created.

Word Processing Documents

The documents created by users of word processors and text editors are stored in a variety of word processing (or document) file formats. Most of these will be user-created documents, but some may be sample documents.

2.4.5.1 Define a File Type Filter

This filter option is available when you know the file formats you wish to block. You define the filter by selecting specific file formats.

Step 1: From the Options drop-down menu, choose “Define File Type Filter.”

Step 2: From this list, select the file types you would like to block by clicking once in the appropriate check boxes, and then click the Close button. In this example, “Empty Files,” “Operating Systems,” “Help,” and "Word Processing Application Resource" are checked. We should expect these files to be blocked after filtering.
Grayed checked boxes indicate that some but not all sub check boxes have been selected. The figure below shows that there are only two Graphic File Types checked. Windows uses these files to display cursors and icons.
2.4.5.2 Filter on File Type

From the Filter drop-down menu, select "On File Type." The file system will be filtered according to which file types you defined in the File Type Filter. A screen similar to this should appear:

You can select the down arrow to the right of All File Types in the left windowpane to see the file types that were blocked. You can view each of these files to confirm that the file types you have selected to include in the filter are indeed those of operating system or software application files.

2.4.5.3 Manually Filtering the Files Passed through the File Type Filter

The Passed files should be examined manually as described in the section Filter Manually:

- View the file types of passed files.
- View the individual files.
- Move operating system and application software files to left windowpane.
• You will not be prompted as to whether you want to add the hash code of the moved file to the OS-App Hash Code Filter, because you are not filtering on OS-App hash codes.

• Continue this process until all that remains in the right windowpane are files that users created, used, or received in carrying out their primary business responsibilities.

• Transfer the blocked files out of the current file system and save them in a TAR file in the Library directory.

• Save the passed files in the your directory in the Work Area directory.

• Save the File Type Filter.

2.5 Arranging Files in a Container

In section 1.1, the description of intellectual arrangement by collection or record group and by series was described. The physical arrangement of containers of records within record series was also described. In this section, arrangement of folders (directories) within a container and records within folders is described. If operating system and application software files need to be filtered out of a container, the filtering activity should be performed before arrangement.

2.5.1 The Arrangement Menu

If a container is opened and the arrangement activity is selected, seven drop-down menus will be available. The File, Activity, View, Options, and Help drop-down menus are the same as explained in the section Exploring a File System and are the same for every activity.

Edit: This menu allows you to create a new top-level folder, a new subfolder, to rename a folder, or to view the properties of a file or folder.
Sort: This drop-down menu allows you to sort the contents of a folder by date last modified, file name, or back into its original order.

Tools: The Auto Extend Folder Title tool will apply to a single folder is the folder is highlighted and all folders if the container name is highlighted. It is based automatic recognition of the record types in a folder. For instance, if all the records in the folder "CORR" are recognized to be White House Correspondence, the title would be extended to be "CORR[ESPONDENCE]."

[Not yet integrated]

2.5.2 Moving a File into a Folder
Suppose that the loaded container appeared as follows. The files have all been viewed and are all WordPerfect 5.x documents. However, two of them are in the root directory.

If you believe that the file in the root directory should be in the WP50 directory, you can highlight the filename of a file in the root directory, hold the left mouse button down, and drag the file to the WP50 directory.

You can do the same for the other file in the root directory.
2.5.3 Reordering the Files in a Directory

You can review the order of the files in the directory. You will notice that they are not in ascending order of Filename. Nor are they in ascending order of Date Last Modified. Suppose that you believe a more logical order for these files is by ascending order of Date Last Modified. Select Sort > By Date Last Modified.

The following screen shows the result. The upward pointing arrowhead in front of Last Modified indicates that the files are in ascending order.

Now, if you click on column header Last Modified, the files will be sorted in descending order of Date Last Modified.
To get the files back into their original order, select the Original Order option from the Sort drop-down menu.

### 2.5.4 Viewing the Properties of a Folder or File

If you highlight a folder name and select *Edit > Properties*, you will see the properties of the highlighted folder.

![Folder Info]

The files in the folder WP50 are sorted in ascending order of Date Last Modified. If you select the downward pointing arrow to the right of the Sort Attribute field, you will see Filename and None, in addition to Last Modified. If you select, Filename or None, the Files in the Folder will be resorted. Similarly, if you select the downward pointing arrow you will see Descending in addition to Ascending. If you select Descending, the files in that folder will be sorted into descending order of the value of Sort Attribute.

You can view the properties of a file by highlighting its filename and selecting *Edit > Properties.*
In this case, the filename of a file that was moved was highlighted. The current path to the file is shown along with its original path. When a container that has been arranged is saved, these attributes will be saved in its manifest file.

2.5.5 Extending a Folder Title

DOS directory names could only be up to eight characters in length plus a three-character extension. In the example of this section, the directory name WP50 was indicative of the application used to create, edit, print, and view the files. If you believe that a directory (or folder name) is needed that is more indicative of the contents of the folder, you can provide this by extending the folder title. Select Extend Folder Title from the Edit drop-down menu.

The files in this folder were taken from the Bush Public Papers; so let us extend the title using that phrase. Unlike renaming, extending involves adding characters to the current title. We will follow the convention used at the Bush Presidential Library of enclosing additions to a folder title in square brackets.
2.5.6 Creating New Folders

Suppose that upon examination of the files one discovers that some were Public Statements, others were Letters of Transmittal, and some were Appointments. You can create subfolders within the WP50 directory.

Step 1: Highlight the folder in which you want to create a new subfolder.

Step 2: Select *New Sub Folder* from the *Edit* drop-down menu.

Step 3: A pop-up dialog box will appear. Type the name of the new subfolder in the field, and click OK.
A new subfolder of that name will be created.

You can then drag and drop the files of those particular types into the subfolders.

If a file system does not have a top-level folder but is just a root directory of files, you can create a top-level folder.
Step 1: Highlight any one of the files in the root directory.
Step 2: Select New Top Level Folder from the Edit drop-down menu.

A pop-up dialog box will appear.

Step 3: Enter the name of the top-level folder, and click OK. A top-level folder of that name will be created.

You can then drag and drop files into that folder. You can also add subfolders to the new top-level folder.

2.6 Preserving Electronic Records
Archivists who used the APT to view and arrange Bush Administration e-records learned that for a variety of reasons there were some files that they could not view. The files that were not viewable included password protected or encrypted files, damaged or corrupted files, and files of a format for which there was no viewer. The Preservation activity of the APT was developed to support archivists in solving these problems.

Personal Computer hardware of 15-20 years ago (IBM AT, IBM PS/2) is now obsolete. The operating systems (DOS and Windows 3.x) that operated on that hardware are also obsolete. The office application software (Word 2.0, Lotus 1-2-3 version 1, Word Perfect 5.x, dBase III+) that operated on that hardware and those operating systems is also obsolete. Many of those legacy software applications will not execute on current computer platforms. Some of the files created by those legacy software applications will need to be migrated to current or standard formats.

Digital preservation involves some type of transform of an original file. The initial research of the PERPOS project was directed toward gaining archival control of records in their native formats. However, some file formats cannot be displayed or viewed without some transformation of the original file. These include files in a format for which there is not a viewer, password-encrypted files, and damaged or corrupted files.

2.6.1 The Preservation Activity Menu

To get support in preserving electronic records, select Preservation from the Activity drop-down menu.

![Archival Processing Tool - C:\Work Area\sl\15002.tar]

Six drop-down menus are available. The File, Activity, View, and Help drop-down menus are explained in the section Exploring a File System and are the same in every activity. Additional options have been added under the under the Action and Options drop-down menus.
The actions under the Action drop-down menu option are Recover Password, Decrypt password encrypted file, Repair file, and Convert file to another format.

The options drop-down menu supports associating file types with applications for recovering passwords, decrypting files, repairing files, and converting files to standard or current file formats.

2.6.2 Recovery of Passwords and Decryption of Password Encrypted Files

During experiments in processing the contents of the Bush hard drives, files were detected by Quick View Plus that had been encrypted using a password. Since the National Archives has the legal and physical custody of these files and is responsible for their preservation, archivists need the capability to recover the password in order to decrypt the files. The recovery of a password for legitimate and practical purposes should be distinguished from cracking of a system or file password for illegitimate purposes such as theft or vandalism. However, the techniques are the same.

There are commercial-off-the-shelf (COTS) products that recover passwords of files protected with encryption procedures built-into software applications such as WordPerfect, MSword, Quattro Pro and Pkzip.

2.6.2.1 Recovery of Passwords
Quick View Plus may recognize that a file is password encrypted. For instance, when an attempt was made to view the file WORD_TST.DOC, Quick View Plus displayed the following message.

To recover the password for this file, select *Preservation* from the *Activity* drop-down menu. Highlight the filename of the encrypted file. Select *Recover Password* from the *Action* drop-down menu.

One or more passwords are recovered by an application associated with the file type of the file. These passwords are displayed on the screen.
The Archivist is responsible for recording the password for later use in decrypting the file.

If there is not a password recovery application associated with the file type of the file, the following message is displayed.

You should notify the local archival computer technician, or the PERPOS staff and an attempt will be made to locate the required password recovery application.

**2.6.2.2 Decryption of Files**

When the password of an encrypted file has been recovered, the file can be decrypted using the following procedure. With the file system containing the encrypted file open, select the *Preservation* activity.

Step 1: Highlight the filename of the encrypted file. Under the *Action* drop-down menu, select *Decrypt*. 
Step 2: You will be prompted to enter the password.

When the password is entered, select OK. The decrypted file will be opened in the application.
Step 3: Select *Options* from the application’s *Tools* drop-down menu. Then select the Save Tab. This will display the save options. You will see that there is a password box that is filled with ‘*’s.
Step 4. Blank out the “Password to open:” box then select OK. This removes the password protection from the file.

Step 5: Select Save from the application's File drop-down menu or press Ctrl+S. This saves the decrypted file where the APT can find it.
Step 6: Select Exit from the application's File drop-down menu. A dialog box will appear asking whether or not you successfully decrypted the file.

Step 7: Select Yes, if you had no trouble decrypting the file. If the password was invalid or there was any other trouble, select No.

If Yes was selected, the decrypted file replaces the original password encrypted file and keeps the same name and extension. The action that was taken is reflected in the right-hand pane of the APT. A new file type will be listed, one
without the word encrypted, and the action Decrypted will be listed with the date and the archivist who performed the action.

2.6.3 Repair Corrupted Files

During systematic processing, archivists using PERPOS sometimes encountered a file that might be an electronic record but could not be read, or displayed, or displayed properly. For example, Quick View Plus may display a message that it cannot display the file, or a "loop" may occur in the document so that moving the cursor down brings you to an earlier part of the record. The file header or function codes in the text of the file have been damaged.

Suppose that you were viewing the sample file whose file name was highlighted in the screen shown below.

It is of unknown file type. The filename extension (DBF) would indicate that the file is a database file. When viewed with Quick View Plus, it is clear that there is some information at the beginning of the file that should not be there. The file has been damaged, in this case purposely, to demonstrate the capability to repair the file.
There are other files in the same directory with the filename extension FPT. FPT is the filename extension for Memo files for FoxPro databases. Sample1.DBF is probably a damaged FoxPro Database with a memo file.
If the file type of SAMPLE1.DBF is changed to "FoxPro Database w fpt," there may be a repair utility that can repair the damaged file. To ensure that we enter the name of the file type correctly we can copy the file type from the Properties of SAMPLE2.DBF to the File type Property of SAMPLE1.DBF. To accomplish this, highlight the filename SAMPLE2.DBF and select View > Properties.

![File Properties](image)

Highlight the value of File Type and right click the mouse. Select copy from the popup menu. Select OK on the File Properties Menu. Then highlight the filename SAMPLE1.DBF. Select View > Properties. Delete the value "Unknown" in the File Type field. Right click the mouse and paste from the clipboard the copied value into the File Type field.
Select **Preservation** from the **Activity** drop-down menu. Find and highlight the filename of the corrupted file. Select **Repair** from the **Action** drop-down menu.

There is a file repair utility for files of type "FoxPro Database w fpt," and that utility examines the file and displays the following message.

You will be asked:
Do you want to replace the damaged file with the repaired file? If no, the repaired file will not be saved.

Yes  No

If you answer Yes, the repaired file will replace the damaged file and will have the same filename as the original damaged file. In the manifest of the container, “Preservation: Repaired File” will be added to the file section for the file.¹

View the file SAMPLE1.DBF to ensure that it has been repaired.

2.6.4 Convert Files in Obsolete Formats to Current or Standard Formats

An archivist may encounter a file with an obsolete format for which there is not a viewer. They should consider converting the file to a current or standard format for which there is a viewer. For instance, there is no viewer for WordPerfect

¹ It might be useful to add an Annotate Record feature that gave the archivist the opportunity to enter notes about any records in the database that might have been lost, or to note that all records appear in the repaired database.
Notebooks. However, Quick View Plus displays them as WordPerfect 4.2 documents as shown below.
Version 3.1 of the APT provides the capability for an Archival Institution, such as the Bush Presidential Library, to implement a Digital Migration (or Conversion) Policy that specifies whether files in a particular format should be converted to another format, and if so what that format should be. The policy is implemented by associating a file type with an application for converting files of that type to files of another file type. This feature is described in section 2.6.5.4.

With the file system open in the APT that contains the file to be converted, select *Preservation* from the *Activity* drop-down menu.

Step 1: Highlight the filename of the file to be converted.
Step 2: From the *Action* drop-down menu, select *Convert*.

