

**Georgia
Tech**



**Research
Institute**



**The Archival Processing Tool (APT):
User's Guide
Version 2.06**

PERPOS Technical Report ITTL/CSITD 03-2
November, 2002
Revised December 2003

William E. Underwood
Marlit Hayslett-Keck
Sandra Laib

ITTL/CSITD
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 ARL Contract No. DAKF11-97-D-0001, Task Order 64. The findings in this paper should not be construed as an official ARL or NARA position unless so indicated by other authorized documentation.

TABLE OF CONTENTS

INTRODUCTION.....	1
ARCHIVAL WORKFLOW.....	1
EXPLORING A FILE SYSTEM.....	2
INTRODUCTION	2
OPENING A FILE SYSTEM.....	2
VIEW A FILE SYSTEM'S PROPERTIES	4
VIEWING A FILE'S PROPERTIES.....	5
OPENING AND VIEWING A FILE	7
QUICK VIEW PLUS OPTIONS.....	8
<i>Viewing Document Files</i>	8
<i>Viewing Archive Files</i>	8
<i>Viewing Bitmap or Drawing Files</i>	9
<i>Viewing Spreadsheet, Presentation and Database Files</i>	9
<i>Viewing File Types that are Unknown to Quick View Plus</i>	10
<i>Exiting Quick View Plus</i>	12
HOW THE APT IDENTIFIES FILE TYPES.....	12
APT OPTIONS	14
<i>Display Storage Areas</i>	14
<i>Associating a File Type with a Viewer</i>	14
<i>Associate Archive File Type with Extractor</i>	16
<i>Associating a File Type with a Repair Utility</i>	17
<i>Associating a File Type with a Conversion Utility</i>	17
CLOSING A FILE SYSTEM.....	17
CHOOSING A PROCESSING ACTIVITY	17
SAVING A FILE SYSTEM IN A TAR FILE	18
VIEWING THE MANIFEST OF A TAR ARCHIVE	19
HELP	20
COMPARISON OF APT EXPLORE WITH WINDOWS EXPLORER	21
EXITING THE APT.....	21
ACCESSION.....	22
LOADING FILE SYSTEMS ASSOCIATED WITH AN ACCESSION	23
STORING CONTAINERS IN ARCHIVAL STORAGE	26
ASSOCIATE CONTAINERS WITH AN ACCESSION	27
DETECTING COMPUTER VIRUSES AND SECURITY CLASSIFIED DOCUMENTS IN RECORD SERIES	27
SCHEDULING WORK.....	28
STARTING WORK.....	29
FILTERING A FILE SYSTEM	30
INTRODUCTION	30
THE FILTERING MENU	30
FILTER MANUALLY	31
<i>Creating an OS/APP Hash Code Filter</i>	31
<i>Unknown File Types</i>	35
<i>Cancel</i>	35
<i>Transferring the Blocked (OS/App) Files</i>	36
<i>Saving your work</i>	36
<i>Saving a Filter</i>	36
FILTER ON OS/APP HASH CODE.....	36
<i>Loading a Filter</i>	36
<i>View OS/App Hash Code Filter</i>	37
<i>Filtering on OS/App Hash Code</i>	38
<i>Viewing Types of Blocked Files</i>	39
<i>Manually Filtering the Files Passed through the OS-APP Hash Code Filter</i>	40
FILTER ON FILE TYPE	41
<i>Define a File Type Filter</i>	44
<i>Filter on File Type</i>	44
<i>Manually Filtering the Files Passed through the File Type Filter</i>	45

FILTER ON OS/APP HASH CODE AND FILE TYPE	45
<i>Define OS-App Hash Code & File Type Filter</i>	46
<i>Filter on OS-App Hash Code & File Type</i>	46
<i>Manually Filtering the Files Passed through the OS-App Hash Code & File Type Filter</i>	47
EPILOGUE.....	47
PRESERVING DIGITAL FILES	48
EXTRACT ARCHIVED FILES	48
DECRYPT PASSWORD ENCRYPTED FILES	50
REPAIR CORRUPTED FILES	51
CONVERT FILES TO OTHER FORMATS	52
ARRANGING A FILE SYSTEM.....	53
INTRODUCTION	53
THE ARRANGEMENT MENU.....	53
MOVING A FILE INTO A FOLDER.....	54
REORDERING THE FILES IN A DIRECTORY	55
VIEWING THE PROPERTIES OF A FOLDER OR FILE.....	56
RENAMING A DIRECTORY (FOLDER).....	57
CREATING NEW FOLDERS	57
REVIEWING A FILE SYSTEM.....	59
INTRODUCTION	59
THE REVIEW MENU	59
REVIEWING A RECORD	60
<i>Open File</i>	61
<i>Close File</i>	62
<i>Redaction</i>	67
<i>Marking for Transfer</i>	72
VIEWING FILE PROPERTIES	74
SAVING A REVIEWED FILE SYSTEM	75
TRANSFERRING FILES MARKED FOR TRANSFER.....	75
CREATING A REFERENCE COPY.....	77
OPENING A CLOSED FILE.....	78
RE-REDACTING A CLOSED FILE.....	78
EDIT REASONS WITHDRAWN	78
EDIT DOCUMENT TYPE LIST.....	78
DESCRIBING A FILE SYSTEM.....	79
ACCESS ARCHIVAL HOLDINGS.....	84
ACCESSING RECORD SERIES IN ARCHIVAL HOLDINGS	85
COPY RECORD SERIES TO WORK AREA FOR PRESERVATION OR REVIEW	85
MOVE RECORD SERIES FROM WORK AREA BACK TO ARCHIVAL HOLDINGS	85
CREATING REFERENCE COPIES FOR PUBLIC ACCESS.....	85
INSTALLING THE APT AND CONFIGURING RESOURCES	87
WINDOWS 2000	87
MICROSOFT OFFICE	87
ANTIVIRUS SOFTWARE	87
WINZIP AND OTHER ARCHIVERS	87
INSTALLING THE APT	88
STORAGE MANAGEMENT	88
QUICK VIEW PLUS	89
PASSWORD RECOVERY SOFTWARE	89
SOFTWARE TO ERASE TEMPORARY AND DELETED FILES	89
INFORMATIK IMAGE DRIVER	90
FILTERS.....	90
TUTORIAL FILES	90
FILE REPAIR TOOLS	90
LICENSES	90
WRITTEN AS A PART OF THIS GOVERNMENT SPONSORED PROJECT	90
FREWARE	91
SHAREWARE	91
COTS.....	91

OPERATING SYSTEM SOFTWARE.....	91
GLOSSARY	92
APPENDIX A: APT ERROR MESSAGES	94
APPENDIX B: FILE TYPES OF OPERATING SYSTEM & APPLICATION SOFTWARE.....	96
NEEDED MODIFICATIONS TO APT USER MANUAL.....	98

Introduction

The Archival Processing Tool (APT) is a prototype software tool designed to support archivists in processing file systems containing digital records. The APT supports filtering, arrangement, preservation, review, and description of record series. The APT prototype also interfaces to modules for accessioning record series, scheduling work, and reporting work status. The APT is one component of PERPOS, or the Presidential Electronic Records Pilot Operating System. PERPOS is a research initiative sponsored by the National Archives and Records Administration and lead by the Georgia Tech Research Institute. The objective of the PERPOS project is to support archivists in gaining intellectual and physical control over the personal computer records received, created and used during the administration of President George H. W. Bush.

This prototype was developed by analyzing the archival work processes of the Bush Presidential Library and Museum and then designing an object-oriented system to support these work processes. It was further refined through experiments in processing the actual Personal Computer (PC) files from the Bush Presidential Administration. It will be refined based on the archivists' experience in processing the Bush PC records. It is currently being extended to include advanced technologies to support archivists in review of electronic records, extension of directory names to more descriptive folder titles, and summarization of record series.

The APT prototype supports the identification of functional and information requirements for processing legacy personal computer records saved in file systems. It represents initial steps in the process of understanding the requirements for the Electronic Records Archives (ERA) of the future.

We appreciate your participation in this research project and welcome your feedback.

Email: William.underwood@gtri.gatech.edu

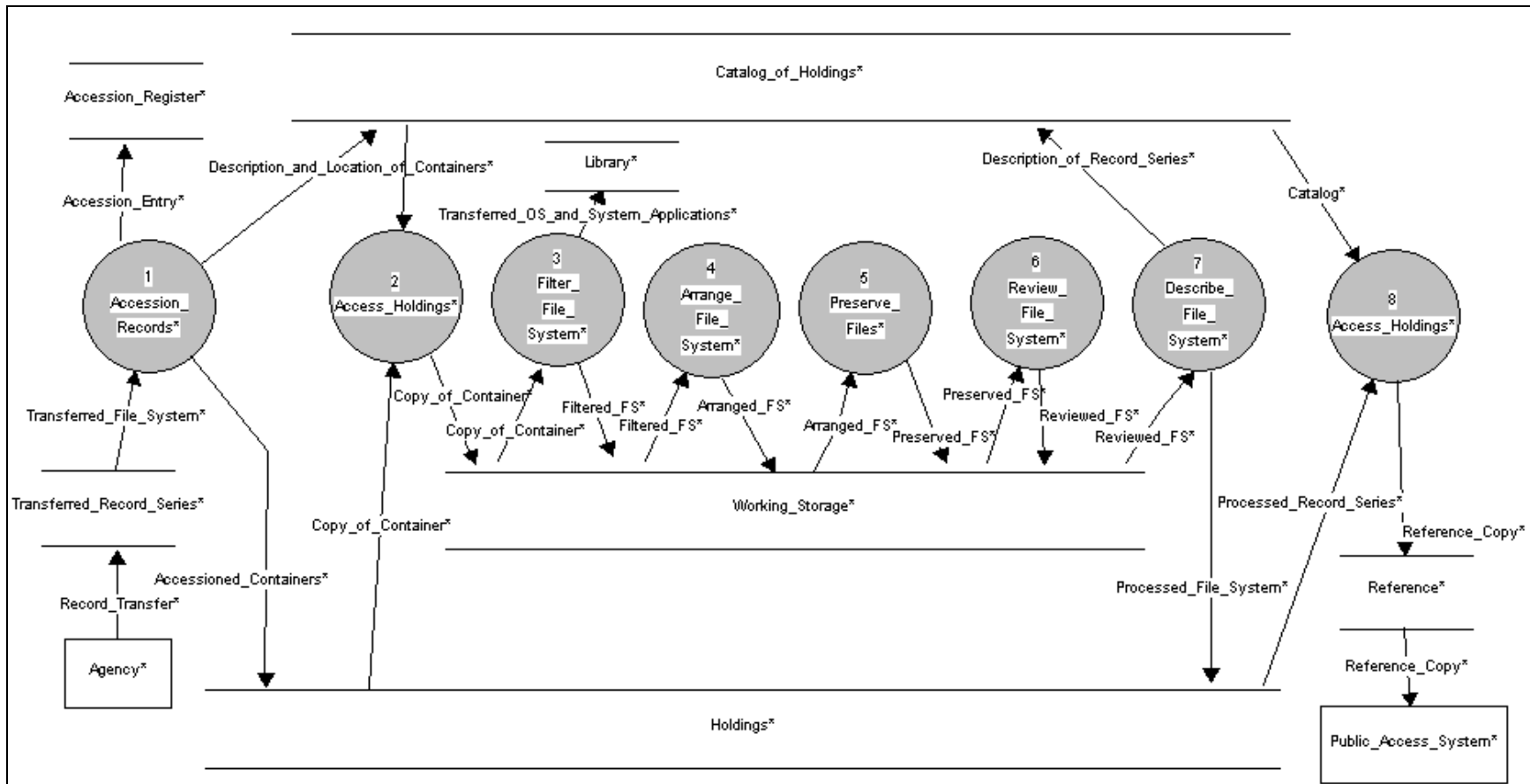
Archival Workflow

Upon receipt of containers for transferred electronic records, a process begins that will result in many of the records being accessible to researchers, but others being temporarily inaccessible to the public due to Freedom of Information Act (FOIA) exemptions or Presidential Record Action (PRA) restrictions.

Archivists at the Presidential Library must accession the record series. Transferred record series are loaded from storage devices, such as floppy and compact disk drives or file transfer areas. The transferred record series (file systems) are loaded into a tool called the Archival Processing Tool and browsing the records. An entry is made in an accession register, and the containers are associated with the accession entry and stored for further processing. This process is illustrated in the following figure. The labeled circles (bubbles) in the figure represent activities in the process. The labeled parallel lines represent kinds of information that are created and used during the process. The labels on directed edges represent the kinds of information that are stored as a result of activities and subsequently used by other activities.

The next activity in processing accessioned electronic record series is for the Supervisory Archivist to schedule archival work. This involves loading the Archival register to determine which archival accessions remain to be processed, loading the current schedule (or Work-in-Process (WIP)), assigning work, saving the schedule, and notifying archivists of assigned work.

Archivists start work, by loading the Work Schedule (WIP) and accession register to identify the accession and containers they are to process. They copy the containers associated with the assigned work into an archivist's subdirectory in a Work Area.



A peculiarity of the Bush PC record series is that they included the entire file system of personal computers, operating system and application files as well as those created. The APT supports filtering files systems by blocking operating system and applications files and passing through user-created files. This activity could be considered a preservation function as it involved separating records and non-records. After each step of work, the archivist saves their work back to the work area, and records the status of their work in the Work-In-Process file.

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 CORR[espondence] folder, 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.

Some PC files may be in obsolete and/or proprietary 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.

PC records must next be reviewed for Freedom of Information Act (FOIA) exemptions on their release to the public. They must also be reviewed for Presidential Record Act (PRA) restrictions on their release.

When archivists have completed the preceding activities, they must describe the record series. Since this involves defining the record creator (organization, office and/or individual records creator, it has traditionally been described as not only description but arrangement of the processed record series. This involves loading the containers containing the record series to view and describe their contents, moving the containers from the work area to archival storage of processed records, and indicating in the work schedule (WIP) that the work is completed. The result is an addition to the catalog of holdings.

Since the master copy that is stored in archival storage may contain records whose access is restricted in whole or in part, it is then necessary to create a reference copy that includes just those records that are open to the public. To do this, an archivist accesses the current holdings and creates a reference copy that is transferred to the Public Access System.

Archivists will also need to access the holdings of processed record series to review those closed files or originals of redacted documents when access restrictions have expired. They will also need to access those holdings for preservation actions such as converting to new file formats when current file formats become obsolete.

The Windows 2000 operating system and software applications such as the APT, Microsoft office, McAfee Virus Scanner, and MKS Tools for checking for security markings are stored on logical hard disk C. The Accession Register, Catalog of holding and PC Files Test Suite are also stored on this disk.

Working Storage which includes the Work Schedule and status and Archivists work areas are stored on logical hard disk D. Archival Holdings, that is accessioned record series and processed record series are stored on logical hard drive E.

The remainder of this manual describes four tools: the Accession Register, Work Schedule and Status, Archival Processing Tool (APT) and Archival Description Tool. Most of the archival processing functions have been integrated into the Archival Processing Tool.



Exploring a File System

Introduction

When working with the Archival Processing Tool, there are a few operations that you will perform regardless of whether you are filtering, arranging, reviewing or describing file systems. Here are the operations you will perform frequently.

- You will **open a file system** to perform every APT activity. File systems are directories or TAR files containing files you will process.
- You will also **view and individual file's properties**. File properties give you information about the date the file was created; and the type of file it is (document, spreadsheet, email, etc). While you may not examine every record's properties, it will be important to know how to do so.
- You will frequently **open and view files**.
- You will **close a file system** in order to open another file system.
- You will choose an **APT activity**. These activities are Explore, Filtering, Arrangement, and Review.
- You will save the results of your work by **saving a file system** as a JAR file.