The following Screen appears.

```
Program to Convert from WP notebook to DBF
By JEP for Doble & Co., Inc.

Getting number of fields...
Analyzing file, please wait.... 116/13
... TMP.DBF, created...
Getting screen format, This may take a few seconds ...
Writing screen...
Moving records from NB to DBF ...
New Checking for Extra Fields... 117
Creating final file and appending from Temporary one...
... sample.DBF, created...
Finished with the conversion of sample.nb

Files created: sample.DBF
sample.SCR
```

C:\PROGRA~1\ARD6CA~1\CONVER~1\MPNB2D~1>
The file is converted to a file of the format specified by the Digital Migration Policy. In this case, it is converted to a dBase III database and a DOS Extended ASCII Text File for the screen image of the notebook. It will be given the original filename and extension followed by a dot and the extension of the new file type placed in the file system, and associated with the original file. "Preservation: Converted File" will be added to the manifest for the file section of this file.

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Path</td>
<td>SAMPLE.NB.SCR</td>
</tr>
<tr>
<td>File Type</td>
<td>DOS IBM Extended ASCII Text</td>
</tr>
<tr>
<td>Length</td>
<td>1390 Bytes</td>
</tr>
<tr>
<td>Last Modified</td>
<td>10/8/2006 6:19:34 PM</td>
</tr>
<tr>
<td>SHA-Digest</td>
<td>079A835700876FD0FF3B1846845E94DCCFB54D57</td>
</tr>
</tbody>
</table>

The files SAMPLE.NB.SCR and SAMPLE.NB.DBF are shown below.
2.6.5 Associate File Type with Application

Before the preservation actions can be taken on a file, the File Type must be associated with a preservation utility. This is accomplished by choosing and **Assoc. FileType with Application** under the Options menu then choose the type of application to associate with the File Type. Extractors and Viewers can be associated with a File Type under any activity. Password Recoverers, Decrypters, Repairers, and Converters can only be associated with a File Type while in the Preservation Activity. These applications are associated with a File Type in the same way as described earlier for viewers and extractors.
2.6.5.1 Associating a File Type with a Password Recovery Utility

You can associate a file type for encrypted files with a password recovery utility by selecting **Associate File Type with Application > Password Recoverer** from the Option drop-down menu. A dialog box similar to the following will be shown.

This dialog box shows the types of encrypted file types that are recognized by the file type identifier. The file type Microsoft Word 6.0 Encrypted Document has been highlighted. Notice that it does not have a MIME Type or Decrypter associated with it. There is a program called msofpass that will recover the password of this file type. You can associate that program with this file type by highlighting the name of the file type and selecting Edit. The following screen is displayed.
The Password Recovery programs are on the path C:\Program files\Archival Processing Tools\Password Recovery. The path to the program can be found by selecting Browse.

The program that we need is the file RecPWMSOFFICE.bat. It is a DOS batch file that calls the program MSOFPASS.EXE with the filename of the file whose password needs to be recovered. That batch file is in the same directory as the program MSOFPASS.EXE.
When you select Open, you are returned to the Edit dialog box with the path entered into the Password Recoverer field.

In the APT Version 3.1, there are Password Recovery programs for the following file types.

<table>
<thead>
<tr>
<th>File Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microsoft Excel 5.0 Encrypted Workbook</td>
</tr>
<tr>
<td>Microsoft Excel 95 Encrypted Workbook</td>
</tr>
<tr>
<td>Microsoft Word 6.0 Encrypted Document</td>
</tr>
<tr>
<td>Microsoft Word 8.0 Encrypted Document</td>
</tr>
<tr>
<td>Microsoft Word for Windows 95 Encrypted Document</td>
</tr>
<tr>
<td>Paradox 3.5 Encrypted Database</td>
</tr>
<tr>
<td>Paradox 3.5 Encrypted Indexed Database</td>
</tr>
<tr>
<td>Paradox 4.x Encrypted Database</td>
</tr>
<tr>
<td>Paradox 4.x Encrypted Indexed Database</td>
</tr>
<tr>
<td>Paradox 5.x Encrypted Database</td>
</tr>
<tr>
<td>Paradox 5.x Encrypted Indexed Database</td>
</tr>
<tr>
<td>Paradox 7.x Encrypted Database</td>
</tr>
<tr>
<td>Paradox 7.x Encrypted Indexed Database</td>
</tr>
<tr>
<td>WordPerfect 5.1/5.2 Encrypted Document</td>
</tr>
<tr>
<td>Zip Encrypted Archive</td>
</tr>
</tbody>
</table>

@Echo off
"C:\program files\Archival Processing Tool\Password Recovery\msofpass\msofpass.exe" %1
Pause
2.6.5.2 Associating a File Type with a Decrypter

When the password for an encrypted file has been recovered, the encrypted file can be decrypted using the office software application that created the file. However, it is necessary that the file type of the encrypted file be associated with the related software application. You can associate file types with software applications for decrypting a file by selecting Associate File Type with Application > Decrypter from the Option drop-down menu. A dialog box similar to the following will be shown.

![Dialog box for associating file types with Decrypter](https://via.placeholder.com/150)

This dialog box shows the types of encrypted file types that are recognized by the file type identifier. A number of legacy software applications that were used to create encrypted files and that can decrypt these files have been collected and installed in the directory C:\Program Files\Archival Processing Tool\Decrypters. In addition, there are some office applications installed on the PERPOS computer platform, such as Microsoft Excel and Microsoft Word, that can decrypt current and legacy encrypted files.

To associate a file type for an encrypted file with a Decrypter, highlight the name of the file type and select Edit. The following dialog box will be displayed.
You can browse to find the path to the filename of the decrypter. In this case, the decrypter, EXCEL.EXE, requires a DOS batch file DEXCEL.BAT to decrypt an Excel Workbook.

"C:\Program Files\Microsoft Office\Office\DEXCEL.BAT" %1 %2 %3

When you select OK, the original dialog box will reappear with the path to the decrypter displayed at the bottom.
In the APT Version 3.1, there are software applications that are able to decrypt the following types of encrypted files.

<table>
<thead>
<tr>
<th>File Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microsoft Excel 2000/XP Encrypted Workbook</td>
</tr>
<tr>
<td>Microsoft Excel 4.0 Encrypted Workbook</td>
</tr>
<tr>
<td>Microsoft Excel 4.0 Encrypted Worksheet</td>
</tr>
<tr>
<td>Microsoft Excel 5.0 Encrypted Workbook</td>
</tr>
<tr>
<td>Microsoft Excel 5.0 Encrypted Worksheet</td>
</tr>
<tr>
<td>Microsoft Excel 95 Encrypted Workbook</td>
</tr>
<tr>
<td>Microsoft Excel 97 Encrypted Workbook</td>
</tr>
<tr>
<td>Microsoft Word 10.0 Encrypted Document</td>
</tr>
<tr>
<td>Microsoft Word 6.0 Encrypted Document</td>
</tr>
<tr>
<td>Microsoft Word 8.0 Encrypted Document</td>
</tr>
<tr>
<td>Microsoft Word for Windows 95 Encrypted Document</td>
</tr>
</tbody>
</table>
2.6.5.3 Associating a File Type with a Repair Utility

Some files that are damaged can be repaired. To do so, it is necessary to associate the file types of damaged files with a repair utility.

Under the Options drop-down menu, you can associate a file type with a repair utility by selecting Assoc. File Type with Application > Repairer.

A dialog box similar to the following will be shown.

This dialog box shows the types of user-created files which might be damaged and for which there may be repair utilities. To associate a file type with a repair utility, select the file type and then click the Add button. If you need to edit or remove an association, use the Edit or Remove buttons respectively.
utility, highlight the name of the file type and select Edit. The following dialog box will be displayed.

![File Type Information](image)

You can browse to find the filename of the required repair utilities. Most of the repair utilities supplied with the APT will be found in the directory C:\Program Files\Archival Processing Tool\Repairers. Select the filename. If there are any required arguments or switches, enter them into the Repairer Field of the dialog box.

![File Type Information](image)

When you select OK, the original dialog box will reappear with the path to the repair utility displayed at the bottom.
In version 3.1 of the APT, there are file recovery utilities for the following file types:

<table>
<thead>
<tr>
<th>File Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>dBase III database</td>
</tr>
<tr>
<td>dBase III database with dbt</td>
</tr>
<tr>
<td>FoxPro Database with fpt</td>
</tr>
<tr>
<td>Zip Archive</td>
</tr>
<tr>
<td>WordPerfect 5.0 Document</td>
</tr>
<tr>
<td>WordPerfect 5.1/5.2 Document</td>
</tr>
</tbody>
</table>
2.6.5.4 Associating a File Type with a Conversion Utility

This option implements an Archival institution's Digital Migration Policy. Under the Options drop-down menu, you can associate a file type with a conversion utility by selecting Assoc File Type with Application > Converter.

A dialog box similar to the following will be shown.

This dialog box shows the types of user-created files for which decisions need to be made as to whether files of these types should be preserved in their original format or converted to a current or standard format. To associate a file type with a conversion utility, highlight the name of the file type.
Select Edit. The following dialog box will be displayed.

You can browse to find the filename of the required converter. Most of the converters supplied with the APT will be found in the directory C:\Program Files\Archival Processing Tool\Converters. Select the filename. If there are any required arguments or switches, enter them into the Converter Field of the
dialog box. For instance, a DOS Batch file named WPNB2DBF.BAT has been created with the following content.

```
@Echo off
"C:\Program Files\Archival Processing Tool\Converters\WPNB2DBF\NB4.exe" %1
Pause
Exit
```

The program NB.exe converts a WordPerfect notebook to a dbase dbf file and the screen associated with the notebook to a text file with extension.scr. It is in the Converters directory. You can browse to find that bat file. It requires an input file parameter %1.

When you select OK, the original dialog box will reappear with the path to the converter displayed at the bottom.
There are approximately 90 file types that are identified by the file type identifier that one might want to convert to other formats. Version 3.1 of the APT has conversion utilities for converting files of the following file types to the target file types.

<table>
<thead>
<tr>
<th>Source File(s)</th>
<th>Target File</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ami Professional Document</td>
<td>PDF</td>
</tr>
<tr>
<td>ASCII 7-bit Test</td>
<td>PDF</td>
</tr>
<tr>
<td>DCA-RFT Document</td>
<td>PDF</td>
</tr>
<tr>
<td>DOS IBM Extended ASCII Text</td>
<td>PDF</td>
</tr>
<tr>
<td>Microsoft Excel 2.x Worksheet</td>
<td>PDF</td>
</tr>
<tr>
<td>Microsoft Word 6.0 Document</td>
<td>PDF</td>
</tr>
<tr>
<td>Microsoft Word 8.0 Document</td>
<td>PDF</td>
</tr>
<tr>
<td>Microsoft Word 9.0 Document</td>
<td>PDF</td>
</tr>
<tr>
<td>Microsoft Word 10.0 Document</td>
<td>PDF</td>
</tr>
<tr>
<td>Microsoft Word for Windows 95 Document</td>
<td>PDF</td>
</tr>
<tr>
<td>MS Rich Text Format Document</td>
<td>PDF</td>
</tr>
<tr>
<td>MS Word for DOS Document</td>
<td>PDF</td>
</tr>
<tr>
<td>MS Word for Windows 2.0 Document</td>
<td>PDF</td>
</tr>
<tr>
<td>Windows Write Document</td>
<td>PDF</td>
</tr>
<tr>
<td>WordPerfect 5.0 Document</td>
<td>PDF</td>
</tr>
</tbody>
</table>
2.7 Reviewing Files in a Container

The Freedom of Information Act provides for public access to holdings of Federal government records. However, it exempts some kinds of information from disclosure. Similarly, the Presidential Records Act provides for public access to Presidential Records, but restricts disclosure of some kinds of information. Access to private papers donated to an archive by individuals or organizations may be restricted due to the conditions of the donor agreements.

Archivists must review Presidential records, Donor records, and Federal records for these access restrictions. Archivists need assistance in reviewing Personal Computer files and in recording the results of their review.

In systematic processing, a file system should be reviewed only after filtering, arrangement, and preservation activities have been performed (if needed). The file system that is loaded will be contained in a TAR archive file with metadata in a manifest indicating the file type and order of records in a folder.

2.7.1 The Review Menu

There are seven drop-down menus available from the main Review menu. The File, Activity, View, and Help drop-down menus are explained in the section Exploring a File System and are the same for every activity.

Action: This menu supports six actions:
- Unmarking a previously marked file,
- Opening a file because there are no restrictions on disclosure,
- Closing a file because there are restrictions on disclosure,
- Redacting a file because some of the information has no restrictions on disclosure but some parts do,
- Marking a file as a PRM (personal record misfile), and
- Marking a file for transfer out of the file system because it is not a record or PRM.

Options: In addition to associating a specific extractor or viewer with a specific file type and editing the list of document types, this menu supports editing the list of reasons withdrawn.

Tools: There is one utility in this drop-down menu. It automatically checks for restrictions in the selected record. Selecting this option currently brings up the “Not yet integrated” message box.
2.7.2 Reviewing a Record

You can review the e-records in a file system by viewing each one, reading it to determine whether there are any restrictions on disclosure, and then using the Action drop-down menu to open, close, or redact the record; or to mark a file as a personal record misfile (PRM), or for transfer to another location, for example, the Library. It is not possible to perform any review action on a record unless the file has been viewed. This rule prevents accidentally taking an action on the wrong file. The APT currently enforces this rule by checking that the file has been viewed during this session. If it has not, it is not possible to select the actions from the Action drop-down menu. If the file has been viewed, each of the actions is possible. A session is defined as the period between the time that a container is opened in the APT and the time that the container is closed or saved.

2.7.2.1 Opening a File to Public Access

Suppose you had viewed file 89030601.doc by double clicking on the file name. You have decided that this is a Presidential message released by the White House and previously published by the Office of the Federal Register, NARA. There are no restrictions on the disclosure of this record. It is a copy of a public document. It can be opened.

---

2 It might be useful to add a review action for situations when the archivist needs further guidance before making a review decision; something like “Pending” or “To Be Determined” that will highlight that the document is outstanding.
To open it, highlight the file name by clicking on it once. From the main review menu, select Action > Open.