Opening a File System

One selects the file system to be processed by opening it. There are two ways to open a file system. When the APT software is started, the first dialogue box prompts you to begin work by opening a file system. You can open a folder, a tar file, or a device, e.g., 3 1/2 floppy, or Compact Disk, containing a file system.

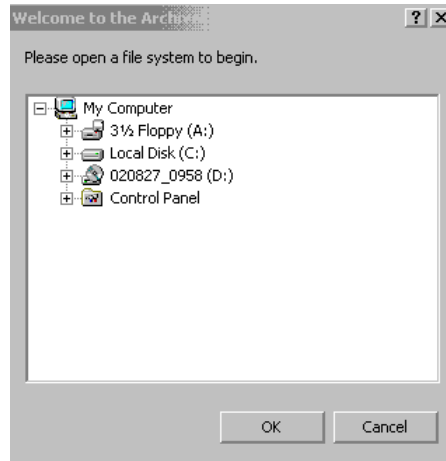
To open a file system:

Step 1: If you wish to work with the contents of an entire drive, go to step 2. If you wish to work with a file system in a folder or a tar file, use the left mouse button to click the plus signs to open folders until you reach the folder or tar file.

Step 2: To select the file system, click it once. It will be highlighted.

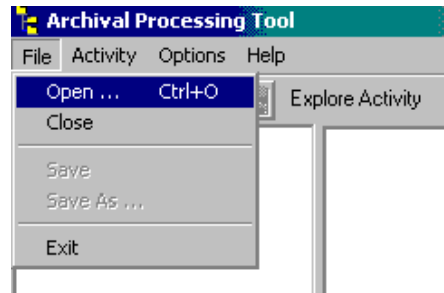
Step 3: Click 'OK'. The file system will be loaded and you can begin the archival processing activity.

Note: Clicking 'Cancel' terminates the selection and opening of a file system. Nothing will be opened.

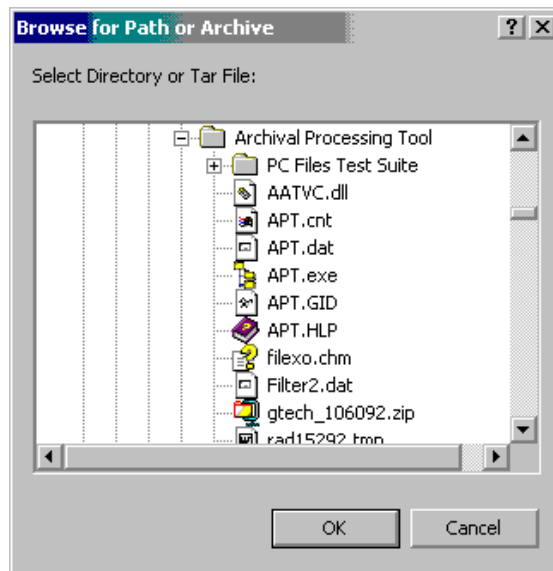


OR

You can open a file system from the File entry on the menu bar. From the menu bar, select File>Open.

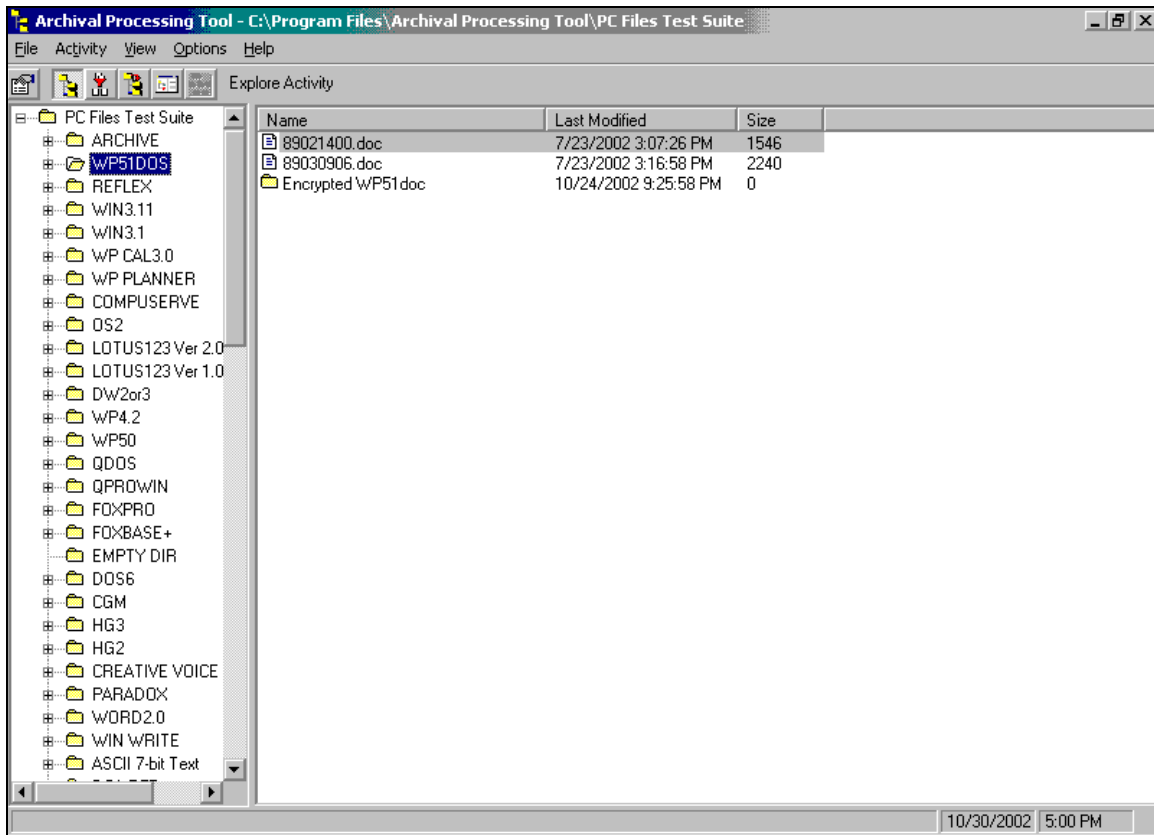


The following dialog box appears.



Follow the instructions for opening a file system given above.

In the following case, a file system containing the PC Files Test Suite that is used to test the accuracy of the file type identification function has been opened.



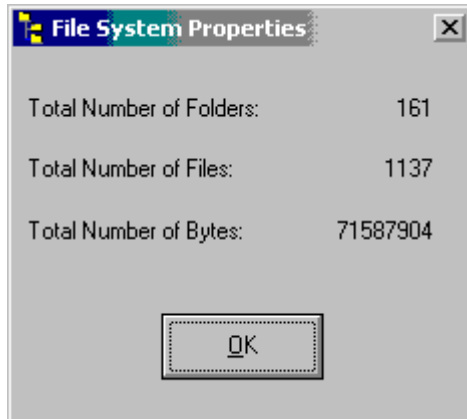
If you highlight a folder title the names of files and folders contained in that folder will be shown in the right windowpane. The date last modified and length of the files and folders will also be shown.

View a File System's Properties

When accessioning a file system, this feature can be used to determine the size of the file system.

Step 1: Select the View Pull Down Menu

Step 2: Select System Properties.

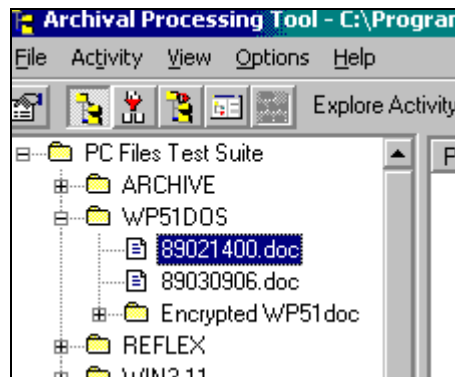


You will see the number of bytes, the number of files, and the number of folders in the loaded file system.

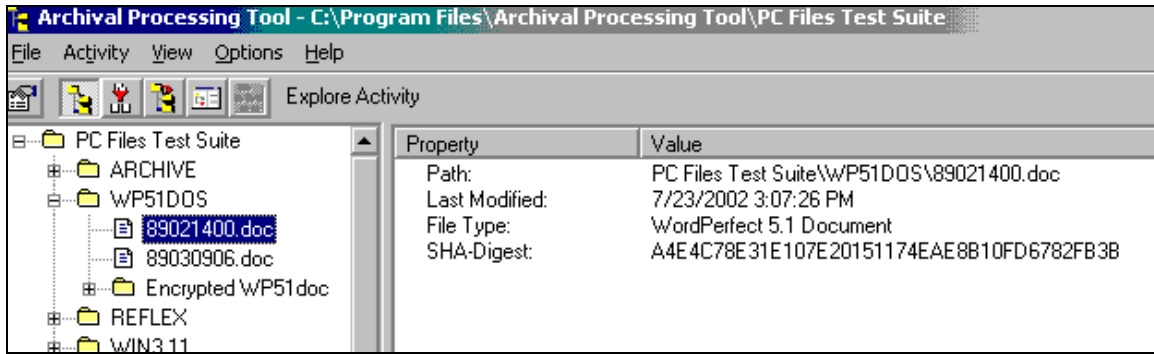
Viewing a File's Properties

A file's properties include its name, its path in a file system, its length, the date it was last modified, its file type, and its sha-1 message digest. To view a file's properties:

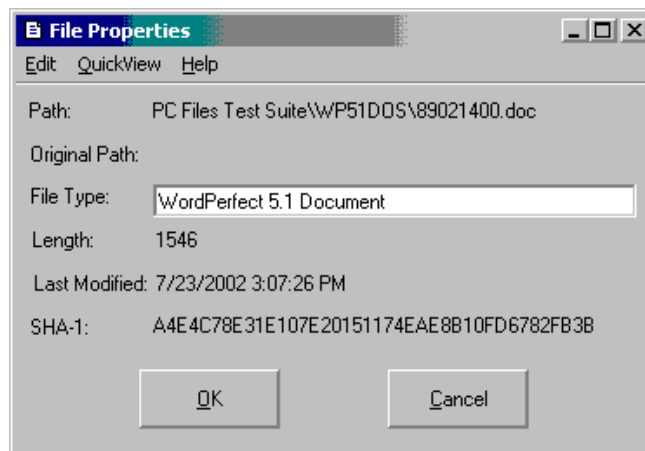
Step 1: Using your left mouse button, click the plus signs to open folders until you reach the file you wish to open.



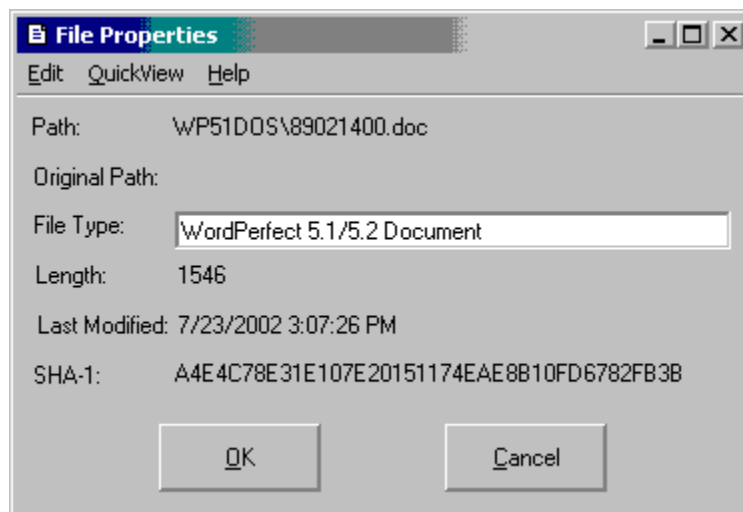
Step 2: Highlight the file name and click it once. In the right plane, you will notice that the file's properties are listed. You can confirm the file's path, the date it was last modified, the file type and the SHA-Digest, or its Secure Hash Algorithm. This SHA is a unique identifier calculated to the file by the APT.



A second way to view a file's properties is highlight the file's name, and click the right mouse button. A menu will appear. Select the properties option and a File Property dialog box similar to the following will appear.



A third way to view a file's properties is to select the properties option from the View pull-down menu on the menu bar.



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.

Opening and Viewing a File

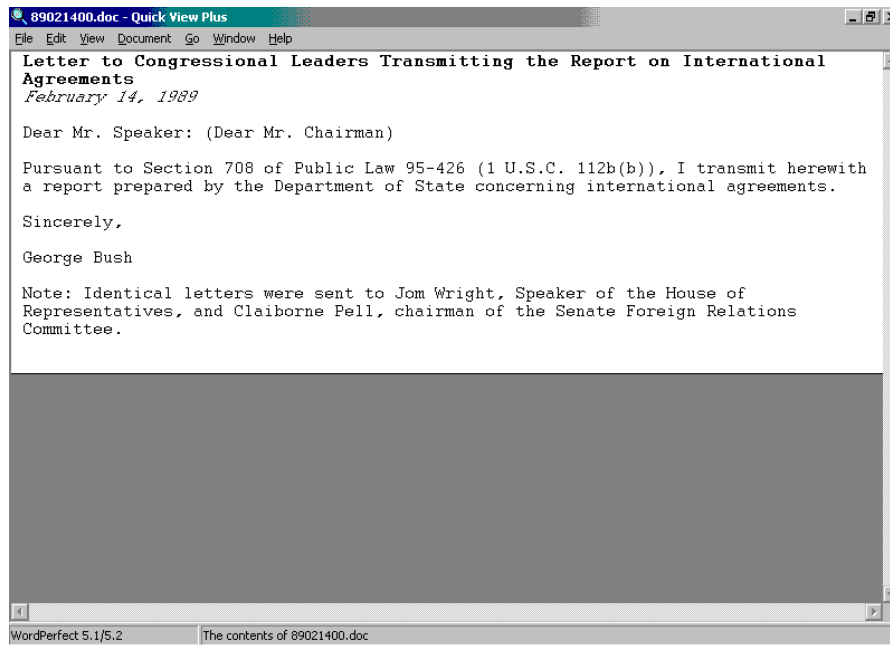
Due to the rapid life cycle of computer technology, hardware and software used to create documents and databases eventually become obsolete. The challenge this poses to archivists is how to view these legacy files without access to the software or hardware on which they were created. The APT uses Quick 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, so additional viewers are required.

The Quick View Plus User's Guide discusses the features and options of Quick View Plus. 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 pull-down menu on the menu bar. The file will be displayed with the Quick View Plus file viewer.