The check in the box preceding the file name indicates that the file has been reviewed. The green document icon indicates that the file has been opened, i.e., there are no access restrictions. The check in the grayed out boxes preceding the parent folders along the path of the document indicate that some of the documents in the parent folders has been reviewed. Whenever all documents in folder are reviewed the folder will have a checked box that has not been grayed out. When all the documents in a container have been reviewed the checkbox preceding the container name, in this case 15014.tar will no longer be grayed out. This is a way to see at a glance whether the container has been fully reviewed. This is also a way to see at glance if top-level folders need to be expanded and their content documents reviewed.

2.7.2.2 Withdrawing a Record due to Restrictions on Disclosure

If there are FOIA exemptions, PRA restrictions, or donor restrictions on the disclosure of information in a record, user access to the record may be denied by withdrawing or closing the record.

Suppose you selected and viewed the file with filename 002.txt that was on the path “Counsel Lee Liberman\NOMINEE\LEIGHTON.”
After reading the record, you conclude that the entire record should be withdrawn because of the Presidential Record Act restriction on disclosure of records involving confidential advice to the President, in this case, a recommendation regarding a judicial nomination. To close this document, you must return to the main window. However, do not close the window that the record is displayed in, because you will need information from this record to enter withdrawal information. You can either

- Click the Minimize button on the title bar of the active document to minimize the document window to a button on the Windows taskbar.
Right-click an empty area of the taskbar, the bar that contains the Start button and appears at the bottom of the desktop. Click Cascade Windows, Title Windows Horizontally, or Tile Windows Vertically to arrange the APT Review window and the Viewer window.

To close (or withdraw) a record:

Step 1: Highlight the filename of the record by clicking on it once.
Step 2: From the main review menu, select Action > Close.

Step 3: The “Reasons Withdrawn” dialogue box will appear. Select the reason(s) for withdrawal and select OK, which will record these reasons for withdrawal.
There is a drop-down menu at the top of Reasons Withdrawn dialog box. This menu allows you to re-open the document if you need to view it again while deciding the proper reasons for withdrawal.

There are no FOIA or Miscellaneous restrictions, but in some cases, the reason for withdrawal will involve restrictions from those Tabs. If you select the second tab of the Reasons Withdrawn dialog box, you will see the FOIA exemptions.

The third tab on the Reasons Withdrawn dialog box is for miscellaneous restrictions. These include Donor Restrictions and Statutes associated with FOIA exemption b(3) Exempt by Statute.
Step 4: The “Withdrawal Info” dialogue box will then appear. Complete the required fields.

You should indicate the Document Type by selecting the downward arrowhead to the right of the Document Type field, and then selecting the document type from that list.
If document type of this document is not in the list, select **Add Document Type** and add the needed document type to the list.

As shown in the figure following Step 4, there are two menu options at the top of the Withdrawal Info dialog box. The two menus are the Edit and the View menu. As with the Reasons Withdrawn dialog box, you can re-open the document to help fill in the withdrawal information.

You can copy information such as correspondent’s name or title and chronological date from the record by highlighting information in the record and selecting **Edit > Copy** from the Quick View Plus Menu bar. Then place the cursor in the field in which you want to paste the information. To paste the information select **Paste** from the Withdrawal Info Edit drop-down menu. Alternatively, right click the mouse button and select **Paste** from the pop-up menu, or use the shortcut keys shown on the menu, which are Ctrl+V.
If you do not know the chronological date, enter “n.d.” in this field indicating, "not dated" (No quotes).

Step 5: Click OK. The check in the box to the left of the filename indicates that the file has been reviewed. The red document icon indicates that the file has been closed. The reason(s) for withdrawal and the closure information for the file are displayed in the right windowpane. If any of this information seems incorrect, you can edit it by highlighting the filename and selecting View > Properties.

2.7.2.3 Redacting Terms or Passages with Restrictions on Disclosure

Suppose that in reviewing a file displayed with the Quick View Plus Viewer, you discover some information that is exempt from disclosure under FOIA or restricted from disclosure under provision of the PRA. However, the document has significant content that is not subject to access restrictions. You might decide to redact those portions that are subject to access restrictions, close the original document and open the redacted document. The APT supports three types of redaction:

- Redaction of records converted to TIF document images
- Redaction of a Portable Document Format (PDF) copy of a document
- Redaction of a document converted to MS Word format
To view the redaction menus, select Redact from the Action drop-down menu. This will drop-down a third menu that allows you to choose the type of redaction you wish to perform.

2.7.2.3.1 Redacting Document Images

The second method of redaction requires that the document be converted to a Tiff image. Suppose that we were viewing the document with filename 057.txt in the “Counsel Lee Liberman\DISTCRT\WASH-WD” directory and deciding that it had an access restriction that we would like to redact.

Step 1: Exit the Quick View Plus viewer by selecting File > Exit QuickView Plus.

Step 2: The file previously viewed in Quick View Plus should still be highlighted. From the APT menu bar select Action > Redact > TIF Document Image.

The contents of the highlighted file will be converted to a TIFF (Group 4) multi-page format file and stored in a Windows temporary directory. The file will then automatically be loaded into a customized Imaging window. By default, thumbnails are shown for each page in a multi-page document.
Step 3: Select Zoom from the drop-down menu to size the displayed page. “Actual size” or “Fit to width” seem to be the best zoom option. Selecting “Scale to gray” from the View drop-down menu sometimes gives a sharper image for black and white text.

Step 4: To redact text, select “block rectangle” from the Redaction drop-down menu or select the black rectangle icon from the toolbar. Place the cross hair to the upper left of the text to be redacted. Click and hold down
the left mouse button as you block out the text to be redacted. If you make a mistake, click on the blocked out text or stamp and select the delete (X) symbol and the black rectangle will be removed.

Step 5: The reason(s) for redaction is indicated with a rubber stamp. Stamps can be selected from the Redaction drop-down menu or by selecting the “rubber stamp” icon. The paragraphs of the PRA restriction and FOIA exemptions are shown in a Reasons Withdrawn dialog box.

Step 6: Check the restrictions that apply to the redacted text, tables, or figures, and select OK.
Step 7: Place the rubber stamp near the redacted text to which the restrictions apply (in the margin or in a white area above the redacted text) and click the left mouse button. The identifier(s) for the restriction or exemption will be inserted at that point. Then select the pointer, or the identifier will be inserted at each point you click the left mouse button. If you make a mistake, click on the blocked out text or stamp and select the delete (X) symbol and the access restriction identifier will be removed.

Step 8: When all the text that has access restrictions has been redacted and annotated with rubber stamps, select Save on the File drop-down menu.

Step 9: Exit the Image Redactor by selecting the close button [X] at the upper right of the window or Exit from the File drop-down menu. A dialog box will be displayed for entering information about the closed original of the redacted document.
Step 10: Enter the information. You should indicate the Document Type by selecting the downward arrowhead to the right of the Document Type field, and then selecting the document type from that list. If document type of this record is not in the list, select Add Document Type and add the needed document type to the list.

You can copy information such as correspondent's name or title and chronological date from the document by highlighting information in the document and selecting Edit > Copy from the Quick View Plus Menu bar. Then place the cursor in the field in which you want to paste the information. Right click the right mouse button and select Paste.

If you do not know the chronological date, enter "n.d." in this field indicating, "not dated." (No quotes).

Step 11: click OK. In the Review Activity window, you will see two files with the same name, except one has the original file format extension. The other has the filename extension ".rdt." The red document icon indicates that the original file has been closed. The blue document icon indicates the redacted file. The reason(s) for withdrawal and the closure information for the file are displayed in the right windowpane. If any of this information seems incorrect, you can edit it by highlighting the filename and selecting View > Properties.

Archivists refer to the documentary form of a record as record type, not document type. This will be fixed in a subsequent version of the APT.
2.7.2.3.2 Redacting PDF Copies of Documents

The third method of redaction requires that the document be a Portable Document Format (PDF) records or that it be converted to PDF Format. It has the advantage that the PDF copy of the document more closely resembles the physical form of the original copy. Furthermore, the readability of the document is often better than that of a Tiff image.

With the document filename highlighted, one selects PDF Document from the Redact option of the Action drop-down menu. This calls Redax 4.0, which is a plug-in for Acrobat 6 and 7 that allows users to select content using the text select tool or by drawing boxes around it. Redax 4.0 permanently removes sensitive content from PDF documents and is a commercial product widely used for redaction of documents. It has been extended to include PRA restrictions as well as FOIA exemptions.
The selected document will be automatically converted to pdf format and loaded into Adobe Acrobat.

Redax adds a menu to the Adobe Acrobat workspace.
DRAFT

The reader is referred to the Redax User’s Guide as to how to perform redaction using Redax [Appligent 2006]. [PDF redaction has not yet been integrated into the Redaction action.]

2.7.2.3.3 Redacting MS Word Documents

This method supports redacting the text of a MS Word document [VA 2006]. This tool was included primarily to demonstrate the issues that arise in redacting MS Word documents, especially the problems of eliminating the redacted text and hidden metadata. This method of redaction should not be used, as it does not remove all metadata associated with a MS Word Document [NSA 2006].

This method of redaction requires that the document be an MS Word document or be converted to MS Word format. It uses MS Word macros. The tool supports only indication of FOIA exemptions, but could be extended to PRA or Donor Restrictions.

To redact text from a record:

Step 1: From the Quick View Plus menu, select Edit > Select All. This results in the whole document being selected.

Step 2: Select Edit > Copy. This results in the text being copied to the clipboard.

Step 3: From the Quick View Plus, drop-down menu select File>Exit Quick View Plus

Step 4: In the APT window, select Action > Redact > MS Word Document. This opens Word and pastes the text you copied into a document. This document has the same name as the original, except the file type extension is “.rdt.”

Step 5: In the new Word document, you will highlight the text that is to be redacted. Using the left mouse button, place the cursor to the left of where you want to begin highlighting. Depress the left mouse button and
keep the depressed as you move to the right highlighting the text. When you reach the end of the text to be redacted, release the left mouse button. The text will have a dark box over it.

Step 6: From the Word toolbar, select the button corresponding to the exemption that applies to the text. The highlighted text is replaced with the reason for redaction and a series of dots. The reason for redaction plus the number of dots equals the number of characters redacted.

Step 7: To add additional exemptions [such as (b)(3) and (b)(6)], select some of the dots behind the first exemption, and select the next exemption button.

Step 8: If you make a mistake, you can use the “UNDO redaction” button in the Word document. This will restore the last redaction to the original text. If two or more exemptions have been inserted in the same place, use the “Undo 2nd” button to remove the second and subsequent redactions, then use the “UNDO redaction” button to undo the last one.

In the Review Activity window, you will see two files with the same name, except one has the original file format as its extension. The other has ”.rdt.” The one
with the original file format is blue, indicating redaction. The one ending in “rdt” is red, indicating this text has been redaction and access to it has been closed.

2.7.3 Marking for Transfer

Upon reviewing a file, an archivist might determine that the file needs to be transferred to some other collection. For instance, a copyrighted software application might need to be transferred to the library, or a system or software application file not created by an individual or office might need to be preserved in another file. For example, the following file in the Bush Public Paper Files is not a document created by a person using Word Perfect to create a document related to their primary business responsibilities.

This document is included with a WordPerfect Installation. It was not blocked during filtering, and needs to be removed from the container of records. To mark a file for transfer to a library of software applications and documentation:

Step 1: Highlight the filename file you wish to transfer.

Step 2: From the review menu bar, select Action > Mark for Transfer.
The checkmark preceding the filename indicates that the file has been reviewed. The document icon painted gray indicates that the file has been marked for transfer. The access property of this file shown in the right windowpane also indicates that the file has been marked for transfer.

The Description Activity of the Archival Repository Tool, supports removing files that are marked for transfer from a container and placing them in another container (See Section 2.8.4)

### 2.7.4 Mark a Record as a Personal Record Misfile

A personal record found filed with Presidential records is referred to as a Personal Record Misfile (PRM). It should be removed from the Presidential records. Suppose that in reviewing records, the archivist encounters the following records in the series "George H.W. Bush's Files."

---

**Miss Ashley Walker Bush**

c/o Mr. and Mrs. Neil M. Bush

Denver, Colorado 80216

Dear Ashley,

On this the first day of your life, your old grandfather sends you his love. Today was the day after my Savings and Loan proposal; the day of my visit to Capitol Hill to see a lot of Congress members; 2 days before my speech to the nation--but on this day of your birth, I'm thinking of you. You have 2 great parents, an older sister who will teach you and brother who will protect you. You have grandparents who love you a lot already. Welcome, welcome to this big loving family--I am a happy Gampy because you're here.

Devotedly,

George Bush

---
Suppose also that because this is a letter to his granddaughter and parents, and does not concern Presidential duties, the archivist concludes that this is a Personal Record Misfile. The archivist can close the viewer, and select PRM from the Action drop-down menu.

The result is that the right windowpane shows that the type of access is PRM (also included in the manifest of the container). The box alongside the filename receives a checkmark and the icon alongside of the filename is colored yellow.

In the Description Activity of ART, files that are marked as Personal Record Misfiles can be removed from a container and placed in another container (See Section 2.8.5).

2.7.5 Reversing Review Actions by Unmarking
An archivist may make a review decision and later decide that they are not certain of the correctness of the decision. A review decision can be reversed by highlighting the filename of the opened, closed, or redacted file, or the filename of a file marked PRM or marked for transfer, and then selecting Unmark from the Action drop-down menu. This changes the status to unreviewed. The check will be removed from the checkbox and the document icon will be white.

The result of unmarking the file is that there is no longer an access property of Open in the right windowpane and in the manifest. If the file had been Closed or Redacted, all closure information would be removed and the Access property would be removed. In addition, the box is no longer checked and the icon alongside of the filename is now white.

2.7.6 Changing a Review Action to Another Review Action
Currently, a file does not have to be unmarked to change its access status. So long as a reviewed file has been viewed in the current session, it is possible to highlight its filename and select a different action from the Action drop-down menu, and its status will be changed to that new action.