Quick View Plus Options

The Quick View Plus Options should be as follows:

In Quick View Plus select View-Options-General

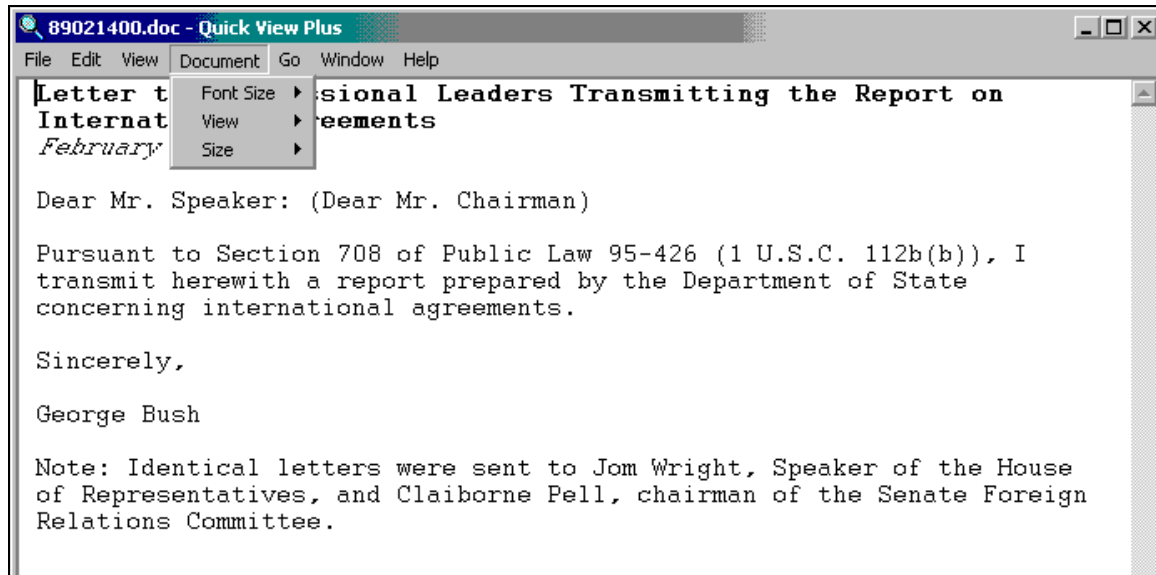
1. *Display Quick View Plus logo on startup* should not be checked
2. *Auto* Should not be checked
3. *New Views have a File Navigation Pane* should not be checked.
4. *New Views have a status bar* should be checked.

In Quick View Plus 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.

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.

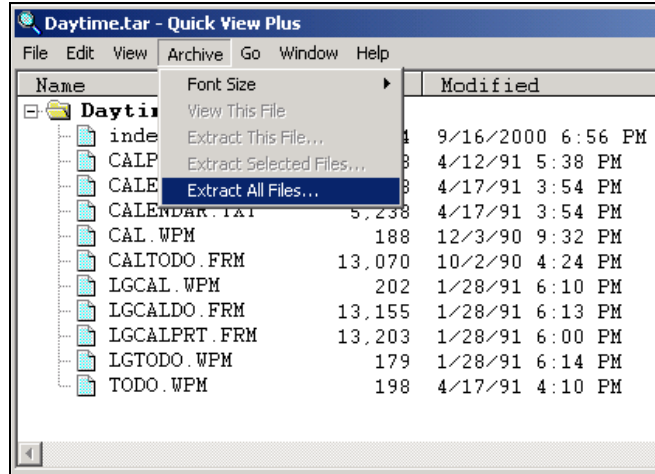


The page size options are Full Size, Fit to Window, and Fit to Window Width. The default is Full Size. Either Full Size or Fit to Window Width is best for reading the text and for creating a document image when redaction is needed.

Viewing Archive Files

Quick View Plus can display the file structure, display the files and extract the files of a number of six types of archive files, namely:

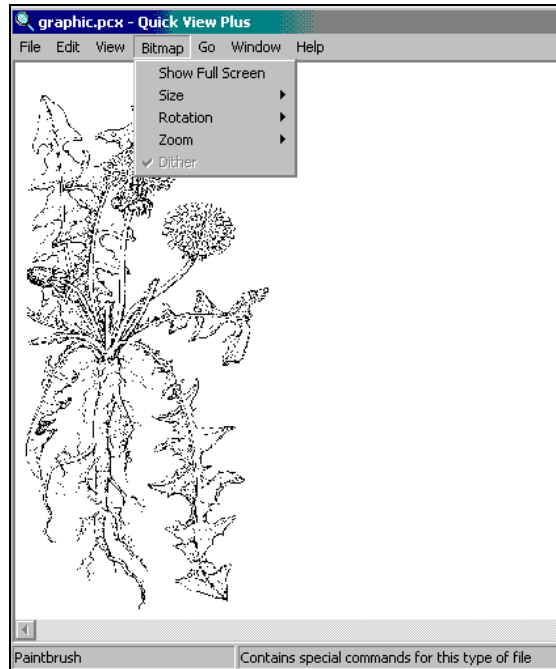
GZIP
 LZH Compress
 Microsoft Binder..... LZA Self Extracting Compress
 Unix TAR.....Zip



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 pull-down menu.

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.



Viewing Spreadsheet, Presentation and Database Files

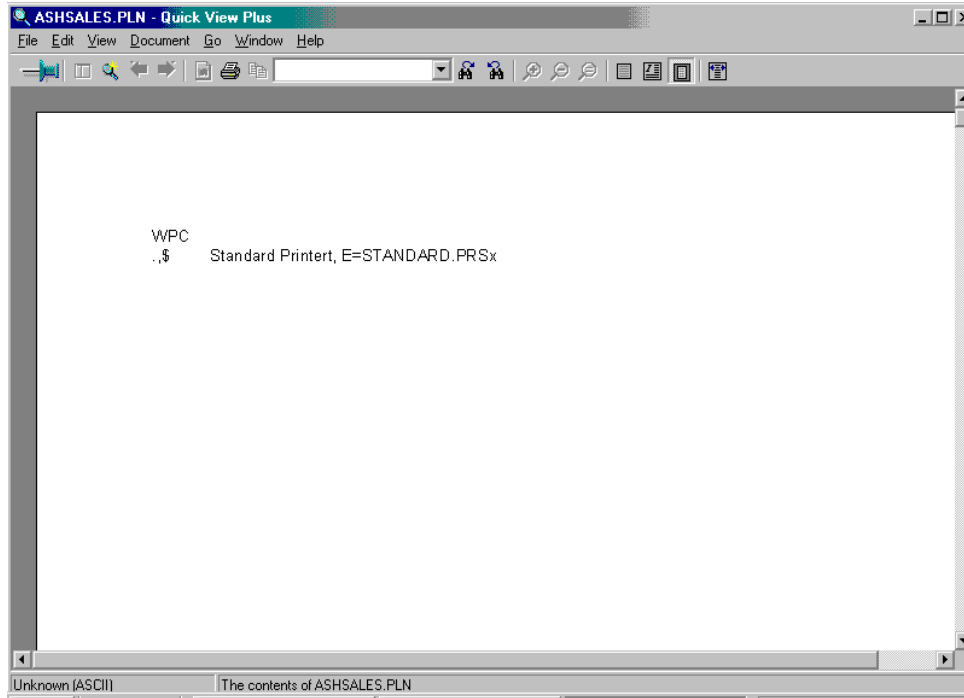
After viewing a spreadsheet (WKS, WB1, WQ1, XLS), presentation () or database files (), you can enlarge or reduce the on-screen font size, show or hide the gridlines, and display any sheet in a multiple-sheet spreadsheet. The figure below shows Quick View Plus's view of a Dbase IV or V database (DBF).

	NAME	READONLY	CKVAL	DATA
1	PREF2.0	COLORSET	COMPQ_COLO	FALSE 57720 (Memo File)
2	PREF2.0	COLORSET	COMPQ_MONO	FALSE 32554 (Memo File)
3	PREF2.0	COLORSET	TOSH_3200	FALSE 20881 (Memo File) 0
4	PREF2.0	COLORSET	COLOR	FALSE 24241 (Memo File) 0
5	PREF2.0	COLORSET	MONOCHROME	FALSE 53450 (Memo File)
6	PREF2.0	COLORSET	T5200_COLO	FALSE 277 (Memo File)
7	PREF2.0	COLORSET	T5200_MONO	FALSE 1967 (Memo File) 1
8	PREF2.0	COLORSET	DEFAULT	FALSE 24241 (Memo File) 0
9	PREF2.0	WINDCMD		FALSE 14165 (Memo File) 1
10	PREF2.0	FILERPOS		FALSE 3494 (Memo File) 0
11	PREF2.0	HELPPPOS		FALSE 60577 (Memo File) 0
12	PREF2.0	VIEWPOS		FALSE 23511 (Memo File) 0
13	PREF2.0	WINDMODIFY	PRG	FALSE 21599 (Memo File) 0
14	PREF2.0	WINDMODIFY	FINDENDS.PRG	FALSE 8720 (Memo File) 0
15	PREF2.0	WINDMODIFY	ACROWORD.PRG	FALSE 16935 (Memo File) 0
16	PREF2.0	WINDMODIFY	WORDACRO.PRG	FALSE 20922 (Memo File) 0
17	PREF2.0	WINDMODIFY	CONSWORD.PRG	FALSE 61305 (Memo File) 0
18	PREF2.0	WINDMODIFY	FINDACRO.PRG	FALSE 37096 (Memo File) 0
19	PREF2.0	WINDMODIFY	WORDHYP.PRG	FALSE 5572 (Memo File) 0
20	PREF2.0	WINDMODIFY	HYPWORD.PRG	FALSE 437 (Memo File) 0
21	PREF2.0	WINDBROW	FILEPLAN	FALSE 14553 (Memo File) 0
22	PREF2.0	WINDMODIFY	WRTFLPLN.PRG	FALSE 62670 (Memo File) 0
23	PREF2.0	WINDBROW	KEYWORD	FALSE 40058 (Memo File) 0
24	PREF2.0	WINDMODIFY	TSTPRCT.PRG	FALSE 4553 (Memo File) 0
25	PREF2.0	WINDMODIFY	DEWEY.PRG	FALSE 19852 (Memo File) 0
26	PREF2.0	WINDMODIFY	BCKTNT.PRG	FALSE 32914 (Memo File) 0
27	PREF2.0	WINDMODIFY	DICTION.PRG	FALSE 29930 (Memo File) 0
28	PREF2.0	WINDMODIFY	FINDTYPE.PRG	FALSE 52053 (Memo File) 0

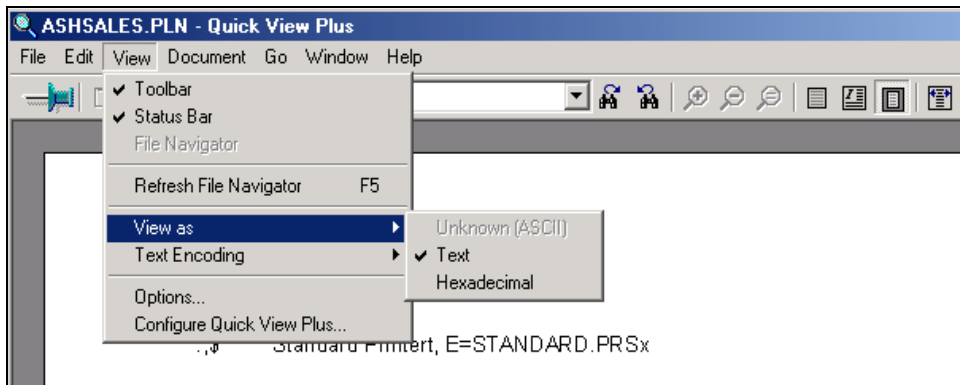
Note that Quick View Plus does not display the values of memo fields that are associated stored in an associated memo file (FOXUSER.FPT). Consequently, the APT uses a different viewer for database files that have associated memo files that also displays the values of memo fields.

Viewing File Types that are Unknown to Quick View Plus

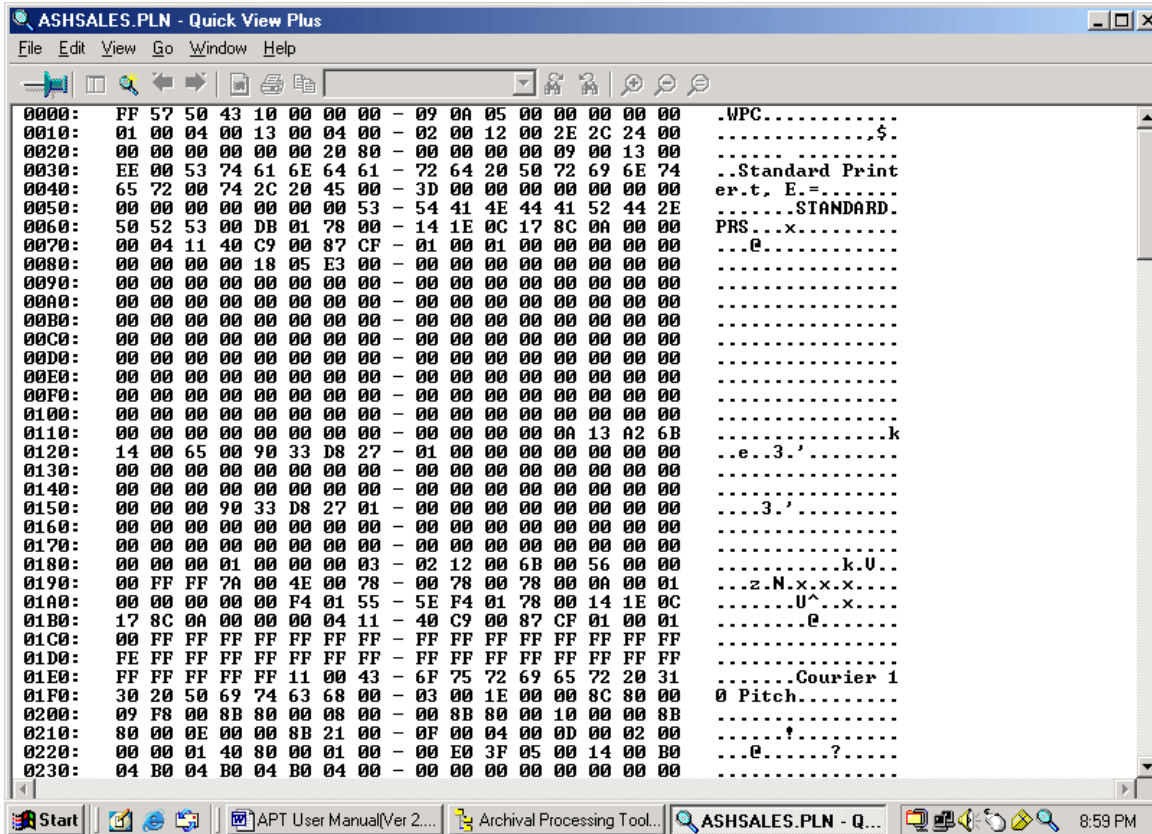
While Quick View Plus can view more than 225 file types there are many that it cannot identify and for which it does not have an appropriate viewer. Such file types are said to be *Unknown* file types. Assuming that Quick View Plus has been configured to view Unknown file types as Text, the file ASHLAND.PLN from the PC Files Test Suite would be displayed as follows.



It is displayed in this form because Quick View Plus can be configured to view file types that are Unknown as Text. This view of the file is not very informative. However, 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.

The point is that 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.

Exiting Quick View Plus

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

- Selecting "Exit Quick View Plus" from the File pull-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 pull-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.

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 document format, but 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 identifier first checks to see if the file is empty, or if the file property indicates that 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:	C8 00 F3 64 00 00 37 01	- 1E 00 00 00 01 00 AB 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	..*...1...A..	U.
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	?...?.5.?.W.?	v.
060:	37 01 44 04 37 01 EA 05	- 37 01 7A 06 37 01 7D 08	?..D.?.?.?.z.?	}.
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
0A0:	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 texts 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."

APT Options

Under the APT Options pull-down menu there are the following options.

Display Storage Areas

This option allows the archivist to see which storage areas are associated with container types or databases. This association is determined at installation time and can only be changed by the system administrator.

Metadata and Record Series Storage	Storage Device
Containers Transferred to Presidential Library	A:; D:
Accessioned Containers	C:\Accessions
Working Storage	C:\Work Area
Containers of Processed Record Series	C:\
Containers of Files Transferred out of File Systems	C:\Transfers
Reference Copies	C:\Reference
Accession Register	C:\Metadata
Archival Catalog	C:\Metadata
Work in Process (Work Schedule)	C:\Metadata

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 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.

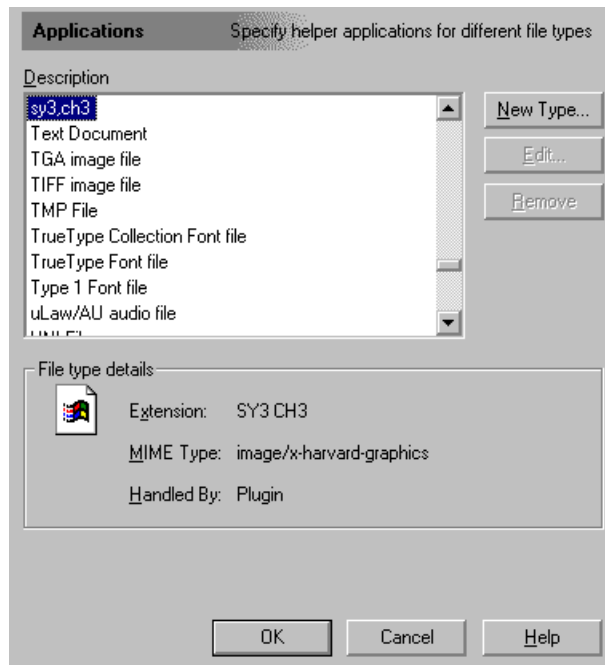
File viewers such as Quick View Plus are needed to view word processing documents, pictures, spreadsheets, and databases. Media players are needed to listen to sounds or play video. Archive file extractors such as WinZip are needed to extract files from archive files.

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 Assoc. File Type with Viewer option under the Options pull-down menu.



To add a new file type and associate it with a helper application:

Step 1: Click the New Type button to display an empty dialog

Step 2: Type the name of the file type in the File Type name field.

Step 3: Type the common file extensions used by the File Extension field.

Step 4: Type the MIME type in the MIME Type field.

Step 5: Type (or use the Browse Button to select) the location of the application capable of interpreting files of this type in the Handled By field.

Step 6: Click OK to store the new information.

To edit an existing association of File Type and 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 file extensions, MIME Type and helper application for the selected item.

Step 3: Edit the MIME Type and/or File extensions.

Step 4: Select an action in the Handled By area of the dialog box. You can designate:

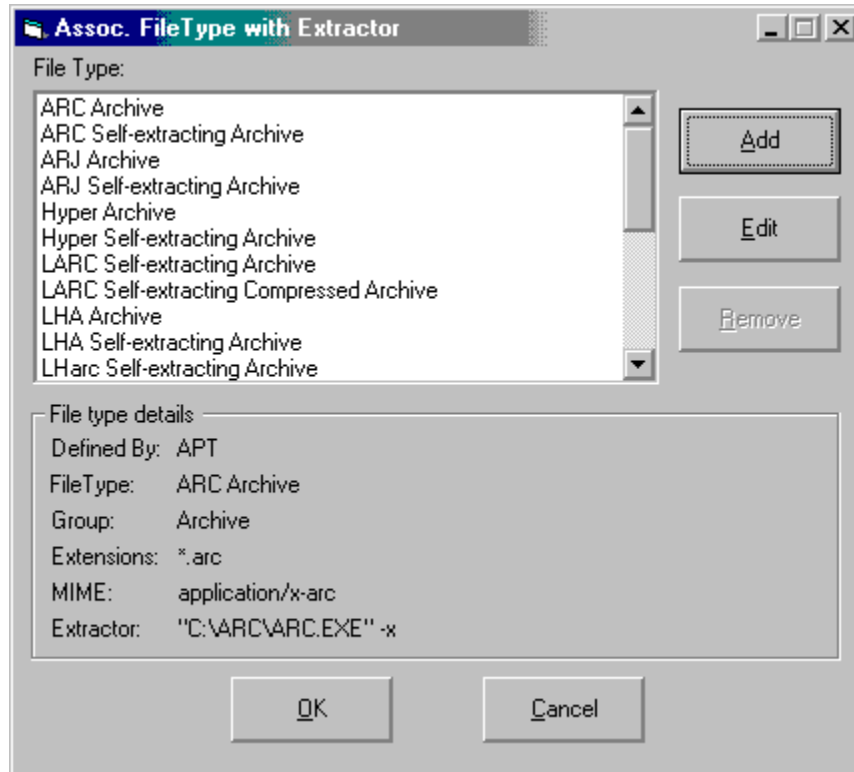
- Quick View Plus
- Application: Type (or use the Browse Button to select) the location of the application capable of interpreting files of this type in the Handled By field.

Step 5: Click OK to store the new information.

Note: that you cannot change the File Type Name, because this is the name associated with a file by the File Type Identifier.

Associate Archive File Type with Extractor

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 pull-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 archives.

The associations are all predefined, but in case the extractor is not available say due to becoming obsolete, one can edit the path to the extractor.

Associating a File Type with a Repair Utility

[To be implemented]

Associating a File Type with a Conversion Utility

[To be implemented]

Closing a File System

Closing a file system is very simple. To close a file system:

Step 1: From the main pull-down menu, select File>Close. The file system you had open will close.

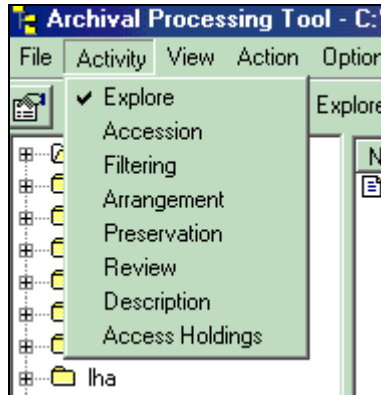
Step 2: Exit the APT or open a different file system.

Choosing a Processing Activity

There are two ways to choose a processing activity.

Using the pull-down menu:

Step 1: From the main pull-down menu, select Activity. The eight activities of the APT appear on the menu.



Step 2: Select the activity you want to perform. Notice the dialogue box and menus change slightly according to which activity you choose.

OR

Step 1: With one mouse click, select the activity you want to perform from the Activity Toolbar. Below the main pull-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 pull-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 pull-down menu. The challenge is learning which button corresponds with which activity.

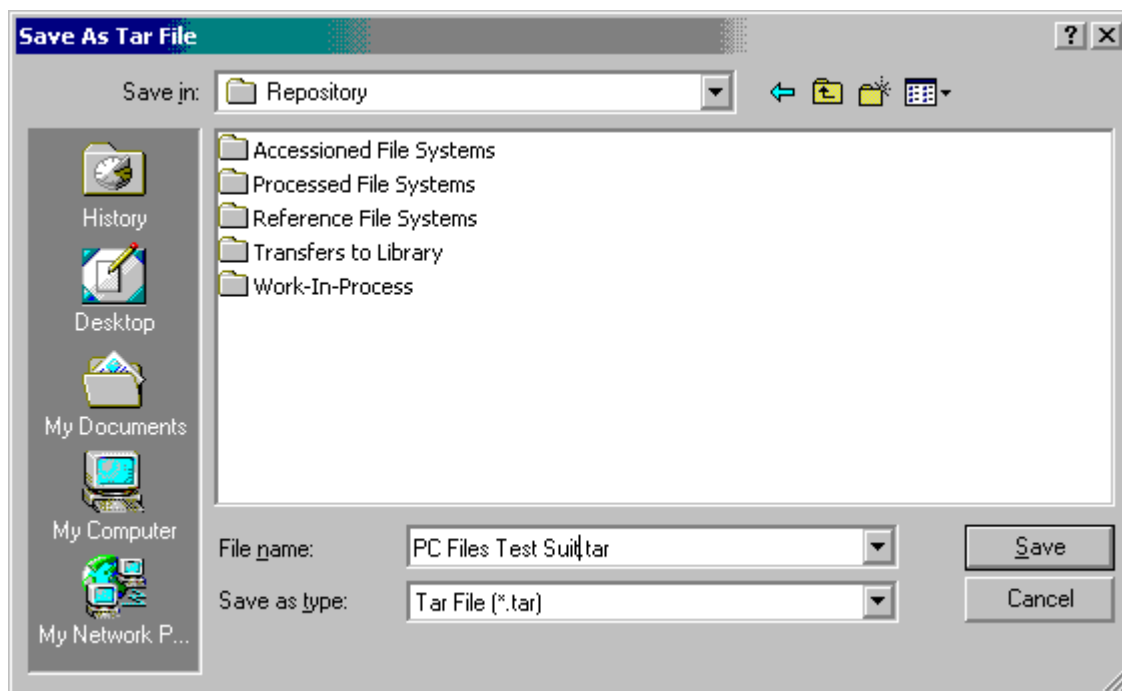
From left to right, the activities are: Accession, Explore, Filtering, Arrangement, Review and Description. You can always confirm the activity as it is written in text to the right of the toolbar.



Saving a File System in a TAR File

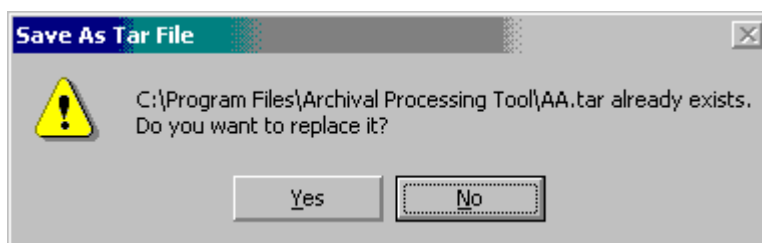
After filtering, arranging or describing a file system, you will save the results of your work.

Step 1: From the main menu of the filtering activity window, select **File>Save As**. You will see a directory named Repository. Open it.



Step 2: In the “File name” field, enter the name you would like to call the file system. Click ‘save’. You will use this file system 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 storing to the same filename that you loaded, the following message will be displayed.



Do not save a file system that has been filtered with the same name as the file system that was loaded. You may save file systems that are being processed with the same filename, but the previous version will be replaced.

All file systems are saved in a single file of file type TAR Archive. In addition, a META-INF directory is added to the file system that contains a file named manifest.mf. (See Viewing the Manifest of a File System

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 also 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.

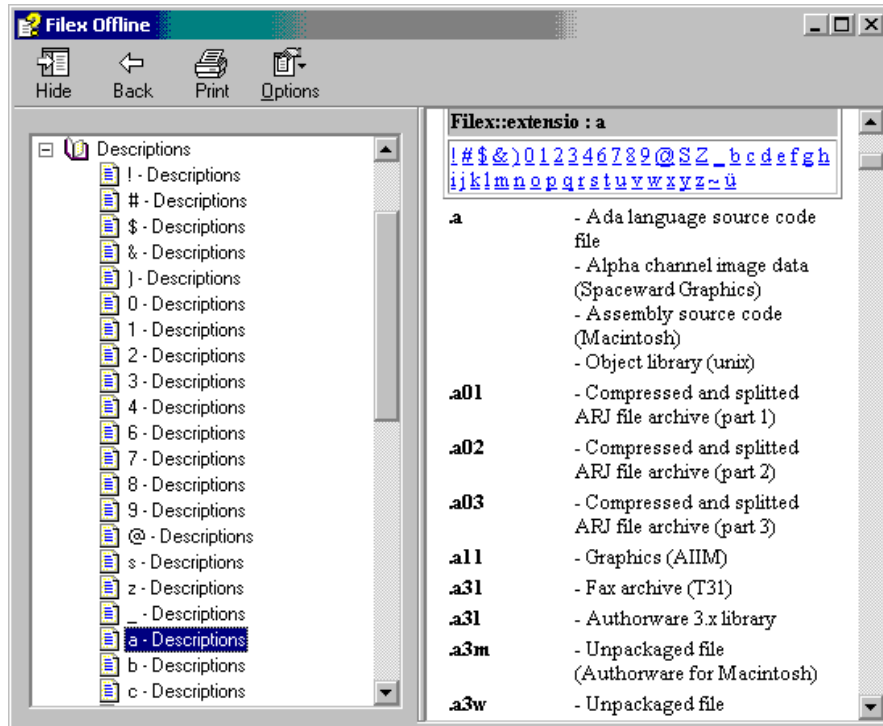
If the file system has been rearranged, there may be an attribute Original Path. If the file system has been reviewed and saved, each file that has been reviewed will have additional attributes: Access (Open, Closed, Redacted), and if Closed, Reason for Withdrawal, Date of Withdrawal, Subject, Author,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

```
Manifest-Version: 1.0
Name: PC Files Test Suite\
FileType: Directory
Name: PC Files Test Suite\ARCHIVE\
FileType: Directory
Name: PC Files Test Suite\ARCHIVE\LZH Compressed\
FileType: Directory
Name: PC Files Test Suite\ARCHIVE\LZH Compressed\wsd2snd.lzh
SHA-Digest: BA8AD40483EFC72187BE7B89789199165E10B229
FileType: Unknown FileType
Name: PC Files Test Suite\ARCHIVE\LZH Compressed\icspred1.lzh
SHA-Digest: 75CB3FA60386A7B90157C70DE074CA751981F989
FileType: Unknown FileType
Name: PC Files Test Suite\ARCHIVE\LZH Compressed\rpmv12.lzh
SHA-Digest: AA913DD8D78261E7160B1C7FDFCFCEDBFB3034D7
FileType: Unknown FileType
Name: PC Files Test Suite\ARCHIVE\SELF-EXTRACTING\
FileType: Directory
Name: PC Files Test Suite\ARCHIVE\SELF-EXTRACTING\ARJ SFX
Executable\
FileType: Directory
Name: PC Files Test Suite\ARCHIVE\SELF-EXTRACTING\ARJ SFX
Executable\arj241a.exe
SHA-Digest: 53BEF16AB9D11F491E33DBF531A412CAC4EF3403
FileType: ARJ 1.0 Self-extracting Archive
Name: PC Files Test Suite\ARCHIVE\SELF-EXTRACTING\ARJ SFX
Executable\arj281.exe
```

Help

The Help pull-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 extension is found.



Comparison of APT Explore with Windows Explorer

Some readers may wonder why a tool was created for exploring a file system when one could just use Windows Explorer, a collection of file viewers such as Quick View Plus and a file archiver such as WinZip to do the same thing. Indeed, when we first investigated archival processing of file systems from Legacy DOS and Windows 3.x operating systems, we used Windows Explorer and Quick View Plus, and tar/untar. Windows Explorer supports opening a file system, viewing a file's properties, associating file extensions with file types and viewers, and closing a file system. These are functions that are needed for managing any file system.

However, what Windows Explorer does not support well is identifying legacy file types, automatically separating user-created files from operating system and application file types, making a record of rearrangements, review and records of review, and archival description. Furthermore, Windows Explorer is too powerful and flexible. It has many functions that are not needed in archival processing, and that can place files that should not be modified at risk of being modified.

Since Microsoft Visual C++ and Visual Basic have classes that can be used to create the needed Windows Explorer functionality, they were used to rapidly create the Explore activity of the APT allowing us to focus on the other requisite functionality.