It is necessary to distinguish changing a review decision before a Record Series or FOIA collection is made available to the public and updating a review decision in conjunction with a FOIA appeal or expiration of a PRA restriction or FOIA exemption. Currently, the APT review activity does not make this distinction. This distinction can be made, by checking whether or not a reference copy has been made for a record series or a FOIA collection. If a reference copy has been made, any changes to review decisions are updates. Updating review decisions is discussed in section 2.7.9.

### 2.7.7 Viewing File Properties

If you highlight a filename and select from the Menu bar **View > Properties**, you will see in the files properties whether there are access restrictions. The value of the Access Properties of a File that has been opened is Open.

An alternative way to view a file's properties is to highlight the file name, right click on the mouse, and select properties or to select the Properties button when the file is highlighted.

The File properties of a file that has been closed or redacted include the reasons for withdrawal or redaction and other withdrawal information.
2.7.8 Saving a Reviewed Container

Often you will not be able to review an entire file system in a single session with the APT. You can save your partial results by selecting File>Save. Moreover, it is recommended that you periodically save your work in case there is a system failure that might cause you to lose your work.

When a file is reviewed, if it is contained in a folder, the box in front of the folder is automatically shaded and contains a check mark. When all the files in a folder have been reviewed, as indicated by a check mark in the box in front of their file name, the shading in the box in front of the folder containing the files is removed and the box remains checked. Since all of the folders have checks in front of them and are not shaded, the entire file system has been reviewed.
2.7.9 Updating Previous Review Decisions

An archivist may need to update, rather than reverse, a previous decision. Examples include when the record has been declassified, opened on appeal, and at the end of the 12-year PRA restriction period, when only FOIA exemptions apply.

2.7.9.1 Opening to Public Access a Previously Withdrawn Record

To open a previously withdrawn record, highlight the filename of the closed record, view and review the record to be sure that it is the record to be opened, and then exit the viewer. With the file name still highlighted, select Open from the Action drop-down menu. The attributes Previous Access and Previous Reasons Withdrawn will be added to the manifest and will appear in the right windowpane that have the values Closed, and the previous values of the Reasons Withdrawn. The Attribute Access will now have the value Open. [Not yet implemented]

2.7.9.2 Redacting a Previously Withdrawn Record

To redact a previously closed record, highlight the filename of the closed record, view and review the record to be sure that it is the record to be redacted. Select Redact from the Action drop-down menu. Redact the record using the image or PDF redactor and fill in the Withdrawal Information. When you save the results, a redacted copy will appear in the file system in the left windowpane that is linked to the withdrawn copy of the record. The attributes Previous Access and Previous Reasons Withdrawn will be added to the manifest and will appear in the right windowpane that have the values Closed, and the previous values of the Reasons Withdrawn. [Not yet implemented]

2.7.9.3 Re-redacting a Previously Redacted Record

To re-redact a previously redacted record, highlight the filename of the previously redacted copy, and view it. Either print the redacted copy or minimize the viewer. Highlight the filename of the withdrawn master copy of the same record and view it. Exit the viewer and with the filename of the withdrawn master copy still highlighted, select Redact from the Action drop-down menu. If you printed a copy of the previously redacted record, you can read it to see what portions of the record might no longer need to be redacted. If you minimized the viewer of the redacted record, you can place the mouse pointer at the bottom of the Desktop in a blank area near the Quick Launch tool bar, right click the mouse and select Tile Windows Horizontally. Both copies of the
record, the previously redacted record and the original to be re-redacted, will be shown. Redact the copy of the original. The withdrawal information is the same as for the previously redacted record. What may differ is the initials of the archivist reviewing the record and the date of withdrawal. Save the redacted copy and exit the Redactor. The filename of the re-redacted copy of the record will be the original filename and extension plus the extension .rdt.rdt indicating that it is a second redaction. It will appear in the file system with the previously redacted copy and the original closed copy of the record.

The attributes Previous Access and Previous Reasons Withdrawn will be associated with the previously redacted copy with replacing the Access and Reasons Withdrawn attributes and having the values of those attributes. [Not yet implemented.]

2.7.10 Edit Reasons Withdrawn

For a variety of reasons, e.g., changes in access legislation, archival coding conventions, or new donor restrictions, it may be necessary to edit the codes and descriptions of reasons for withdrawal or redaction. The reasons for withdrawal can be edited by selecting Options > Edit Reasons Withdrawn.

This dialog box allows you to select a Category of access restrictions (PRA, FOIA or Miscellaneous). Then you can select a Restriction or Description and edit it. You can also select a row and use the Delete button to delete the Access Restriction. Finally, you can add a Restriction Code and description using the Add Button.
The following dialog box shows FOIA exemptions from disclosure including subcategories of b(7) Law Enforcement Investigations. One uses the right scroll bar to see other b(7)-b(9) exemptions. Relevant paragraphs of statutes for b(3), Exempt by Statute, are defined in the Misc(ellaneous) dialog box.

The following dialog box shows an example of a donor restriction and some of the relevant paragraphs of statutes where there are exemptions from disclosure per FOIA exemption b(3).
2.7.11 Edit Document Type List

You need to indicate the document type of a closed document. A short list of document types is included with the initial installation. You can see this list, edit it and add to it by selecting Options > Edit Document Type List.

2.7.12 Checking for Access Restrictions

An experimental prototype tool is being developed that will check a file (or all the files in a container) Access Restrictions. The tool will interpret the document and apply a set of rules to determine whether is a Presidential record or a Personal/Political Record Misfile (PRM). If it is a Presidential record, it will attempt to determine whether the entire document or only passages or phrases have PRA restrictions or FOIA exceptions to public access. The restrictions currently checked are

a(2) - Appointments to Federal Office
a(5) - Confidential Advice
a(6) or b(6) - Personal Privacy

A window will popup indicating whether no restrictions were found, the record seems to be a Personal Record Misfile, or seems to have one of the restrictions listed above. This tool will be integrated with the review activity under the Tools drop-down menu.
There are many potential benefits of such a tool.

1) It might identify an access restriction not identified by the reviewer, thus reducing the risk of opening a record or passage of a record whose access should have been restricted.

2) It might be used as a tutor during training of review archivists.

3) Novice reviewers could use the tool to check their work.

4) The tool might provide additional evidence in case a reviewer's judgment was uncertain, or point out uncertainties, where the reviewer thought the decision was certain.

5) It might give a rapid review to records responsive to a FOIA request to estimate the workload in terms of the number of restrictions and types of restrictions likely to apply.

6) It might estimate which unprocessed electronic record series are likely to have many restrictions, and which are likely to have few or no restrictions. The systematic review of those with no or few restrictions could result in more records being opened to the public at an earlier date.

7) Experienced reviewers are eventually promoted or retire and NARA and Presidential Libraries lose their expertise. The tool might accumulate review knowledge so that the knowledge resource is not lost.

8) The tool will support PRA and FOIA review decisions for Presidential records, so it would also support review of Federal Records for FOIA exemptions.

9) Since most states have Open Record Acts, state records need to be reviewed for access restrictions before disclosure to the public. The technology might be transferred to support archivists performing review of state government records.

10) Although the records being considered in this study are unclassified records, the technology might transfer to declassification review.

2.8 Describing a Record Series
In section 2 of this Manual, it was explained how the Description Activity supported the description of the arrangement of records at the collection and office levels. This section describes the Description of records at the record series level. The Description Activity does not yet support description of records at the file unit (folder or directory) or item (e-record) levels. However, for records that are withdrawn due to access restrictions, the record type, correspondent(s), subject, and chronological date are captured and saved in a withdrawal record.

Each of the other activities supported by the APT, arrangement, preservation and review, capture information at the record level about archival actions. In the Arrangement activity, if a file is rearranged, the original path as well as the current path is captured. In the Preservation activity, if a file is converted to another format, decrypted, or repaired, information is captured as to the action, by whom it was performed, and the date it was performed. In the Review activity, closure and redaction decisions are captured as well as by whom it was decided and the date the decision was made. Other review decisions such as opening a file, marking a file as a PRM and marking it for transfer are captured, but the person making the decision and the date of the decision are not explicitly captured. The name of the archivist making those decisions can be inferred from the name (or initials) of the archivist assigned to the Systematic case or the FOIA case. Physical information about records, such as file format, is determined at the time of accession and is stored in the manifest of accessioned containers.

Description of Record Series is the last step of systematic processing. This means that filtering, arrangement, preservation, and review activities for the record series should be complete, and the containers associated with the record series should have been moved from the archivist's work area back to Holdings.

1. In ART, select *Description* from the *Activity* drop-down Menu. Select the name of an office and the series that you wish to describe.
2. Find one of the containers you have processed, highlight its name, and open it by selecting Open from the File drop-down menu. The contents will be displayed in the Archival Processing Tool.

3. Place the mouse pointer in a blank area at the bottom of the screen and click the right mouse button. Select one of the three options—Cascade, Tile Horizontally or Tile Vertically. For instance, if Tile Horizontally were chosen, the Archival Processing and Description windows would be tiled as shown below.
In the APT, you can open the folders and files to read the contents of files in order to determine the content of the record series. You can then switch to the Description Activity of the Archival Repository Tool to describe the record series.

4. To describe the record series, highlight the name of the record series and select Properties from the Edit drop-down menu of ART. The Edit Series dialog box will be displayed. There are two ways to open the Edit Series dialog box — Right clicking the mouse while it is sitting on a series or selecting the Properties button while a series is selected.
The description at the record series level includes information about the content (Scope and Content Note, time span), intellectual access (arrangement, e.g., alphabetical, chronological), and physical description (Extent or Volume in bytes).

NARA's Lifecycle Data Requirements Guide [NARA 2005] provides the following guidance for describing record series in a Scope and Content Note: "At each level of description, usually indicate the level being described by using an introductory phrase such as 'This series consists of' or 'This file unit contains.'"

"The Scope and Content Note should contain information about: who created the archival materials, who the archival materials are about, (i.e., to what person or organization they relate,) who contributed to the production or authorship of the archival materials and what their relationship is to the activities documented; what the archival materials are generally about, what the main topics or subjects mentioned are, and what unusual or historically significant topics are mentioned in addition to the main topics presented; where the action or events take place, what specific geographic places or areas are mentioned; how the information is recorded, what record types are included, and how the information is presented."
5. To determine the *Inclusive Dates* of files in the file system and the *Extent* of electronic records in the Record Series, select the *Container Properties* option from the *View* drop-down menu in the Archival Processing Tool. You will see a display similar to the following.

The dates, and number of files and number of bytes can be copied to the Edit Series dialog box by highlighting the dates or dates, clicking the right mouse button selecting copy, and then pasting into the corresponding fields of the Edit Series dialog box.

If there is more than one container in the record series, it is necessary to view each container, select Container Properties and determine the Earliest and Latest File Dates, and accumulate the number of Bytes of each container.\(^4\)

\(^4\) The next version of the APT will determine the Earliest and Latest File Dates and Total Volume in Bytes for all containers in the record series. In addition, the tool for estimating the
For electronic records, volume is measured in files or bytes (or kilobytes, megabytes, gigabytes). It is recommended that one enter the abbreviation of the unit (ft, files, or bytes, KB, MB, GB). The exact number of bytes of the files in this container is 4291. It is recommended that bytes only be used if the number of bytes is less than 1 KB, i.e., 1024 bytes and that the volume be approximate, not exact. For instance, in this case one would enter 4.2 KB. Alternatively, one can also enter the number of files. The PERPOS Project staff is working on a method to estimate as well the number of pages of electronic records in a container. To determine the inclusive dates and extent of all containers in a record series, it is currently necessary to accumulate these manually and enter the cumulative results. [Note: It is planned to add the capability to automatically determining the inclusive dates and extent of the files in multiple containers in a record series.]

6. Arrangement in this case refers to the arrangement of folders in the record series. If there were more than one folder, the arrangement would probably be Chronological.

7. One selects OK, to save the record series description.

8. For each container in the series, the archivist should highlight the container name, select Properties from the Edit drop-down menu, select the Processing Status, and indicate that the container has been described. As with the record series, there are the same two ways to access the Properties dialog box.

In a prior version of ART, if one highlighted a series name, one could add file unit (folder, directory) titles. It is planned to automatically generate for containers the file units in the container and show these in the finding aid for a processed and opened record series.

2.8.1 Moving a Container to a Different Record Series

After reviewing the contents of a container, an archivist may determine that the record series that it was associated with at the time of accession is incorrect. To move the container, the archivist must know the name of the office and the title of the record series to which they want to move the container. To move the number of pages in a FOIA collection will be extended to estimate the number of pages in a record series. The data element for Extent (or Volume) of the record series will be extended to include a data elements for volume in bytes and Volume in pages.
container, in the Description Activity, highlight the name of the container and select Move Container from the File drop-down menu. A dialog box similar to the one used to add a container during Accession would be displayed.

Select the name of the office to which you want to move the container. The record series in that office can then be selected, or if it is a new record series title, it can be entered. Then select OK. You will be returned to the Description Window where it will be observed that the container has been moved. [Not yet implemented]

2.8.2 Creating Reference Copies for Public Access

The Bush Presidential Library separates closed and opened files in a record series of paper records by creating shadow folders and keeping the closed records and their shadow folders in a secure area, while the opened records are kept in another.

Many of the record series processed using the APT and stored in the Archival Repository will not only contain open records, but closed records, original copies of redacted records, and redacted copies. It may also contain files marked for transfer to the Library and Personal Misfiled Records. You must create a Reference copy of record series for public access that does not include closed records, originals of redacted records, records marked for transfer, or PRMs.

A Reference Copy for a series of records cannot be created until the Processing Status of all containers in the series is Filtered, Arranged, Preserved, and Reviewed. In addition, a check is made that the record series has a Scope and Content Note, Inclusive Start and End dates, and a value for Extent.

1. In ART in the Description Activity, browse the tree of organizations, organizational units, and series titles until you find the record series for which a reference copy is needed. Highlight it.