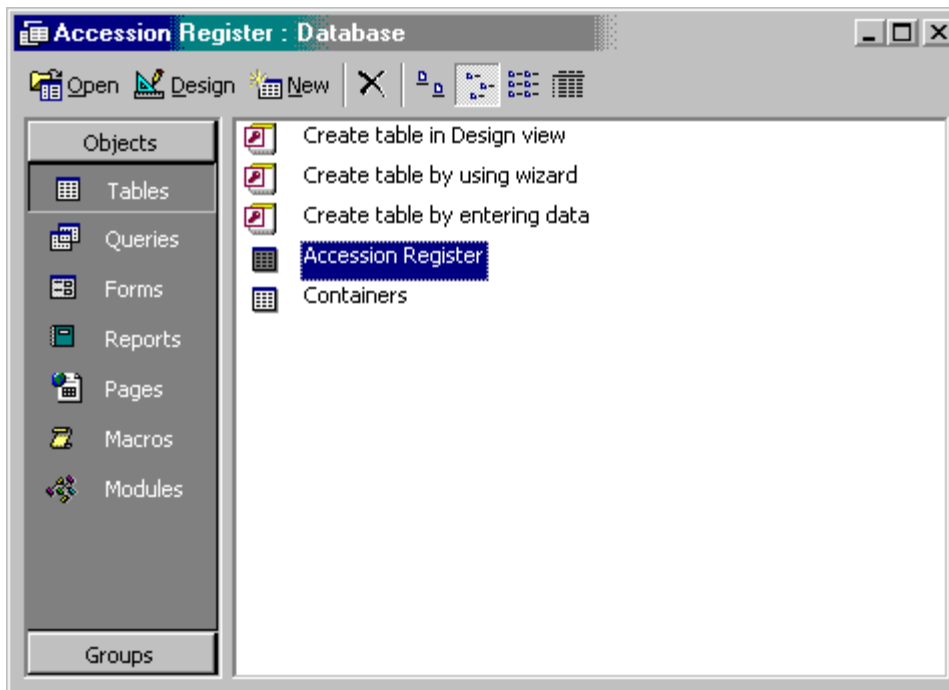
Exiting the APT

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.



Accession

Two tools are used in the process of accessioning electronic record series, the Accession Tool previously developed by the Bush Presidential Library and the Archival Processing Tool. From the desktop select the Accession Tool. A screen similar to the following will be displayed. The Accession Register contains databases and forms to use during the accession process. The same tool is used for accessioning electronic records as is used for accessioning electronic records.



One should first select Forms Objects and then Accession Register. The following form (shown in two parts) will be displayed. This is a form developed by the Bush Presidential Library with minor modifications for accessioning electronic record series.

BUSH PRESIDENTIAL LIBRARY ACCESSION REGISTER			
Accession Number:	2003.0001	Accretion Number:	
Date of Receipt:	5/6/1996	Date Logged In:	
		Logged By:	
Donor Information			
Donor:	A		
Affiliation (If Applicable):			
Street:			
City:		State/Country:	
		Zip:	
Telephone Number:		Donor Representative:	
Accession Information			
Identification:	0		
Brief Description:	0		
Inclusive Dates:	1988-1989	Bulk Dates:	
		Approximate Volume:	1

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 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 these may be determined later to be Personal Records that will be offered back to their creator, or may be donated to the Library by the person who created them.

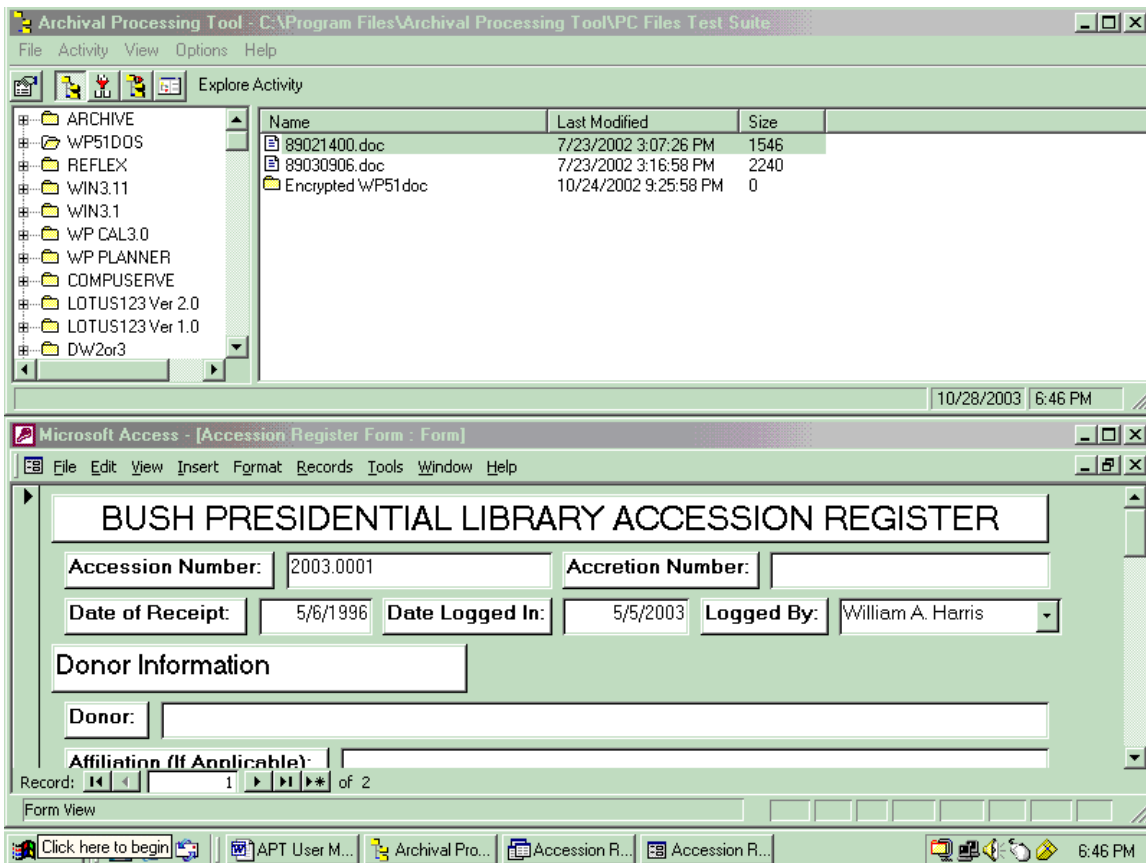
The next section, *Accession Information*, has as its first field, Identification. This is a tentative title of the record series being accessioned. The *Brief Description* is a description of the records series, for example, a description of the responsibilities or activities of the Office or Person who created, received and used the records and the document types.

The contents of these two fields is determined from the Records Transmittal Form, other documents associated with the transfer, or by using the APT to browse and view the loaded records. In the case of the Bush PC Files, there is an inventory of 500+ file systems. A Bush Presidential Library archivist reviewed the inventory of the files that have been transferred and identified the offices associated with the individually who purportedly used the computer from which the file systems were obtained. The accessioning archivist must determine whether they want to accession all of the file systems as a single accession, to accession each file system individually, or to accession the file systems by office. Suppose that the first accession is for the Office of the Vice President.

Loading File Systems Associated with an Accession

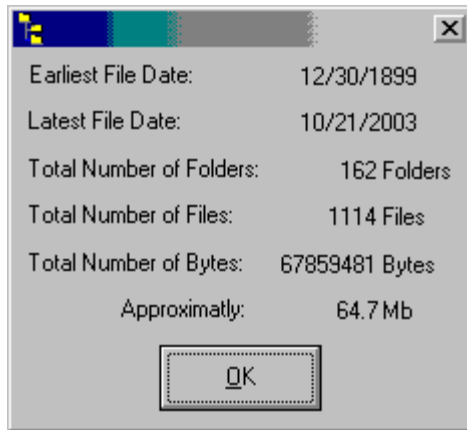
The archivist should now select the Archival Processing Tool icon from the desktop, and select *Accession* from the pull-down menu. Then select *Open* from the File pull-down menu. Currently, the APT assumes that the transferred record series will be loaded from CD-ROM or Floppy Disk. Referring to the Inventory of Transferred Files, find the CD-ROM containing the first of the file systems purportedly created by a person in the Office of the Vice President. When that file system is selected from the CD-ROM, its file system namespace will be displayed in the left window of the APT.

The Archival Register, the Inventory of Transferred File Systems, 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. For instance, the Figure below shows the APT and Accession Register Windows Tiled Horizontally.



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.

To determine the *Inclusive Dates* of files in the file system and the *Approx. Volume*, of electronic records in the accessioned file system, select the System Properties option from the View pull-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). 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 this file is 67859481, 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

So in this example, what would be entered is 64.7 MB, since

67859481 bytes \approx 66,269 KB \approx 64.7 MB

The following figure shows the bottom half of the Accession Register Form. 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 Office of Management and Budget 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.

Restrictions:
 PRA FOIA Deed of Gift Deposit Agreement Other

Associated Accessions:

Date of Deed of Gift/Donor: **Date of Deed of Gift/Archivist:**

Date of Deposit Agreement: **Date Returned:**

Transfer Information

Audiovisual **Audiovisual Department Accession Number:**

Museum **Museum Accession Number:**

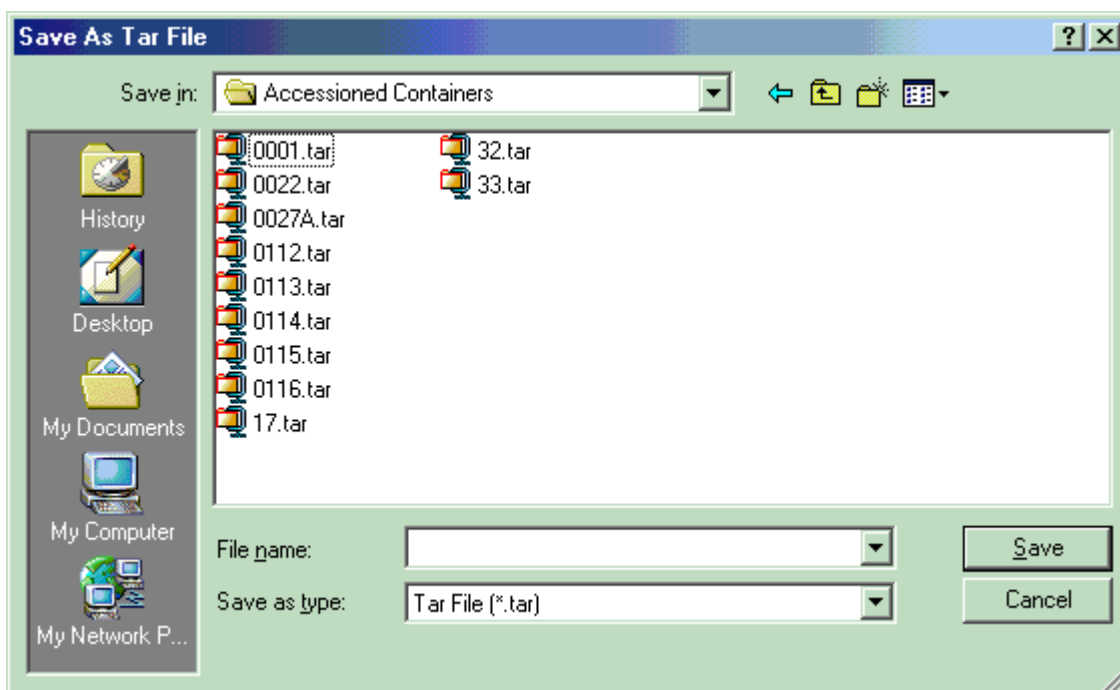
Additional Transfer Information:

Additional Information

Notes:

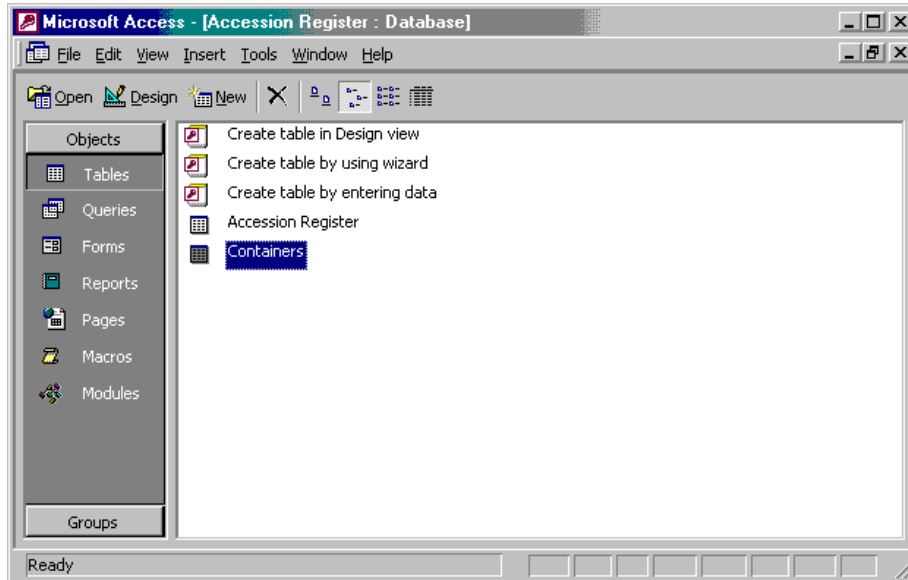
Storing Containers in Archival Storage

When all relevant information has been entered in the Accession Register entry, the file system should be stored. From the Files pull-down menu, select *Store*. You will be prompted for an accession number and container file name. The accession number will be stored in the manifest of the container file. Do not replace a previous accessioned container with the current container.



Associate Containers with an Accession

The next step is to associate containers with an accession. At the bottom of the Desktop, you will see a button labeled Accession Register: Database. Point at that button with the mouse and left-click the mouse button. Select the Tables object, and the following window will be displayed.



Select the Containers Data Table and a window similar to that in the following figure will be displayed. Enter the accession number and the names of the containers associated with the accession. By container is meant a package or archive containing a file system, e.g., a tar file. Enter the volume of the file system in Bytes (KB, MB GB) and/or Files. When all of the containers for a particular accession have been stored, add the volumes and enter the total in the Approx. Volume filed on the Accession Register Entry.

An 'A' will automatically be entered in the Status field indicating that the container has been *Accessioned*. When a container associated with an accession is being processed, an 'I' will be added, e.g., AI, indicating that the container is *In Process*. When archival processing is completed, that is, the files in the container have been arranged, preserved, reviewed and described, a 'C' will be added, i.e., AIC, indicating that processing is complete.

The screenshot shows the Microsoft Access application window titled "Microsoft Access - [Containers : Table]". The menu bar includes File, Edit, View, Insert, Format, Records, Tools, Window, and Help. The toolbar contains various icons for data manipulation. The main workspace displays a data table with the following columns: Accession Number, Accretion Number, Container File Name, Volume-Byte, Volume-Files, and Status. The table contains five rows of data, with the last row marked with an asterisk (*).

	Accession Number	Accretion Number	Container File Name	Volume-Byte	Volume-Files	Status
	2003.0001		0112.tar	64.7 MB	1114 Files	A
	2003.0001		0113.tar	12 MB	401 Files	A
	2003.0001		0022.tar	35 MB	820 Files	A
	2003.0001		0027A.tar	22.5 MB	656 Files	A
*						A

Detecting Computer Viruses and Security Classified Documents in Record Series

The APT prototype was developed to support the processing of sensitive, but unclassified records. Nevertheless, the loaded records must be checked for security markings to test the assumption that the records are indeed unclassified. Failing to do so can result in a security violation requiring isolation of the media on which the records are stored including record series that do not include records with security markings.

Records from PC file systems, E-mail systems and Records Management Applications should have been checked for computer viruses. However, electronic records accessioned must be again checked for viruses to test the assumption 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 at a later date when files are distributed to researchers.

Consequently, any record series read from floppy or CD-ROM is automatically checked for security markings and computer viruses. Furthermore, any record series 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.

This capability is demonstrated by

1. Insert the "PERPOS Accession Demo" Disk (or CD-ROM) in the Floppy Drive (or Compact Disk Drive).
2. Select Accession from the Activity pull-down menu.
3. Select Load from the File pull-down menu.
4. Then select folder "Virus Detection Demo"

To demonstrate the capability to detect security markings in files to be accessioned,

1. Insert the "PERPOS Accession Demo" Disk (or CD-ROM) in the Floppy Drive (or Compact Disk Drive).
2. Select Accession from the Activity pull-down menu.
3. Select Load from the File pull-down menu.
4. Then select folder "Security Markings Demo".

Scheduling Work

The next activity in processing accessioned electronic record series is for the Supervisory Archivist to schedule archival work.

1. Open the Archival Register and the Container table to determine which archival accessions remain to be processed.
2. Load the current Work-in-Process (WIP) or Work Schedule by clicking on the Work-In Process icon on the desktop.

WIPID, accession number, container file name, archivist, date, status (P, A, R, D)
Comments

	A	B	C	D	E	F	G	H
1	Task ID	Accession Number	Container Name	Archivist	Date Assigned	Status (PARD)	Date Completed	Comments
2	2003-1	2003.0001	0112.tar	WEU	8/20/2003			
3	2003-1	2003.0001	0113.tar	WEU	8/20/2003			
4	2003-1	2003.0001	0022.tar	WEU	8/20/2003			
5	2003-1	2003.0001	0027A.tar	WEU	8/20/2003			
6								
7								

3. Assign work by entering the year and a sequential task number. Enter the Accession Number and the containers from that accession to be processed. Enter the name (or initials) of the archivist to whom the work is assigned.
4. To save the schedule, select save from the File pull-down menu.
5. Enter in the container list, the status 'I' indicating that the contain is In processing. Close the Accession Register.
6. Notify the archivists of new work assignment.

Starting Work

Archivists start work by performing the following steps:

1. Load the WIP (Work Schedule) by selecting the WIP icon on the desktop
2. Load accession register to identify the accession and file names of the containers they are to process.
3. Select the Archival Processing Tool.
4. Select Accession from the Activity pull-down menu.
5. Select *Create Working Copy* from the File pull-down menu.
6. Copy the containers associated with the assigned work into the Work Area subdirectory labeled with the Archivists Name.

The archivist is now ready to process the contents of the containers. This consists of preservation, arrangement, and review. 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, and records the status of their work in the Work-In-Process file.

	A	B	C	D	E	F	G	H
1	Task ID	Accession Number	Container Name	Archivist	Date Assigned	Status (PARD)	Date Completed	Comments
2	2003-1	2003.0001	0112.tar	WEU	8/20/2003			
3	2003-1	2003.0001	0113.tar	WEU	8/20/2003			
4	2003-1	2003.0001	0022.tar	WEU	8/20/2003			
5	2003-1	2003.0001	0027A.tar	WEU	8/20/2003			
6								
7								

Filtering a File System

Introduction

This activity involves separating the files of a DOS or Windows file system into user-created files and system and software application program files. A **file filter** is a pattern that is matched against files. A blocking filter is a filter in which files that match the pattern are blocked from passing through the filter. The filtering activity sorts the files in a file system into two sets. The files that match the pattern are said to be **blocked**. The other set is said to be **passed**.

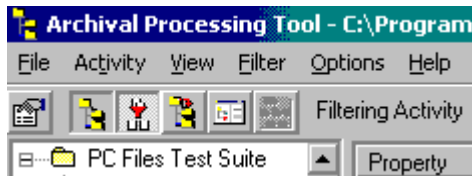
A useful analogy to help understand filtering is a coffee filter. A coffee filter blocks the grounds from passing through the filter and allows the coffee infusion to pass through resulting in an enjoyable cup of coffee. It blocks what you do not want (the coffee grounds), and allows to pass what you do want (the coffee brew).

After you filter a file system, there will be some files you want to keep for further processing and some files you may no longer need. At this point, you will transfer the files you do not want to keep.

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 not need to perform the filtering activity.

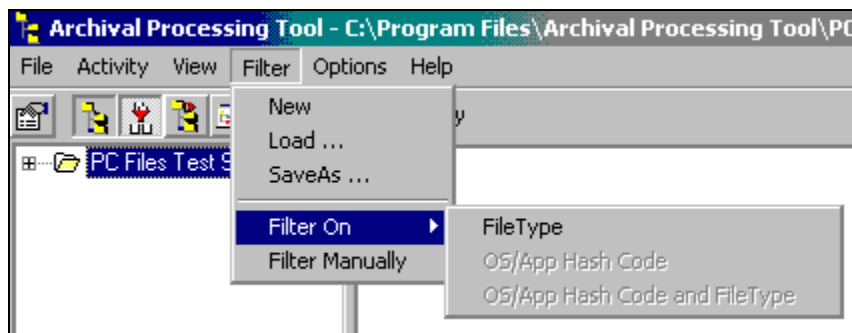
The Filtering Menu

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



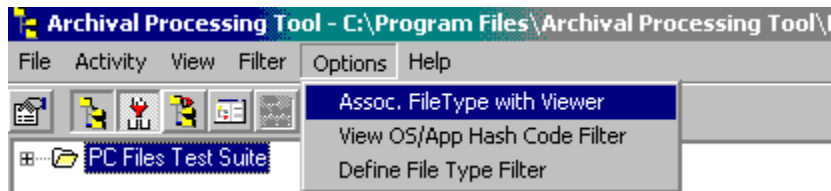
The File, Activity, View, and Help pull-down menus are the same as explained in the section *Exploring a File System* and are the same for every Activity.

Filter: This menu provides the main function 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.



Each of these functions is described below.

Options: This menu allows you to associate a specific viewer with a specific file type; view a specific hash code filter, and define a file type filter.



The option to **Define File Type Filter** shows the entire list of file types than can be automatically identified. The list of identifiable file types is also listed in an appendix to this User's Guide.

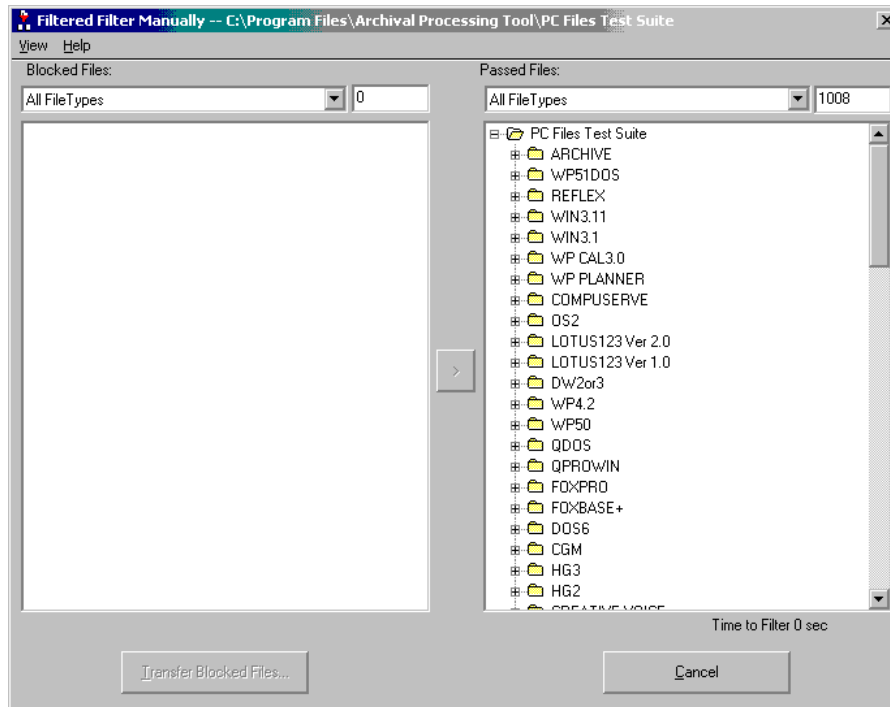
Filter Manually

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.

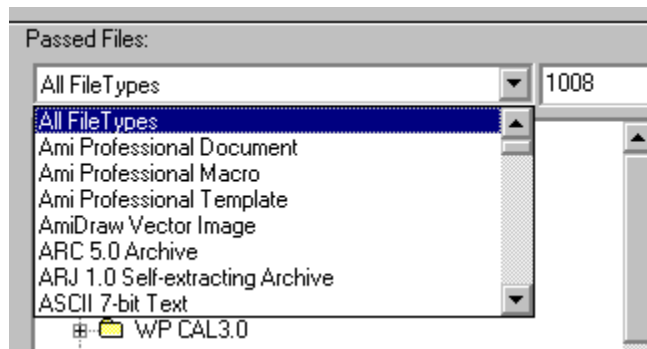
In this mode, the archivist will act as the filter, viewing each file or instance of a file type, deciding what is a user-created file (record) and what is an operating system or software application that might have been used to create the records. The APT will support this process with a user-interface 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, 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 us that filter, and the archivist will not have to make the same decisions twice.

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.

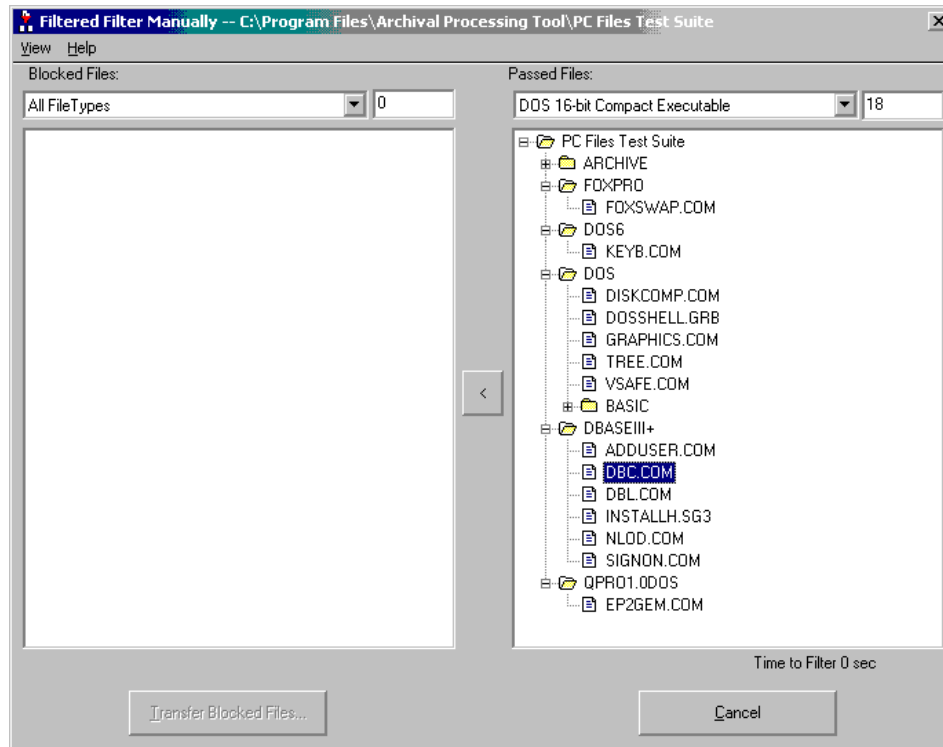


The field at the top of the right pane shows the number of files in the file system. If one selects the down-arrow to the left of the field All File Types, one sees a panel showing the file types identified in the file system.

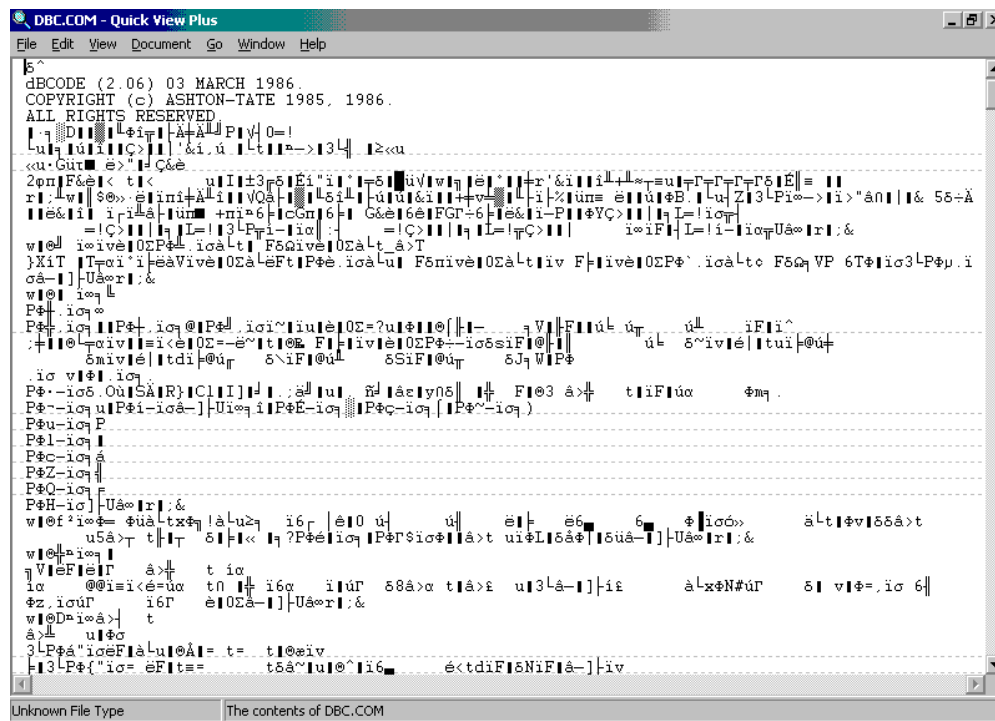


One can scroll down the list to see additional file types that are identified in this file system. The entire list of file types that can be identified is shown in an appendix to this User's Guide.

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.

```
dBCODE (2.06) MS-DOS/PC-DOS ***
COPYRIGHT (c) ASHTON-TATE 1985, 1986
AS AN UNPUBLISHED LICENSED PROPRIETARY WORK.
ALL RIGHTS RESERVED.

*** dBCODE: illegal option *** END RUN   dBCODE
dBASE III RunTime+ Pseudo-Compiler

Usage: dbc [ -cfile -i[prefix] -oprefix -rfile -sprefix file...]

      FLAGS (order arbitrarily, don't concatenate!)

      -c  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 inRan out of memory onPremature end of input
onRead error onWrite error onUnable to open responseSource -
output filename conflict -Unrecognized command inIllegal DO
statement inIllegal RunTime+ SET option inIllegal command macro
(&) inIllegal condition in ON statement inIllegal action in ON
statement in file 0123456789ABCDEF Completed.
*** .prg file not generated ***
Errors:ENDTEXT
*** expected          END

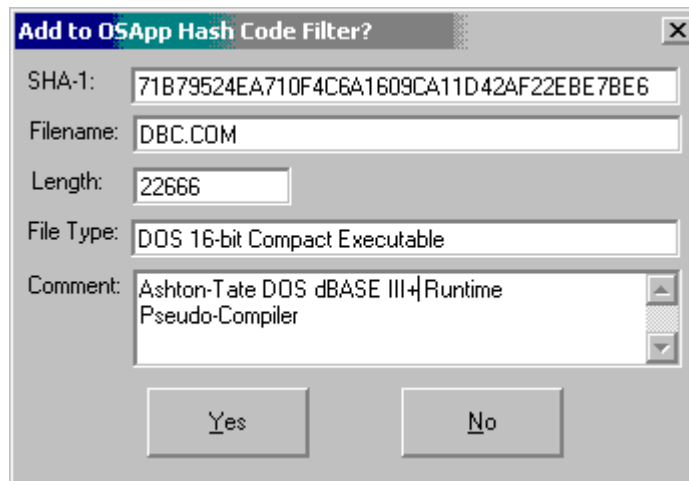
||                               $| .dbg`| .prgh| .src| *** not found!
```

This file is a dBASE III Runtime Psuedo-compiler, not a user-created record. Close the viewer.