2. From the File drop-down menu of ART, select Make Ref. If not all the conditions for making a Reference Copy are met, the Make Ref option will not be available.

Copies will be made of each opened record, redacted record, transfer sheet, and withdrawal sheet in a container. These will be placed in a new container with the same filename as the original. This will be done for each container in the record series. The new containers will be stored in the Reference directory. A screen similar to the following will be displayed.
Each container in a record series has a Status attribute that can be “InProcess,” “Processed” or Blank. If a reference copy has been made, the string “Reference copy has been made” will be displayed under the Status line as shown in the figure above.

3. To verify that the reference copy of a container has been properly created, highlight the container name, and then select Open Reference Copy from the File drop-down menu. The APT will display the reference copy as read only.

2.8.3 Create Finding Aid
Yet to be created is a function to create the finding aid for a record series. This finding aid can be used in the reference room for finding containers associated with offices and record series.

2.8.4 Transfer Files Marked for Transfer
During the Review Activity, some of the copies of containers may have been found to contain non-records, and should have been marked for transfer. To remove the files so marked:

1. In ART in the Description Activity, browse the tree of organizations, organizational units, and series titles until you find the container that contains files marked for transfer. Highlight the name of the container and then select **Open Master Copy** from the **File** drop-down menu. The file system structure of the processed record series will be displayed in the APT.

2. In ART, from the File drop-down menu, select **Transfer Marked Files**.

The files marked for transfer will be removed from the container and placed in a container with the same file name as the original, and it will be placed in the Library directory. If there is a file in the Library that already has that container name, the files marked for transfer will be appended to that file. A transfer sheet is created for transferred files and included in the original container. [Not yet implemented]

2.8.5 Transfer Personal Misfiled Records
During the Review Activity, some of the master copies of containers may have been found to contain Personal Misfiled Records (PRM). To remove the PRM files:

1. In ART in the Description Activity browse the tree of organizations, organizational units, and series titles until you find the container that contains PRMs. Highlight the name of the container and then select **Open Master Copy** from the **File** drop-down menu. The file system structure of the processed record series will be displayed in the APT.

2. In ART, from the File drop-down menu, select **Transfer PRM**.

The files marked PRM will be removed from the container and placed in a container with the same filename as the original. The container will be placed in the Personal Records directory. [Not yet implemented]

2.8.6 View Processed and Opened Record Series
This view includes a catalog of e-collections that have been opened for public access and includes the capability to browse the catalog and open containers associated with systematically processed record series. The catalog looks like the left panel of the Description Activity, except it includes just those series that have been systematically processed and for which a reference copy and finding aid have been created. When a series title is highlighted, the finding aid for that series is displayed in the right panel. The archivist can browse the finding aid seeing the folder titles within containers. If they see a folder title for which they want to view the records, they highlight the name of that container in the left panel and open it. The container is opened in a version of the APT that only allows them to explore the contents of the container. [Not yet implemented]

2.8.7 View Software Library

This view includes a catalog of containers of operating system software and office application software that was blocked during filtering of the contents of containers. It may also include software and other non-records that were marked for transfer during review and transferred to the container. When a container id is highlighted, it is opened in the explore activity of the APT. This view is currently implemented by opening the APT in Explore mode, browsing to the C:\Library directory and selecting a container id of a container to be opened.

2.8.8 View PRM Containers

When PRM containers are created, the manifest will include the name of the person who created, received, or used the personal records. Currently those containers can be viewed by opening the APT in the Explore Activity, browsing to the C:\PRM directory, selecting a PRM container id, and opening the container. In the future, a catalog of the PRM containers could be displayed in the left panel of ART, and when a container was highlighted, the metadata from the container manifest could be displayed in the right panel. [Not yet implemented]

2.8.9 Public Access to Opened Collections

A reference tool is needed for the Presidential Libraries that could be used by researchers in the reference room. This tool would include the features described in section 2.8.6, View Processed and Opened Record Series. It would also include the features described in section 3.9.2, View FOIA Collections. It could also contain the capability to search for records in the Library of opened
Presidential e-records. It could also include the capability to view the Software Library described in section 2.8.7

Such an interactive catalog and access tool is easy to construct from existing components of ART and the APT. A Reference Catalog would have a structure similar to the Repository Catalog in the Description Activity of ART and the access and viewing capability correspond to the Explore Activity of the APT.
3. FOIA Processing

When a request for records under the Freedom of Information Act is received from a researcher, a FOIA Case number is assigned and it is logged into the Presidential Library Database using a Bush Presidential Library Reference Request Form. A paper copy of this form goes into a yellow folder labeled with the requestor’s last name and the assigned FOIA case number. Then all applicable databases are searched for records relevant to the request. The results are logged on a Bush Presidential Library Reference Search Form and a paper copy of this form is placed in the same yellow folder. The requestor is notified of the volume of records (in pages) that is responsive and an estimate is made of the time needed to process them. An archivist will then review just those records that are relevant, not considering an entire record series or container, but often just the contents of some folders within several containers. While the archivist might also perfect the arrangement and perform preservation actions on those records reviewed, they often do not fully preserve, arrange and describe the contents of an entire container or record series. The requestor is notified of the availability of the requested records. This process is called FOIA processing.

3.1 FOIA Processing Dataflow

The diagram below illustrates the dataflow of FOIA Processing activities that are supported by the PERPOS tools. The numbered, labeled circles are activities. The two parallel lines represent data stores. The rectangles represent entities external to the PERPOS system, and the labeled, directed arrows are data flows. Stepping through the diagram in the numerical sequence of the activities, one sees the dataflow.
Holdings consist of accessioned containers of e-records. The PERPOS Tools first support FOIA processing by providing the capability to Index Holdings (Activity 1). This function creates an index of all the terms in the textual records in Holdings.

The Bush Presidential Library Database is a Microsoft Access database of tables, forms, reports, and queries. The database includes an accession register, location register, folder title list, Reference Request Form and a Reference Search Form. The Presidential Library Database is external to the PERPOS Tools. To perform a search of the Staff Member and Office electronic records in Holdings, a FOIA case is created (Activity 2) that has the same FOIA Case Number as that on the Reference Request Form. An archivist translates the FOIA request into a FOIA Query, which is submitted to FOIA Search (Activity 3) that returns a result set of pointers to records that are relevant to the FOIA query. The result set is associated with the FOIA case. The requestor is notified of the volume of records (in pages) that is relevant and an estimate is made of the time needed to process them (not shown in the data flow). When an archivist is ready to begin work on the FOIA case, they check out one of the containers associated with the case (Activity 4) to an archivist's work area. The archivist then uses the review activity of the Archival Processing Tool (APT) to review just those records in the container that are relevant to the query, not considering an entire record series or container (Activity 5). Review actions
include opening a record for public access; withdrawing or redacting a record because of access restrictions; marking a record as a Personal Record Misfile or marking a file for transfer to the Library because it is a nonrecord. When the records that are in the result set for a container are reviewed, the archivist checks the container back into Holdings (Activity 6). Once all containers with relevant e-records are reviewed, the Archival Repository Tool is used to make a FOIA Reference Collection (Activity 7) and Finding Aid (Activity 8). Records Marked for Transfer are transferred to the Library (Activity 9). The Public Access System has not yet been developed.

The following sections describe each of these support functions in detail and show the user interface.

### 3.2 Index Holdings

Before one can search for electronic records relevant to a FOIA case, one must create an index of e-records in the Repository (Holdings). Only containers in Holdings that have been filtered will be indexed. Filtering is the process of removing operating system files, office software application files, system or software documentation, or sample application files from accessioned records. These files are not Presidential records and indexing them would result in an archivist having to consider non-records during the review process. Filtering of containers of record series is discussed in section 2.4 of this Reference Manual. Archive files that have not been expanded, and password-protected files, image, and audio files will not be indexed.

To index the contents of all accessioned and filtered containers, select **FOIA Case Mgmt** from the Activity drop-down menu. Next, the archivist selects **Index** from the Tools drop-down menu.
The message “Indexing Containers …” will appear in the status bar at the bottom of the screen. Depending on the number of containers and number of files in each container, this process will take from a few minutes up to several hours. If unfiltered containers are encountered while indexing, a dialog box will appear that lists the OAID of the containers that are not indexed because they are not marked as filtered.

The archivist can get a printout of these container ids by selecting the Print button at the bottom of the dialog box. Select the OK button at the bottom of the form to complete the process.

It is necessary to re-index the containers in the holdings area in the following cases.

1. Since the index was last created, new containers have been accessioned and/or filtered.
2. Since the index was last created, non-record files have been transferred from containers to the Library during the arrangement, preservation, or review process.
3. Removal of Personal Record Misfiles (PRMs).

In the first case, records relevant to a FOIA search may have been added to Holdings, but will not appear in a result set because they are not indexed. In the second case, a FOIA search may return a pointer to a record that is no longer in its original container. In the third case, a FOIA search may return a pointer to a record that is no longer in its original container because upon review it is found to be a PRM and is subsequently transferred out of the container. It is not
necessary to re-index holdings before each search of holdings, if these conditions have not occurred.5

3.3 FOIA Case Management

The Archival Repository Tool is used to manage the processing of electronic records relevant to a FOIA Case. To do this, select FOIA Case Mgmt from the Activity drop-down menu shown below.

The archivist then selects Add FOIA Case from the Edit drop-down menu.

This causes the Add FOIA Case dialog box to appear.

---

5 When archivists attempt a FOIA search, they are notified when they need to reindex holdings. A container table has a field that indicates the date a container was last indexed (or is unindexed), and the date when files were last transferred out of the container, including the date of removal of PRMs.
The Archival Repository Tool automatically fills in the Archivist field with the user name and the Date field with the system date. In the dialog box, the archivist enters the FOIA Case number from the Bush Presidential Library Reference Search Form and selects OK. There is a check to be sure that there is not a collision with an existing FOIA case number. The Scope and Content Note should not be filled out until the FOIA Case is complete and it is time to create a FOIA Reference copy.⁶

After OK is selected, the archivist is returned to the FOIA Case Mngmt activity with the FOIA Case Number highlighted in the left pane and the FOIA Case properties in the right pane.

---

⁶ A check could be made as to whether review of the FOIA case was complete. If not, it would not be possible to enter the Scope and Content Note.
3.4 Performing a Search

To search for records relevant to a FOIA request, the FOIA Case Mngmnt Activity must be the current activity. To begin a FOIA Search, the archivist must first select a FOIA Case. When a FOIA Case is highlighted, the archivist selects Search from the Tools drop-down menu.

This starts the FOIASearch application. The FOIASearch’s main screen appears with the FOIA Case No in the screen title.
The archivist translates the FOIA request from the Reference Request Form into an Oracle Text with a word query. For instance, if the request had been for "Any materials related to Iraq and Kuwait," the Oracle Text query might be "Iraq & Kuwait." The archivist enters the search criteria into the Query box.

Oracle Text uses the Boolean operators AND (&), OR (|) and NOT (~). Parentheses are used for grouping expressions. For example '(George | Barbara) & Bush' is for all records containing the words either 'Barbara' or 'George' and the word 'Bush'.

A root word prefixed with a dollar sign ($), e.g., $broadcast, will find all documents containing its root word (stem) or derivatives, e.g., broadcasts, broadcasting, or broadcaster. The EQUIV operator (=) is used to indicate that two or more words are equivalent, for instance (91=1991).

Using the ACCUM(ulate) (,) and weight (*) operators, one can increase the score for documents that match a query by weighting terms differently. For instance, in searching for documents related to the Clarence Thomas nomination to the Supreme Court, the expression

(justice, judge, Supreme Court*5, Clarence Thomas *10)

will increase the score of the term Supreme Court by 5 times and the term Clarence Thomas by 10 times. This signifies that documents related to Clarence Thomas and Supreme Court are most relevant to the query. The ACCUM operator gives the highest scores to documents that contain the terms within the scope of the operator; for instance, ACCUM (dog, pet, Millie) will give the highest score to documents that contain all three terms.
One can search for terms that are in close proximity with the NEAR operator. For example, to find all documents where Soviet is within six words of Revolution, the following query would be issued.

\[ \text{NEAR((Soviet, revolution), 6)} \]

The default and maximum value for the NEAR operator is to search for terms separated by no more than 100 words. In conjunction with Boolean operators, the NEAR operator constrains the scope of a query. Used with the section-searching operator WITHIN, the NEAR operator can constrain the search to predefined zones (sentence, paragraph, HTML sections). Examples of FOIA requests and the corresponding Oracle Text queries are shown in Appendix C.

Once the query is entered, select the **Search** button.

Once a search has been performed, the **View Document**, **Limit Results**, and **Save Results** buttons are enabled. The result set is displayed as a table with the column headings Rank, OAID, and FilePath. The Rank is the relevance value assigned by Oracle Text Search. The OAID is the unique identifier the archivist assigned to the container when it was accessioned. The FilePath is the path within the container to the file (record).

To determine whether a record is relevant to the FOIA request, an archivist performing the FOIA search is able to view individual records with the search terms highlighted. The View Document screen also displays the OAID along with the path within the container in the title of the screen. To do this, the archivist selects the row in the table containing the path to the file he/she wishes to view and selects the **View Document** button.
The archivist can limit the results of the query to any result with a rank greater than one of their choosing. To do this, the archivist selects the **Limit Results** button.

However, in this case, it is not necessary to limit the result since the second record is relevant. The archivist can remove all results that do not have rank greater than the seven by selecting the **OK** button. The archivist can also choose to change the search criteria by simply changing the query and selecting the
search button. The old results will be discarded and the results of the new search will be displayed.

When the archivist selects the *Save Results* button at the bottom of the Search screen, the query is returned to the Archival Repository Tool along with the query results and is associated with the appropriate FOIA Case for further processing. At any time, the archivist can leave the FOIASearch application by selecting the *Cancel* button at the bottom of the dialog box. If this button is selected, neither the query nor the result list will be returned to the Archival Repository Tool and associated with the FOIA case.