You can move the file and its path from the current file system to the left pane by

Step 1: Highlighting its filename.

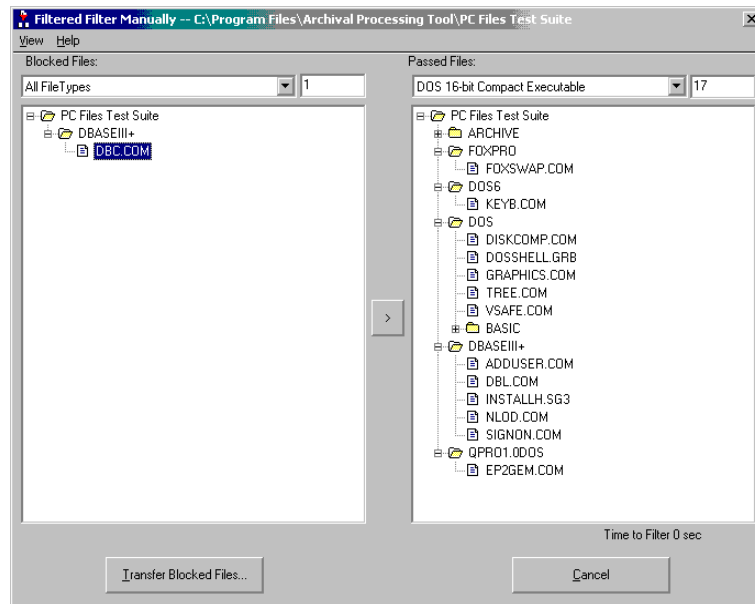
Step 2: Clicking on the arrow between the windowpanes that points to the left windowpane. A dialog box will pop up that shows the properties of the file.



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 comments 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 add button, then that 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 picks cancel 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, 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.

Unknown File Types

[To be written]

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 file system. 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 that you saved your work to. 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.

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. The system and software application files will be saved in a package and the Blocked File System will be closed. You 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.

Saving your work

After filtering the file system, and transferring files you do not want, you should save the remaining passed files.

Step 1: From the main menu of the filtering activity window, select **File>Save As**.

Step 2: In the "File name" field, enter the name you would like to call the file system. Click 'save'. You will use this file system in future activities such as Review and Description. Be sure to note where you store it.

Saving a Filter

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

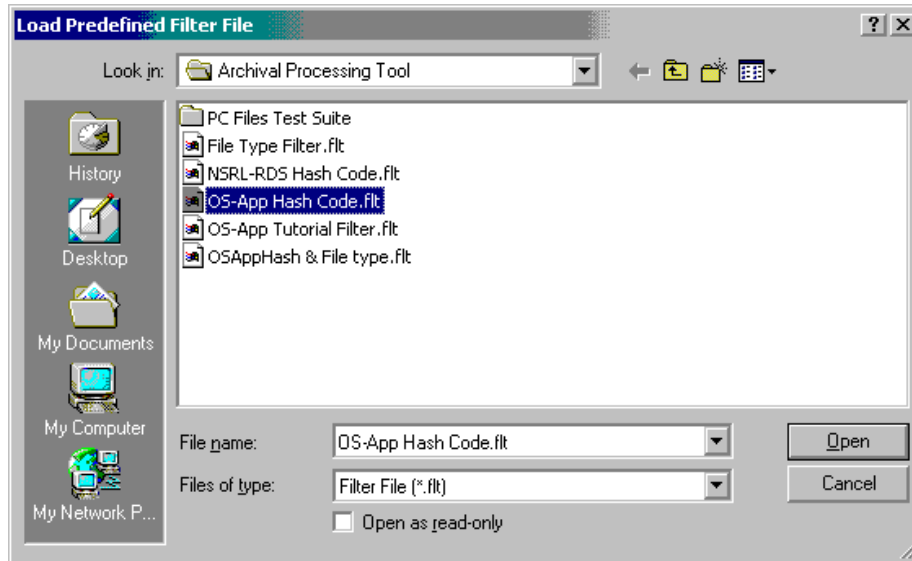
In the next section you will learn how filtering a filing system using the OS/APP Hash Codes Filter can reduce the manual effort required to filter operating system and application software from file systems.

Filter on OS/App Hash Code

Suppose you have loaded 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 don't match will be passed.

Loading a Filter

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



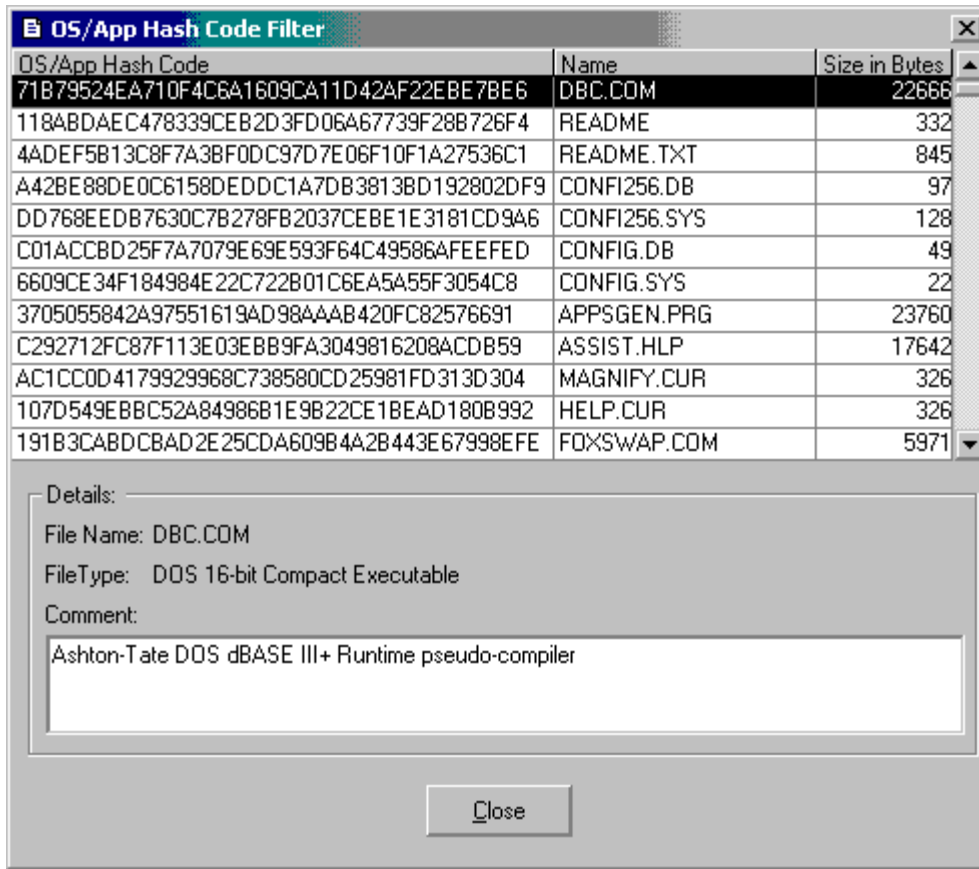
The filter *OS-AppTutorial Filter.flt* is the one created during the process of understanding the Manual Filtering process.

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. It is recommended that you use and extend this filter, since doing so will save you a lot of time.

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).

View OS/App Hash Code Filter

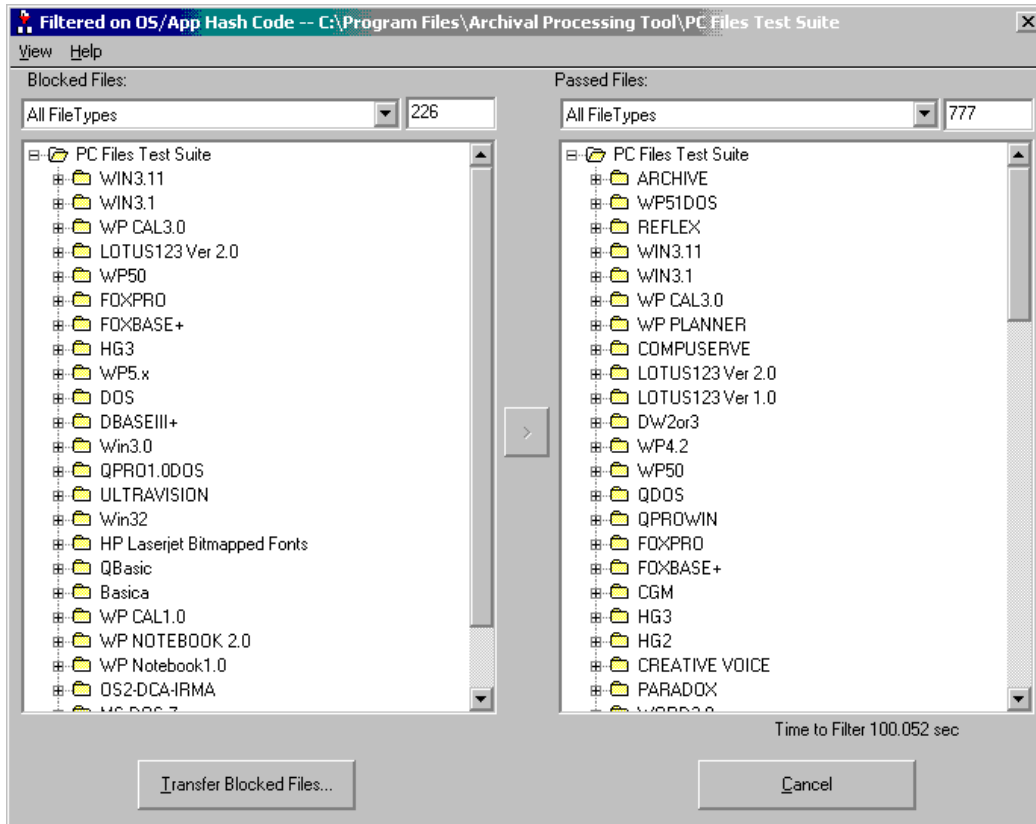
To view the loaded OS/App Hash Code Filter, select Options>View OS/APP Hash Code Filter. You will see a display similar to the following.



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.

Filtering on OS/App Hash Code

From the Filter pull-down menu, select "Filter On > OS/App Hash Code". The file system will then be filtered using the loaded OS/App Hash Code filter. The blocked and passed files will appear.



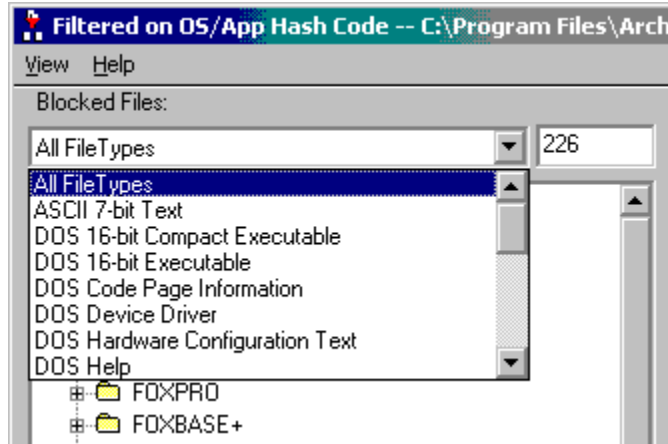
The "Blocked files" field is found above the left pane. These files have been filtered out of the original filing system because their hash codes matched hash codes in the OS-App Hash Code Filter. The number of blocked files is also given.

The "Passed files" field is found above the right pane. These files have passed through the filter. Note the number of passed files.

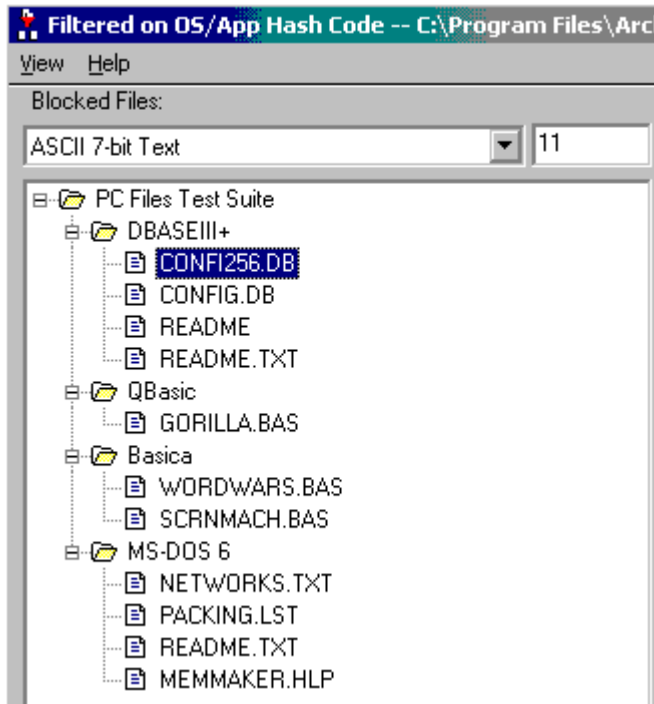
"Time to Filter" is the lower right-hand corner of the screen displays the length of time in seconds necessary to filter the entire file system.

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.

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*:

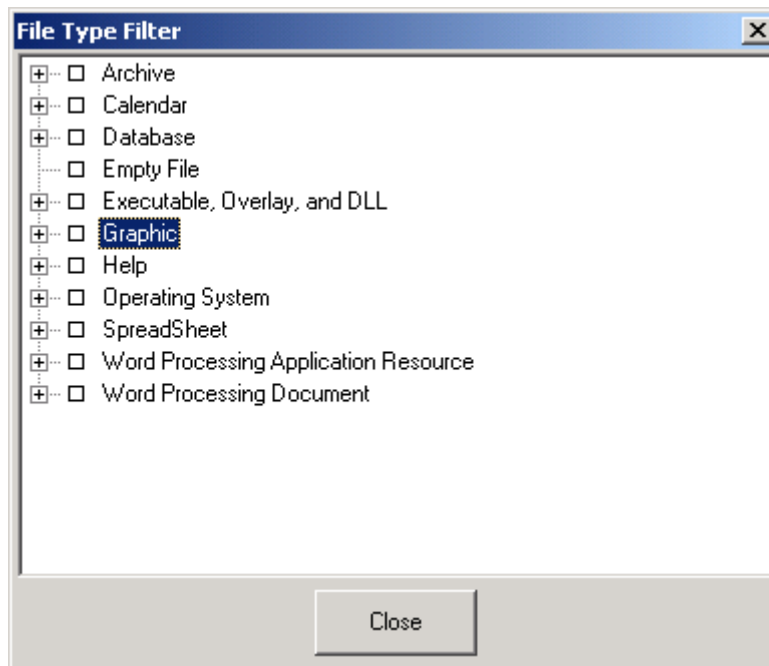
- View the file types of passed files.
- View the individual files.
- Move operating system and application software files to left windowpane.

- Add them to the OS-App Filter.
- Continue this process until all that remains in the right window pane 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 Work-in-Process directory.
- Save the updated OS-App Hash Code Filter.

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. There are other file types that 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 File Type filter from the Options pull-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 or "unarchive" 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 data base 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 data base 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. Unlike source program files, executable files cannot be read by humans. 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, 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.

Lzexe 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 almost always operating or software application files. However, there are important exceptions. For instance, there are some software applications that 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.