If the FOIA results search is saved, both the query and the results will be displayed in the right windowpane of the Archival Repository Tool when the FOIA Case is highlighted. The containers that contain the relevant records will be listed under the FOIA Case. The results will be displayed in a list with the headings 'Status', 'OAID', and 'Path'. The status of a record can be 's', 'f', or blank. The status will be 's' for a record that has been systematically processed, so will not need to be reviewed. It will be 'f' for a record that has been processed for a previous FOIA case, so will not need to be reviewed. It will be blank for a record that is unprocessed, so will need to be reviewed. Under the result list, the properties *Processed Files*, *Unprocessed Files*, *Processed Pages*, and *Unprocessed Pages* are listed. The Archival Repository Tool calculates the *Processed Files* and *Unprocessed Files* properties. These property values are updated each time a container is returned to holdings, even if the container was checked out for another FOIA case or systematic processing.

---

7 Currently, the FOIA Case Management activity and FOIA Search capability only support saving the result set of one query for a FOIA Case. For complex FOIA requests or for follow-up searches, it will be necessary to have the capability for additional queries and result sets to be associated with a FOIA case. It is planned to accomplish this by extending the FOIA Case No to have subcases, e.g., 2006-0020-F[1], 2006-0020-F[2], and creating FOIA search criteria and results sets for each subcase. (See Appendix D, FOIACase Table)

8 It is not clear that this information (Status, OAID, and Path) needs to be displayed. It is needed in the results table by ART and the APT to identify the containers that have records that need to be reviewed, the specific records that need to be reviewed, and their status for preparation of the finding aid. An archivist processing paper records in folders needs such a list because they have to manually find particular folders in particular containers. But that manual capability is not needed in this FOIA processing environment.

9 This status information is used to calculate the number of processed and unprocessed files, which gives an archivist an idea of the amount of work needed to fulfill the FOIA request. This information is also used, after review of the unprocessed records, to prepare the Finding Aid for the FOIA collection.
3.5 Estimating the Number of Pages to be Reviewed

An archivist responding to a FOIA request must enter into the Reference Search Form the estimated number of pages to be reviewed (unprocessed) and those that are already reviewed (processed). Files of the same size but different file format can correspond to very different numbers of pages. For example, a Microsoft Excel file may take up a relatively small amount of file space, but generally converts to a large number of pages. Whereas, an image file may have a large number of bytes but correspond to a single page. Another factor is the size of the header that some file formats have. An experimental method has been developed for estimating the number of pages of processed and unprocessed records relevant to a FOIA case.

To use this method, highlight the FOIA Case number, and then select Estimate Pages from the Tools drop-down menu.
When the Archival Repository Tool has estimated the number of pages in the FOIA Case, it updates the currently selected FOIA Case and displays the estimates in the right pane.

3.6 Print Reference Search Form

The kinds of information appearing on a Library Reference Search Form can be printed for inclusion in the Yellow folder labeled with the Requestor's Name and FOIA Case Number. With the FOIA case number selected on the FOIA Case Management Window, select Print Ref Search Form from the File drop-down menu. [This option has not yet been implemented.]
3.7 Reviewing Records for FOIA Cases

To begin the review of records relevant to a FOIA case, an archivist opens the Archival Repository Tool. The archivist then selects FOIA Case Mngmt from the Activity drop-down menu. The archivist selects the FOIA case to be processed. The FOIA case’s properties appear in the right panel. If FOIA search is performed, the query criteria and the result list will appear in the property list. Lists of containers in which there are records relevant to the request are displayed beneath the FOIA Case in the left pane. Click on the ‘+’ symbol beside the FOIA case, to view the list of containers. The archivist then selects a container to review. If the container is already checked out for processing by another archivist, the Archivist, Processing Type, and Case No properties will appear in the container’s property list in the right panel. Containers can only be checked out to one archivist at a time.

3.7.1 Checkout a Container

If the container is not already checked out to another archivist, the archivist can check it out by selecting Checkout Container from the File drop-down menu. The archivist who is checking out the container must be the archivist who created the FOIA case.

A dialog box will appear to allow the archivist to indicate where the checked out container should be placed. Recommended practice is to put it in the Work Area directory in a folder titled with the initials or name of the archivist.
When the container is checked out, the container's manifest is modified to allow the Archival Processing Tool to know which records in the container belong to the FOIA case and which archivist has checked it out. Then, a copy of the container with the modified manifest is placed in the location indicated in the dialog box. The original container with its original manifest is kept in the repository as a backup.
When save is selected, the archivist is returned to the FOIA Case Mngmt screen. The Archivist who checked out the container, the Processing Type under which the container was checked out, and the Case No of the case under which it was checked out are displayed in the right pane of the Archival Repository Tool. Containers can only be checked out to one archivist at a time.

### 3.7.2 Open a Container for Review

When a container is checked out, only the archivist who checked out the container or a user with administrator privileges can open the container for processing. Anyone else who opens the container will be opening a read-only copy that remains in the Holdings area. The archivist opens the container using the Archival Processing Tool. The archivist selects Review from the Activity drop-down menu, and selects Open from the File drop-down menu.

![Open Container](image)

From the archivist's folder in the Work Area, the container can be opened.

---

10 The archivist who checks out the container can also open the working copy of the container from inside the Archival Repository Tool by selecting Open from the Files drop down menu.
The filenames of the records associated with the FOIA case and the folders that contain the records appear in **boldface**.

### 3.7.3 Review Records Related to FOIA Case

To review a record, the archivist opens the file whose file name is boldfaced and it is displayed in a viewer. The archivist reads the document to determine whether it is a non-record, a Personal Record Misfile (PRM), or a Presidential record. Review of Presidential records, whether as a part of FOIA Processing or Systematic Processing, is to determine whether there are PRA restrictions or FOIA exemptions that apply.

The record shown below is a memorandum from President Bush to Brent Scowcroft, his National Security Advisor. PRA restriction a (5), Confidential Advice, would have applied to this record, had the President not waived that restriction for this record which he included in his book.\(^\text{11}\)

\(^{11}\) All the Best, George Bush: My life in letters and other writings, 1999, pp 490-491.
By selecting an action from the Action drop-down menu, an archivist can Open, Close, or Redact the record, or Mark it as a Personal Record Misfile or Mark it for Transfer to the Library (because it is a non-record). Suppose that an archivist decides this record should be closed because of PRA a (5). He selects Close from the Action drop-down menu.
If a record is closed, the archivist must indicate the reason for withdrawal, in this case PRA a (5). There are no FOIA exemptions.

The archivist must also enter withdrawal information—record type, correspondent's name(s), subject or title, and chronological date. Some of this information can be copied from the record and pasted into the withdrawal form. The Archivist's username is automatically captured, as is the date of withdrawal.
A box next to the filename of the record is color coded to indicate the type of access—green for open, red for closed, blue for redacted. Yellow indicates a PRM and grey indicates marked for transfer. The access type, access restrictions, and withdrawal information for a record are displayed in the right panel of the window and are stored in a manifest file in the container.

A check in a grey box next to a folder (or container) symbol indicates that some of its records have been reviewed. When all of the records in a folder (or container) have been reviewed, the check will appear in a white box.

If a record is redacted, a document image is created for the record, and the archivist can block out text in the image, select the reason for withdrawal and stamp an area near the text with the reason for withdrawal.
The review support capabilities are essentially the same as were developed for systematic processing, except that only the records relevant to the FOIA query need to be reviewed. The support capabilities for review of records are described detail in section 2.7 of this Manual.

3.7.4 Add to or Remove Record from a FOIA Case

Two additional actions have been added to the Action drop-down menu in the Review Activity. These actions are Add to Case and Remove from Case. It is possible that some records were not returned in the result set of a FOIA search because the records were not indexed. This situation can occur when files have not been extracted from an archive file, a file is password protected or is an image or audio file. It is also possible that a record read by the reviewer could be determined to be relevant to the FOIA request, but was not relevant to the FOIA query.

3.7.4.1 Add a Record to a FOIA Case

While viewing files in a container that is related to a FOIA case, an archivist can view files that are not in the result list of the FOIA case. If after viewing a file that is not in the results list of the FOIA case, an archivist decides that the record is relevant to the FOIA case, he can add the file to the results list of the current FOIA case container by highlighting the filename and selecting Add to Case from the Action drop-down menu.

The Case Document property will be set true for the file and the filename will appear in boldface indicating that it is a member of the FOIA case. If there are any other versions of the record, e.g., redacted or in a different file format, they will have the same GroupID, and they will appear in boldface.
3.7.4.2 Remove a Record from a FOIA Case

After viewing a record, an archivist may decide to remove it from the result list of the FOIA case. This might occur because the file was an operating system or software application file, program documentation or a sample text file that should have been filtered and transferred to the Library.\(^\text{12}\)

The following figure shows record “EO\121.doc” with the file name and the path boldfaced and the Case Document property set “True.”

To remove this record from the FOIA Case, the archivist first highlights the filename and selects Remove from Case from the Action drop-down menu.

\(^{12}\) This function may not be needed. If the only reasons for removing a file from a case are that it is a nonrecord or a PRM, then the archivist should mark it for transfer or as a PRM. Having done so, the results list is not modified, but when the FOIA reference copy is created, it will not contain files that are in the results list that are marked as PRMs or are Marked for Transfer.
The Case Document property will be set false, the file name will not appear in boldface, and the Case Document property will not be displayed. If there were any other versions of the record, e.g., redacted or in a different file format, they will have the same GroupID, and their boldface will disappear.

During review, when there is both an original record and a copy of the record that has been converted to a different file format, if either record is marked closed, both records are marked closed and given the same reason for closure. When any record in the group is marked Open, PRM, or for Transfer, then all copies of records with the same GroupID are marked with the same type of access. When a record is redacted, both the original and any copy converted to a different file format are marked closed.

3.7.5 Check in a Container

When an archivist has finished reviewing all the records in a container that are part of a FOIA Case, he checks the partially reviewed container back into the archival repository (including both reviewed and un-reviewed records). This is
accomplished in the Archival Repository Tool by selecting Checkin Container from the Files drop-down menu.

When a container is checked back into the repository, the FOIA case under which the container was checked out and all other FOIA cases that contain one or more of the same records are updated. Their Processed Files and Unprocessed Files properties are updated. The Case’s Processed Pages and Unprocessed Pages properties are not automatically updated. It is the responsibility of the archivist to update these properties by selecting the Estimate Pages tool. The Status field, which displays Results, is also updated for this case and any other case that contains the reviewed records. Any additional records, created during redaction or conversion, are added and any path properties that were changed due to changes in arrangement are modified. Temporary information (Archivist’s name and results list for this container) is removed from the manifest and the reviewed container and manifest replaces the container and manifest in holdings.
When a record is removed from a FOIA case, all other records with the same GroupID are removed from that case, but no records are removed from any of the other cases. The records removed from one FOIA case are still relevant to other FOIA cases to which they belong.

Once all pertinent updates have been made, the temporary information that was added to the container’s manifest is removed and the container replaces the original container that had been kept as a backup in the repository.

When an archivist has processed a container, he can proceed to the next container that is not checked out to another archivist. All the relevant records in some of the containers may have already been processed. If this is the case, the container does not need to be checked out. The archivist can tell if a container has any unprocessed records associated with the FOIA case by looking at the FOIA case’s result list. The status of a record is listed beside the filename of the record. They can stop review at any time and resume review later by selecting the relevant FOIA case number from the Archival Repository Tool.

### 3.7.6 Undo Checkout

The working copy of a container may become corrupted or be inadvertently deleted. In this case, the archivist does not want to, or cannot, check the
container back into the repository. In fact, the archivist should undo the whole check out process and start over. This can be accomplished in the Archival Repository Tool by selecting *Undo Checkout* from the *Files* drop-down menu.

There is a check that the archivist who is undoing the check out is the archivist who checked out the container. The container reverts to the copy that remained in holdings. It is the responsibility of the archivist who undoes the check out to delete the copy in his work area.\(^\text{13}\)

The *Archivist*, *Processing Type*, and *Case No* are removed from the container’s property list and will not be displayed in the right pane of the Archival Repository Tool and the container becomes available for checkout once more.

\(^{13}\) Searching the work area for the name of the container could eliminate this secondary responsibility. If it was found, then it could be deleted. If not found, it had been inadvertently deleted.
3.8 FOIA Case Description

When an archivist has reviewed the records in all the containers and folders that are associated with a FOIA Case, the FOIA Case should be described. To describe the FOIA Case an entry is made in the FOIA Case’s Scope and Content Note property. To accomplish this, select the FOIA Case in the FOIA CASE Mngmnt Activity, and then select Properties from the Edit drop-down menu.

![Edit FOIA Case](image)

Once the Scope and Content Note has been entered, select the OK button. The archivist will be returned to the FOIA Case Mngmnt Activity screen. The Scope and Content Note will be displayed with the other properties in the right pane. After a description has been entered into the Scope and Content Note field, the Make FOIA Ref menu option on the File drop-down menu will be enabled.

3.9 Arrangement and Finding Aid of a FOIA Reference Container

The last step in FOIA Processing is creating the FOIA Collection Reference Container and Finding Aid. The Archival Repository Tool creates them at the same time. The arrangement of the information in the Finding Aid is the same as the arrangement of the records in the Reference Container.

The contents of the FOIA Case Reference container and the Finding Aid are arranged following the Bush Presidential Library guidelines for arranging a FOIA
Collection. The FOIA Collection Reference container is arranged by Collection with the Bush Presidential Records: Staff Member and Office Files being the first collection. Under each collection, the offices are arranged alphabetically. Under each office, the series are arranged alphabetically by the staff member’s last name. Under each series, the containers are arranged numerically by OAID number (smallest to largest). Within OAID number, the directories and records are in the order in which they appeared in the original container, unless there was some rearrangement. The collection of Quayle Vice Presidential Records are arranged next in alphabetical order by office, then alphabetically by staff member name, and then in OAID number order.

If no staff person is identifiable for a series, the series are arranged alphabetically after the last staff person for an office. If Federal records are processed in response to a FOIA request, they are arranged by Record Group at the very end of the FOIA Collection.

3.9.1 Make a FOIA Reference Container and Finding Aid

To create a FOIA Reference container select the FOIA Case in the FOIA CASE Mngmnt Activity, then select Make FOIA Ref from the File drop-down menu. If the Make FOIA Ref menu option is not enabled, then not all files in the result set have been reviewed or the Scope and Content Note for the FOIA Case has not been created.
The FOIA Reference Container will contain opened records, redacted records, FOIA markers for relevant records that are in another FOIA Collection or in a Reference container for a systematically processed record series, withdrawal sheets and transfer sheets. It will not contain closed records, originals of redacted records, PRMs, or transferred records.

When the FOIA Case is selected, the case containers are shown under the FOIA Case and the contents of the Scope and Content Note along with query and the result list are displayed in the right pane. The only difference is that the Open FOIA Ref Copy and the Open FOIA Finding Aid option of the File drop-down menu are enabled when a FOIA Case is highlighted that has a Reference Container.

3.9.2 View FOIA Collections

This view includes a catalog of FOIA collections that have been opened for public access and includes the capability to browse the catalog and open FOIA collection reference containers. The catalog is similar to the left panel of the FOIA Case Management Activity, except it includes just those FOIA cases whose review is complete and for which a reference copy and finding aid have been created. When a FOIA Case id is highlighted, the finding aid for that FOIA collection is displayed in the right panel. The archivist can browse the finding aid seeing the folder titles and filenames. If they want to view the records, they highlight the name of that container in the left panel and open it. The container is opened in a version of the APT that only allows them to explore the contents of the container. While this view is not yet implemented, the following two sections describe how it is currently possible to open a FOIA collection Finding Aid and FOIA Reference Container.

3.9.2.1 Open a FOIA Collection Finding Aid

To open the Finding Aid of a FOIA Case, first select the FOIA Case. Then select Open Finding Aid from the File drop-down menu.
The Finding Aid for the FOIA Case is created as an HTML document so that it can be published on the Bush Presidential Library Web Site.

**3.9.2.2 Open a FOIA Reference Container**

To open the FOIA Reference container, select the FOIA Case. Then select *Open FOIA Ref Copy* from the *File* drop-down menu. This will open the FOIA Reference Container inside the Archival Processing Tool.
The “wds” extension on the highlighted filename above indicates that it is a “Withdrawal Sheet.” In right pane, the Access property is shown as “Closed.” The grayed out Activity buttons on the tool bar indicate that the container is open as read only. It is still possible to open a record in a read-only viewer. To view the Withdrawal Sheet, select QuickView from the drop-down View menu.

Withdrawal Sheet

This item has been withdrawn.

FOIA Case: 2006-0001-F
OAID: 15004
Name: MEMO0013.txt
Reasons: a(1); b(1)
References


[VA 2006] MSWord redactor freeware from the Veterans Administration.

http://www.va.gov/foia/redactor/
Glossary

application software
Programs designed for the end user, such as word processors, database systems, spreadsheet programs, graphics programs, communications software, and games.

archive
(v) 1. To copy files to a storage medium for purpose of backup. 2. Put into an archive
(n) 1. A disk, tape, or file that contains files that have been backed up. 2. A depository containing historical records and documents
(adj.) A file attribute that marks files that have been modified since the last backup.

archive file
A collection of data files that have been packaged together, for example, a zip, arc, tar, or rar file.

blocked file
A file that does not pass through a filter because it is an operating system office application file.

closed file
A file unit or series to which access is restricted or denied.

cluster
To be able to keep track of the data on a partition, the file system divides each partition into small blocks called clusters. A cluster is the smallest area, which can be allocated from the disk and its size depends on the file system and on the size of the partition.

cluster tip
The unused area at the end of the last cluster allocated by a file is called the cluster tip (or the slack space). This unused area is present in most files because space can be allocated only as cluster sized blocks and the contents of the file rarely completely fill all allocated clusters.
compressed file
A file in which redundant information has been eliminated and thus is shorter and requires less storage space than the original file.

computer virus
A self-replicating computer program that spreads by inserting copies of itself into other executable code or documents.

corrupted file
A program, text, or data file that has been altered accidentally by a software or hardware failure. Hence, it is unreadable by the hardware, or readable but not executable, displayable or playable by the software.

distiller (Adobe)
Commercial conversion software that accepts source formats, such as Postscript or MSWord, and outputs a document in Portable document format (PDF)

file blocking filter
A file filter in which files that match the pattern are blocked from passing through the filter. Those files that do not match the filter are said to be passed files.

file decryption
The process of converting a file that has been encrypted back into its original form.

file format
A format for encoding information in a file. A file format defines a file type. The file format specifies first whether the file is a binary or ASCII file, and second, how the information is organized.

file format conversion
Changing the contents of a file from one file format to another.

file system
The system of an operating system that can be used to organize and keep track of files. For example, a hierarchical file system is one that uses directories to organize files into a tree structure. The file system, such as FAT (File Allocation
Table) and all its variants (FAT12, FAT16, FAT32) and NTFS (New Technology File System), keeps track of the filenames, dates, size and the physical location on a partition of a disk.

file type
A category of digital files determined by file format (or filename extension) that is used or created by a software application and that is used to determine which program can be used to view the contents of the file.

file viewer
A program that enables you to read a file and display its contents. A WordPerfect 4.2 viewer, for example, enables you to read WordPerfect 4.2 files and display their contents.

hash function
A function (algorithm) that maps a bit-string of arbitrary finite length to a bit-string of fixed length referred to as a hash-code, hash-value, hash, or message digest. Hashing is used to index and retrieve items in a database because it is faster to find the item using the shorter hashed key than to find it using the original value. The hash-codes of a operating system and software application files that are not user-created files can be uses to determine that other occurrences of those files with the same hash-codes are also not user-created files. In addition to faster data retrieval, hashing is also used to encrypt and decrypt digital signatures (used to authenticate message senders and receivers).

integrity check
A mechanism to verify that the present state of data has not been tampered with or modified, often using check sums or hashing algorithms.

magic number
Special data located at the beginning of a binary data file to indicate its type to a system or application program. Under Unix, the system and various applications programs distinguish between types of executable file by looking for a magic number. Some magic numbers are, in fact strings, like the '!<arch>' at the beginning of a Unix archive file or the '%!' leading PostScript files. Sometimes referred to as a file signature.

MIME
(Multipurpose Internet Mail Extensions) The most common method for transmitting non-text files via Internet e-mail, which was originally designed for
only ASCII text. MIME encodes the files using one of two encoding methods and decodes it back to its original format at the receiving end. A MIME header is added to the file that includes the type of data contained and the encoding method used. The MIME "type" has become the de facto standard for describing files on the Internet. See MIME type.

MIME type
A file identification notation based on the MIME encoding system. The MIME type has become the de facto standard for identifying content on the Internet. For example, in order to identify the type of attachment sent in an e-mail message, its MIME type is embedded in the message header. Web servers send the MIME type to the requesting browser so that it can launch the appropriate helper application or plug-in. The MIME "Content Type" has a type and subtype separated by a slash; for example, text/plain and image/gif. The major types are application, audio, image, text, and video. Application refers to a variety of formats; for example, application/x-pdf refers to Adobe Acrobat documents, and application/octet-stream refers to an .EXE file.

non-records
Materials that are neither Presidential Records nor personal/political records. For electronic media, this includes computer operating system files, office application files (word processing, spreadsheets, data base management systems, help files), sample office application files, computer game files, program documentation, and reference e-publications such Webster's Dictionary.

operating system
Software on a computer that manages the operation of system resources and software application programs. Operating systems perform basic tasks, such as recognizing input from the keyboard, sending output to the display screen, keeping track of files, and directories on the disk, and controlling peripheral devices such as disk drives and printers.

overwriting pass
The number of overwriting passes determines how many times an area on the disk is to be overwritten.

partition
A hard disk can be divided into several logical drives called partitions. The size of the partition and the file system used determine the cluster size. Usually it is desirable to keep the cluster size small to reduce the amount of wasted (or slack) space on the partition.
passed file
A file that is not blocked by a file-blocking filter.

password recovery
The process or method of determining the password that was used to encrypt a file.

period length of PRNG
The length of a pseudorandom sequence; the amount of numbers that can be generated with a PRNG before the sequence starts from the beginning.

cryptographically strong pseudorandom number generator, PRNG
An algorithm that provides a sequence of numbers that appears to be random. All “random” data created by arithmetical means is called pseudorandom.

secure hash algorithm-1 (SHA-1)
A standard hash algorithm that maps a file (a strings of bits) into 60-bit hash code or message digest.

disk space
The space on a partition not used for storing data. Consists of cluster tip areas of the files on the partition and the available free space.
# Appendix A: APT Error Messages

<table>
<thead>
<tr>
<th>Error Message</th>
<th>Description</th>
<th>Corrective Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>APT Run-time Error 429 Active X Component Can't Create Object</td>
<td>If this error occurs, it will occur just after selecting the Archival Processing Tool (APT). This error can occur when there is a new version of the VBAAT.dll that has been copied in to replace an earlier version. There is incompatibility between the parameters of the prior version and the new version of the DLL.</td>
<td>It is necessary to unregister the DLL and then reregister it. Select 'Start' and then 'Run'. Browse to find VBAAT.DLL. &quot;C:\Program Files\Archival Processing Tool\VBATT.dll&quot; and select OK. Enter Regsvr32 –u in front of the quoted path to the DLL. Select OK. You should see the message: DllUnregisterServer in C:\Program Files\Archival Processing Tool\VBATT.dll succeeded. Remove the -u parameter from Regsvr32 command but leaving the same quoted path to the DLL, and select OK. You should see the message: DllregisterServer in C:\Program Files\Archival Processing Tool\VBATT.dll succeeded. Retry selecting the Archival Processing Tool.</td>
</tr>
<tr>
<td>ART Run-time Error 429 Active X Component Can't Create Object</td>
<td>If this error occurs, it will occur just after selecting the Archival Repository Tool (ART). This error can occur when there is a new version of the Vbaar\dll that has been copied in to replace an earlier version. There is incompatibility between the</td>
<td>It is necessary to unregister the DLL and then reregister it. Select 'Start' and then 'Run'. Browse to find the path to VBAAR.DLL. Select OK. Enter Regsvr32 -u in front of the quoted path to the DLL. Select OK. You should see the message: DllUnregisterServer in C:\Program Files\Archival Repository Tool\VBAAR.DLL succeeded. DllregisterServer in C:\Program Files\Archival Repository Tool\VBAAR.DLL succeeded. Retry selecting the Archival Repository Tool.</td>
</tr>
<tr>
<td>Archival Processing Tool\APT is not accessible</td>
<td>This message may occur after exiting the APT. The APT does not have full control of the Quick View Plus Viewer and a copy may remain open after exiting the APT. The APT attempts to erase all temporary files upon exit, and Quick View Plus may still have one of them open. This causes the error message.</td>
<td>There is a Quick View Plus Icon (magnifying glass) in the lower right-hand corner of the desktop. Point to it with the mouse, right click the mouse and select Exit Quick View Plus. Alternatively, ignore the error message. The APT erases the files when it reenters when Quick View Plus no longer has control of the files.</td>
</tr>
<tr>
<td>&quot;Bad Argument&quot; APT aborts</td>
<td>Error in File Type Identifier</td>
<td>Make a copy of the file system that was being filtered. Break the file system into two roughly equal parts and run the file type Filter on each part. Typically, the error</td>
</tr>
</tbody>
</table>
DRAFT

| Initializing Container Error: Must Define a Collection before adding a container to an Accession | ERROR Accession Activity of ART when attempting to associate a container with an accession. | The Description Activity of ART must be used to create collections and Record Groups before a container can be associated with an Accession. Select the Description Activity of ART and create collection names. |
| "Invalid Key" APT aborts | Error in File Type Identifier | See Error Message "Bad Argument" |
| Run time error '9' Subscript out of range APT Aborts | Error in File Type Identifier | See Error Message "Bad Argument" |
### Appendix B: File Types of Operating System & Application Software

The file types listed below are checked in the File Type Filter. They are file types of operating system or software application files that will never be created by users of the operating system and software applications.

<table>
<thead>
<tr>
<th>File Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ami Professional Macro</td>
</tr>
<tr>
<td>Ami Professional Template</td>
</tr>
<tr>
<td>DOS Batch Command Text</td>
</tr>
<tr>
<td>DOS Code Page Information</td>
</tr>
<tr>
<td>DOS Device Driver</td>
</tr>
<tr>
<td>DOS Hardware Configuration Text</td>
</tr>
<tr>
<td>DOS Help</td>
</tr>
<tr>
<td>DOS System Data Country</td>
</tr>
<tr>
<td>DOS System Data Keyboard</td>
</tr>
<tr>
<td>Empty File</td>
</tr>
<tr>
<td>OS/2 16-bit Dynamic Link Library</td>
</tr>
<tr>
<td>OS/2 Virtual Device Driver</td>
</tr>
<tr>
<td>Unix Shell Script Text</td>
</tr>
<tr>
<td>Windows 16-bit Dynamic Link Library</td>
</tr>
<tr>
<td>Windows 3.1x Installation Information Text</td>
</tr>
<tr>
<td>Windows 3.x Help</td>
</tr>
<tr>
<td>Windows 32-bit Dynamic Link Library</td>
</tr>
<tr>
<td>Windows 95 Installation Information Text</td>
</tr>
<tr>
<td>Windows Collect of Virtual Device Drivers</td>
</tr>
<tr>
<td>Windows Cursor Resource</td>
</tr>
<tr>
<td>Windows Icon Resource</td>
</tr>
<tr>
<td>Windows NT Installation Information Text</td>
</tr>
<tr>
<td>Windows Program Information File</td>
</tr>
<tr>
<td>Windows Program Initialization File Text</td>
</tr>
<tr>
<td>Windows Program Manager Group File</td>
</tr>
<tr>
<td>Windows Virtual Device Driver</td>
</tr>
<tr>
<td>WordPerfect Application Resource Library</td>
</tr>
<tr>
<td>WordPerfect Block</td>
</tr>
<tr>
<td>WordPerfect Column Block</td>
</tr>
<tr>
<td>WordPerfect Device Driver</td>
</tr>
<tr>
<td>WordPerfect Dictionary</td>
</tr>
<tr>
<td>WordPerfect Display Resource</td>
</tr>
<tr>
<td>WordPerfect Equation Resource</td>
</tr>
<tr>
<td>WordPerfect External Dictionary</td>
</tr>
<tr>
<td>WordPerfect Help</td>
</tr>
<tr>
<td>WordPerfect Hyphenation Code</td>
</tr>
<tr>
<td>WordPerfect Hyphenation Data</td>
</tr>
<tr>
<td>WordPerfect Hyphenation Lex</td>
</tr>
<tr>
<td>WordPerfect Keyboard Definition</td>
</tr>
<tr>
<td>WordPerfect Macro</td>
</tr>
<tr>
<td>WordPerfect Macro Resource</td>
</tr>
<tr>
<td>WordPerfect Mouse Driver</td>
</tr>
<tr>
<td>WordPerfect Overlay</td>
</tr>
<tr>
<td>WordPerfect Printer Resource ALL</td>
</tr>
<tr>
<td>WordPerfect Printer Resource PRS</td>
</tr>
<tr>
<td>WordPerfect Rectangular Block</td>
</tr>
<tr>
<td>WordPerfect Setup</td>
</tr>
<tr>
<td>WordPerfect Spell Code Rules</td>
</tr>
<tr>
<td>WordPerfect Spell Code Word List</td>
</tr>
<tr>
<td>WordPerfect Thesaurus</td>
</tr>
<tr>
<td>WordPerfect Vax Keyboard Definition</td>
</tr>
<tr>
<td>WordPerfect Video Resource</td>
</tr>
</tbody>
</table>
Appendix C: Oracle Text (Word Queries) for Examples of FOIA Requests

1. Documents pertaining to United States Information Agency (USIA), Voice of America (VOA) and China. (98-0001-F)
   $china AND ((({USIA}|united states information agency|u.s. information agency) OR ((VOA)|voice of america))

2. Documents pertaining to Radio Marti or TV Marti and broadcasting to Cuba. (98-0001-F)
   (radio marti) OR (tv marti) OR ($cuba AND $broadcast)

3. Documents pertaining to aid to nonpublic schools, e.g., tuition tax credits and vouchers during the George Bush Presidency. (98-0002-F)
   (aid=tuition=credit=credits=voucher=vouchers NEAR private=nonpublic=religious NEAR education=educational=school=schools)

4. Documents pertaining to the Domestic Policy Council. (98-0004-F)
   (domestic policy council)

5. I am requesting a copy of all documents which concern POW/MIAs in Southeast Asia, including Vietnam and Laos. I am a family member of an MIA.(98-0029-F)
   (southeast asia|vietnam|laos) AND ((({POW}|prisoner of war) OR ({MIA}|missing in action))

6. Records, briefings and reports held by the Bush Library which pertain to the bombing of Pan Am Flight 103 over Lockerbie, Scotland on December 21, 1988. (98-0034-F)
   (pan am|flight 103|lockerbie) AND (bomb|bombing|terrorist|terrorism)

7. Materials pertaining to the nomination of former Senator John Tower of Texas as Secretary of Defense. I am interested particularly in Tower's nomination fight for defense secretary. (98-0041-F)
   (nominate|nominee|nomination) AND (john tower|senator tower|(defense AND tower))

8. Materials pertaining to Human Immunosuppressant Virus or HIV, and Acquired Immune Deficiency Syndrome or AIDS. [HIV is actually Human Immunodeficiency Virus] (98-0091-F)
   ({HIV}|human immunodeficiency virus|{AIDS}|acquired immune deficiency syndrome)

9. Files dealing with the following subject matter: refugee / asylum procedures for Central America (e.g., El Salvador, Nicaragua, and Guatemala). (98-0097-F)
   ($refugee OR asylum) AND (central $america|$nicaragua|$guatemala|el $salvador)

10. Records relating to US military intervention against Iraq (i.e., Desert Shield, Desert Storm) (98-0099-F)
11. Records relating to US military intervention in Somalia (i.e., Operation Restore Hope). (98-0101-F)
(somalia NEAR military) OR (operation restore hope)

12. Records pertaining to US military intervention (or non-intervention) in Bosnia. (98-0102-F)
($bosnia)

13. All documents and materials that relate to the NATO Heads of State Summit at Rome on Nov 7-8, 1991. This is also referred to as a meeting of the North Atlantic Council at Rome on Nov 7-8, 1991. This includes, but is not limited to speeches and formal statements regarding the Rome Summit in general, and specifically, the NATO Strategic Concept presented at that summit. (98-0142-F)
(rome summit|heads of state summit|north atlantic council) OR ((NATO) AND strategic concept)

14. Documents or material dealing with March 14, 1991: Meeting between President Bush and French President Francois Mitterand. (98-0142-F)
(march,14=14th,91=1991,meeting,Mitterand,Bush) AND (Mitterrand AND {March})

15. Records and references relating to George Bush and Iran-contra independent counsel Lawrence Walsh. (98-0189-F)
(Iran,Contra,affair) AND (Walsh|Independent Counsel)

16. All records related to President Bush trip to Hungary during July 11-14, 1989 and his diplomatic efforts in Hungary. (98-0194-F)
Hungary AND ($trip|$visit|$diplomacy|diplomatic|July NEAR 89=1989)

17. Papers pertaining to the nomination of US Supreme Court Justice Clarence Thomas. (98-205-F)
(nominate=nominee=nomination,justice,judge,Supreme Court*5,Clarence Thomas*10,(Thomas NEAR nomination)*10) AND (Thomas) AND (supreme court)

18. A request for materials relating to the making, revising, adopting of the FY 1990 Federal Budget or what White House aids called the "budget deal of the century. (98-0251-F)
(FY=fiscal=budget) NEAR (90=1990)

19. A request for materials relating to or reflecting President Bush's concern for ethical conduct in the Federal executive, specifically, the creation, operations, and staffing of the Office of Government Ethics as a separate agency in October 1989. (98-0251-F)
(Office of Government Ethics)

20. I am interested in the various foreign and domestic policy experts that were assembled to assist President Bush assume power. I would like to request access to any materials
which link President Bush to the following think tank: Council on Foreign Relations. (98-0255-F)
(Council on Foreign Relations)

21. Documents related to the creation of a solution to the Savings and Loan crisis between the date of the Bush Administration coming into office until August 9, 1989 (the latter date being the passage of the Financial Institutions Reform, Recovery and Enforcement Act. (98-0356-F)
($saving (and) $loan) AND DOCNAME < 89080900

22. Documents concerned with US foreign relations with Pakistan. (98-0387-F)
($pakistan)

23. Documents concerned with nuclear proliferation with respect to India or Pakistan. (98-0387-F)
(nuclear=proliferation=nonproliferation) NEAR (India=Pakistan)

24. Information on U.S. efforts to convene an Arab-Israeli peace conference in 1991. The conference ultimately occurred in Madrid, Spain, and opened on October 10, 1991. The conference was orchestrated by the United States and included Palestinian representatives as well as officials from Egypt, Israel, Jordan, Saudi Arabia, Syria and a number of other states.

I am requesting information on the efforts to convene the conference and convince all relevant parties to attend. I believe this primarily took place between February and October, 1991 (Though U.S. discussions might have begun as early as late 1990).

One final note: according to Secretary Baker the decision to hold the conference in Madrid only came in the last few weeks before the conference was held. Thus, it is unlikely that most file headings or titles will refer to Madrid. (98-0497-F)
(peace,conference,Madrid,(October Near 91=1991)) AND ($Arab|$Israel|Israeli|Middle East)

25. In general the topic is NAFTA (North American Free Trade Agreement)
--Any business involvement in the passage of NAFTA - labor involvement.
--During Bush's Presidency, the evolution and history of how and why NAFTA came about.
--Other peoples, advisors involved.

In general, I am looking for both the internal process of how NAFTA passed and also external support/non-support of it. Mexico's point man was Jaime Serra de Puche. I can't remember the US's point man - it may be Robert Mossbacher's [Mosbacher]. (99-0062-F)
{(NAFTA)!North American Free Trade Agreement}

26. Request for records re: NASA, specifically. U.S.-Russia joint efforts including but not limited to Shuttle-MIR. (99-0093-F)
(States) NEAR (Soviet=Russia) NEAR (NASA=space)
27. I am working with former Defense Secretary Caspar Weinberger on his autobiography, and we are interested in any Bush presidential records relating to Secretary Weinberger - in general, but particularly regarding the pardon, which President Bush granted to Weinberger in December of 1992. (99-0099-F)
(caspar|weinberger) ACCUM (caspar weinberger)

28. I would like to file a Freedom of Information Act request to gain access to all documents that focused on the internal situation in South Africa between 1948 and 1994. I am interested in studies that examine the internal political dynamics of South Africa and attempt to predict the actions of the white government and the black opposition as they struggled over apartheid. (99-0103-F)
(apartheid,South Africa,internal,situation,political,struggle) AND ((apartheid) OR (South Africa))

29. Materials relating to the Bush Administration's role in supporting NIH research (including material on the Human Genome Project, and AIDS Research). (99-105-F)
({NIH}|National Institutes of Health) AND ($research|$funding)

30. Material relating to the Bush Administration's consideration of legislation to expand access to healthcare. (99-0105-F)
(health care) NEAR ($legislation=$expand=$access)

({UNCED}|earth summit) OR (Rio NEAR 92=1992)

32. Records related to the White House council on competitiveness (V. P. Quayle). (99-0129-F)
(competitive=competitiveness NEAR council=Quayle)

33. Documents regarding the decision to invade Panama on 12/20/89. I will not require any documents dated after 12/20/89. I would like any documents relevant to the decisional process. (99-0186-F)
(military=decision=situation=89=1989) AND Panama AND DOCNAME < 89122100

34. All documents that relate in whole or in part to President Bush and General Secretary Gorbachev's meeting in Malta on December 2-3, 1989. (99-0273-F)
(Gorbachev) AND (Malta|NEAR((November=December,89=1989),5))

35. Documents related to U.S. - Soviet relations, particularly, arms control negotiations - START I and START II (99-0302-F)
(((U.S.}|United States) AND ($Soviet|$Russia) AND (arms control)) OR (START I=II)

36. All materials pertaining to attempted coup against Gorbachev, 8/91 (99-0303-F)
($coup AND $Gorbachev)
37. All materials pertaining to breakdown of the Soviet Union, 12/91 (99-303-F)
   (($revolution$ NEAR ($Soviet=$Russia)) OR ((December NEAR 91=1991) AND ($Soviet=$Russia)))
   AND DOCNAME > 91119999

38. Documents and materials that pertain to early policy decisions on the administration's
    human rights approach in foreign policy. [Early is interpreted as 1989] (99-0461-F)
    (human $right$) AND DOCNAME < 90000000

39. Documents and materials held by the Bush Presidential Library that pertain to Human
    Rights as a part of foreign developmental aid, particularly for the years post-1989. (99-0461-F)
    (human $right$ NEAR foreign=development=aid)

40. Documents and materials that pertain to Human rights in the bilateral relations between
    the U.S. and China. (99-0461-F)
    (human right=rights) NEAR (china=chinese)

41. Documents and materials that pertain to Human rights as an issue at the Drug Summit
    San Antonio in February 1992. (99-0461-F)
    (human $right$ AND drug summit AND San Antonio AND (February NEAR 92=1992))

42. Documents and materials that pertain to trade issues in the bilateral relations between
    the U.S. and China. (99-0461-F)
    (trade NEAR china)

43. Documents and materials pertaining to Foreign Financial and Military Assistance issues in
    the bilateral relations between the U.S. and Turkey. (99-0461-F)
    (turkey NEAR finance=financial NEAR assist=assistance) OR (turkey NEAR military NEAR
     assist=assistance)

44. Documents and materials pertaining to Arms Exports issues in the bilateral relations
    between the U.S. and Iraq. (99-0461-F)
    (export=exports=arms) NEAR Iraq

45. Copies of all documents that relate to the Central Valley Project Improvement Act of 1992
    (Public Law Number 102-575). This law was introduced in 1991 as H.R. 429. (99-0551-F)
    (Central Valley Project)

46. Records relating to SII (Structural Impediments Initiative) (99-0584-F)
    ([SII]$ Structural Impediments Initiative)

47. Records relating to SUPER 301 / Omnibus Trade and Competitiveness Act of 1988 (99-0584-F)
    (Super 301)
48. Records relating to Japan (Trade and Economic Policy) (99-0584-F)
   ($Japan=Japanese NEAR $trade=$economic)

49. Records relating to Bilaterals (Ministerial Meetings)
    3/2-3/90 Bush - Kaifu Ministerial Meetings (Palm Springs, CA)
    4/4/91 Bush - Kaifu Ministerial Meetings (Newport Beach, CA)
    7/11/91 Bush - Kaifu Ministerial Meeting (Kennebunkport, Maine)
    9/1/89 Bush - Kaifu Ministerial Meetings (Washington)
   (99-0584-F)
   (Kaifu,Bush,$minister,ministerial,$meeting,$discuss,$discussion,(March NEAR 90=1990),(April
   NEAR 91=1991),(July NEAR 91=1991),(September NEAR 89)) AND (Kaifu)

50. Request for records pertaining to Hurricane Andrew (99-0727-F2)
    ($hurricane,$andrew,($hurricane $andrew)*10) AND ($hurricane AND $andrew)

51. Material on "Millie", Mrs. Bush's deceased canine. (200-0590-F)
    (dog,pet,canine,Millie) AND Millie