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 Document

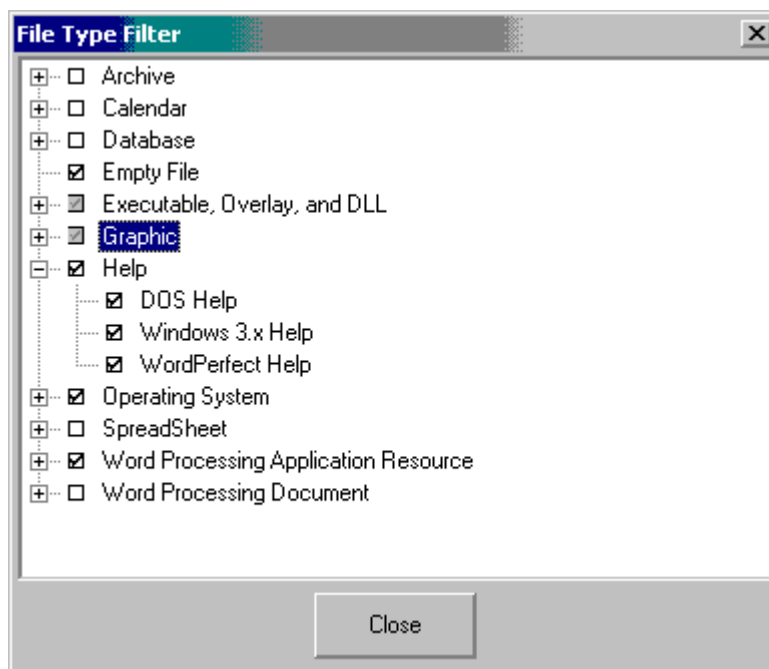
The documents created by users of word processors and text editors are stored in a variety of word processing (or document) file formats. Most of these will be user-created documents, but some may be sample documents.

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.

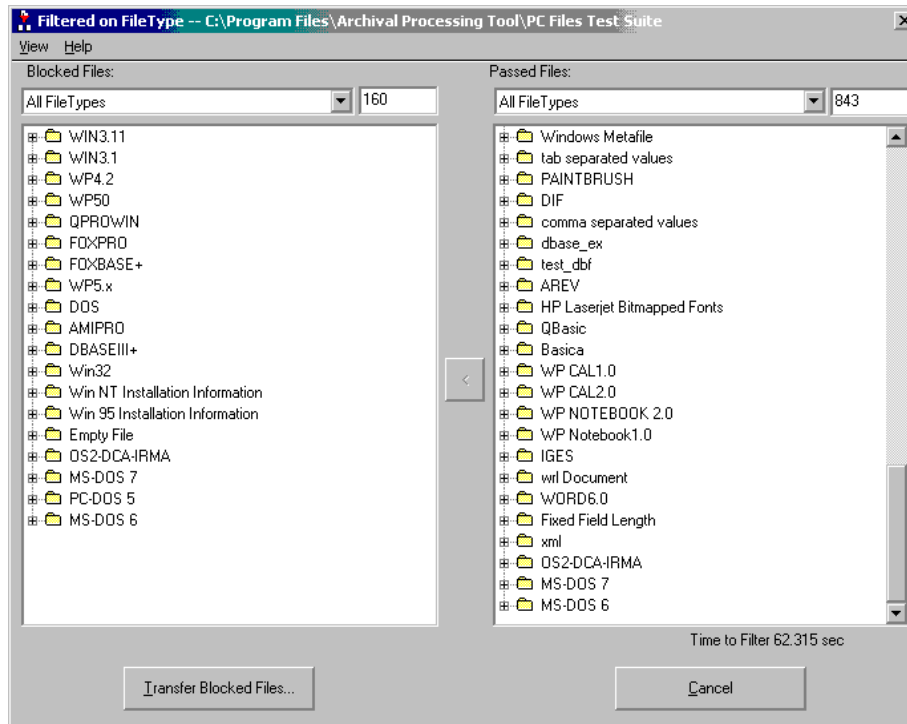
Step 1: From the Options pull-down menu, choose "Define File Type Filter".

Step 2: From this list, select the file types you would like to block by clicking once in the appropriate check boxes, then click Close. In this example, "Empty Files" and "Operating Systems" are checked. We should expect these files to be blocked after filtering.



Filter on File Type

From the Filter pull-down menu, select "On FileType". The file system will then 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.

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 window pane 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.
- Saving the passed files in the Work-in-Process directory.
- Save the File Type Filter.

Filter on OS/App Hash Code and File Type

If the goal is to separate operating system and application software files from files created by users in the course of performing the primary business responsibilities, one will find that there are some files that are created by users for communicating with the operating system or

software applications. These are not files that need to 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:\wbwb4;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:\MKS
set TMPDIR=c:\temp
SET TMP=C:\TEMP
SET TEMP=C:\TEMP
SET BW=C:\BW
SET WINPMT=%s%H%7m%K(Windows Shell)%p%[m%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:mvcdd001 /m:10 /v /e
C:\mouse\mouse.COM
LH /L:1,6384 C:\DOS\DOSKEY.COM
win
```

Other examples of files created to communicate with the operating system are other DOS batch command files, 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).

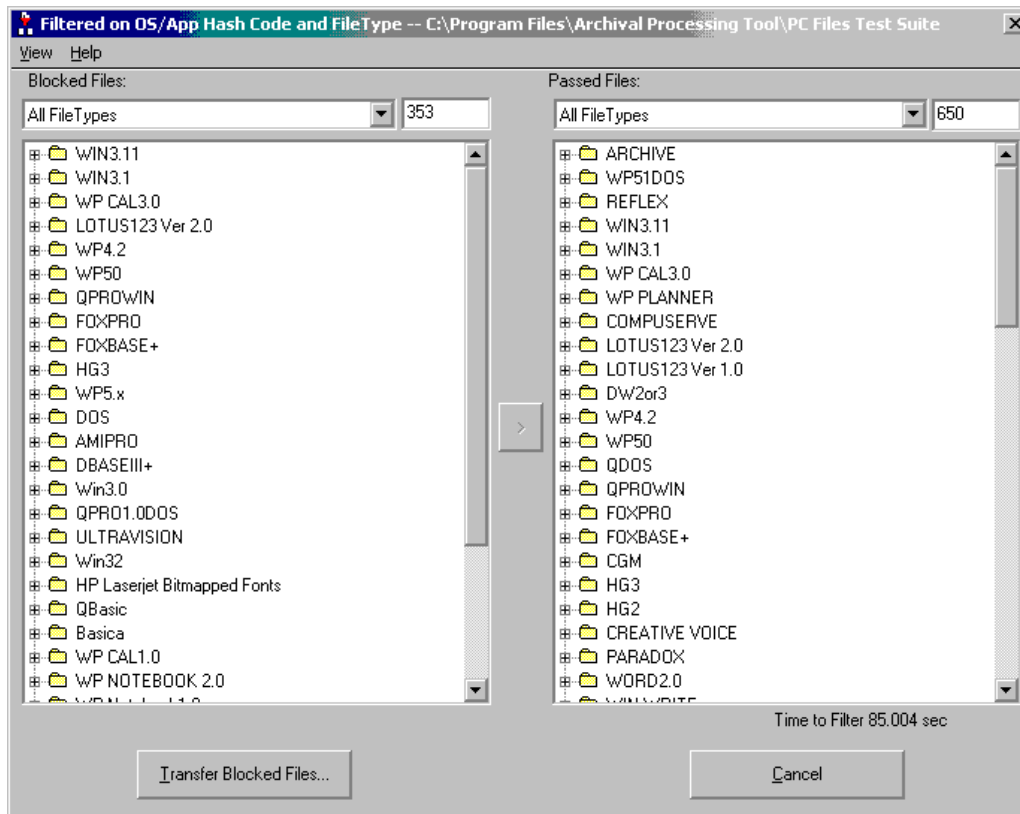
One could block these filters using the OS/App hash code filter, but that would be an ineffective method of filtering as each filing system filtered would likely have a different AUTOEXEC.BAT file. A mixed strategy of filtering on OS/APP hash codes and the DOS Command Batch Text and DOS Hardware Configuration files is a better approach.

Define OS-App Hash Code & File Type Filter

You can construct an OS-App Hash Code & File Type Filter by loading an OS-App Hash Code Filter and then defining a File Type filter. The File Types that can be defined are listed in an appendix. The Filter should be saved with a file name such as OS-App hash code & File type to distinguish it from the other two kinds of filter.

Filter on OS-App Hash Code & File Type

If one selects filter on OS/App Hash Codes and File Types, both the OS/App Hash Codes and the checked file types in the file type filter will be used to filter the loaded file system. If the hash code of the file matches the hash codes in the OS/APP hash code filter, or the file is among the formats in the list of checked file types, then file is blocked and included in the left window pane, else the file is passed and included in right window pane.



Note that this combination filter blocks more than either filter would individually block.

Manually Filtering the Files Passed through the OS-App Hash Code & 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 be prompted as to whether you want to add the hash code of the moved file to the OS-App Hash Code Filter (which also includes File types).
- 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 Work-in-Process directory.
- Save the OS-App Hash Code & File Type Filter.

As a result of experiments with the different kinds of filter, it is concluded that this is the best kind of filter to use in separating 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.

Epilogue

A utility program is being created that reads all of the TAR files containing operating system and software application files filtered from the original files systems. It creates a TAR file containing a single copy of operating system and software application files. It should be reviewed to ensure that no user-created files that should be preserved were incorrectly identified as operating system or software applications that do not need to be preserved.

Preserving Digital Files

Personal Computer hardware of 10 years ago (IBM AT, IBM PS/2) is obsolete. The operating systems (DOS and Windows 3.x) that operated on that hardware are also obsolete. The application software (Word 2.0, Lotus 1-2-3, 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 forms.

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

Extract Archived Files

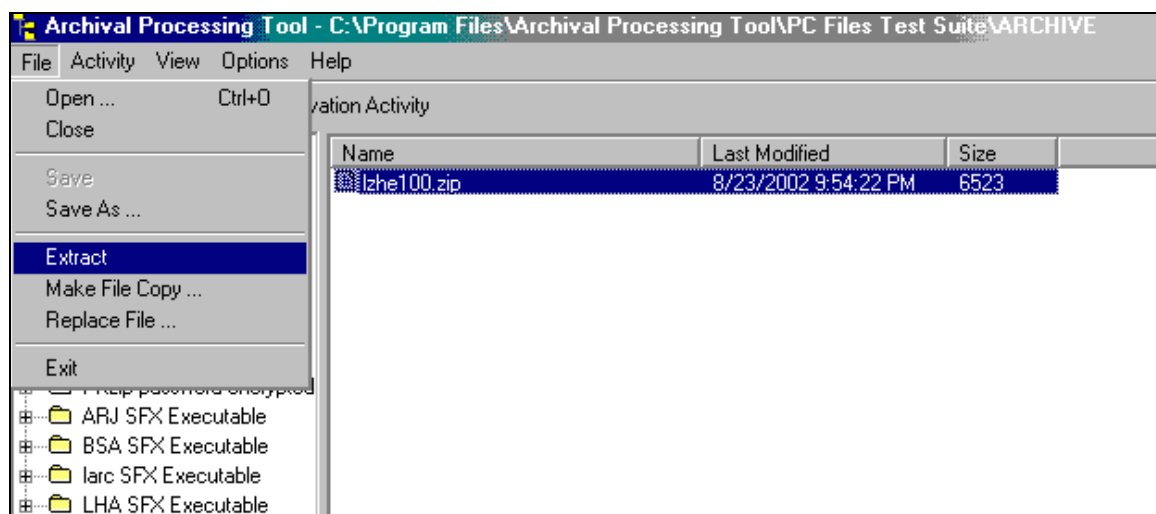
User-created files must be extracted from archive files, e.g., PKZIP, ARC and ARJ files, in order to be viewed. Quick View Plus supports extracting files from Zip, ARC, ARJ, and LHA files. However, Quick View Plus cannot extract other types of archive files, including self-extracting archive files. All archive files should be extracted using the following procedure.

The file system containing the archive file must be loaded into the APT.

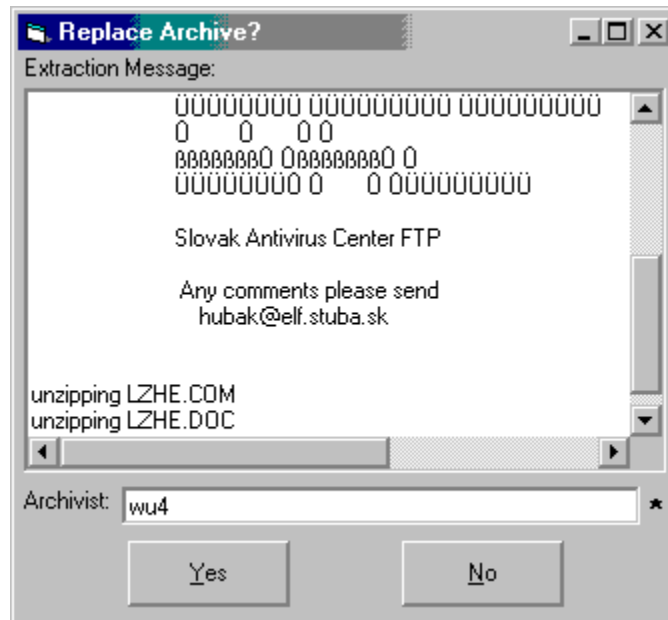
Step 1: Select Preservation from the Activity pull-down menu.

Step 2: Highlight the archive file from which files should be extracted.

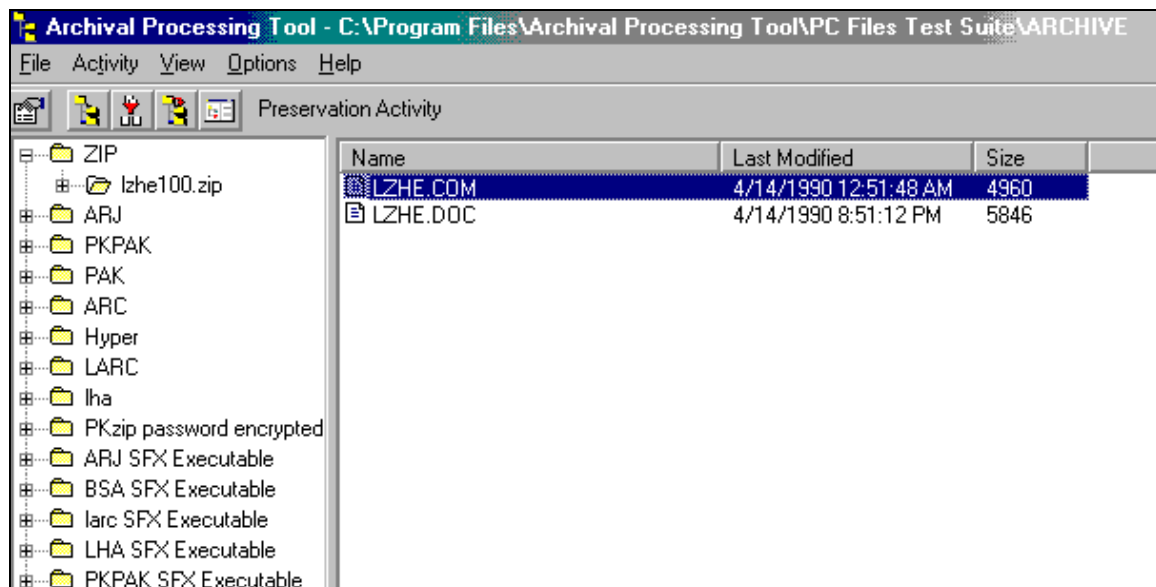
Step 3: Select Extract from the File Pull-down menu.



A dialog box will be displayed that indicates whether the extraction can be performed and if so that shows the filenames of the extracted files.



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.



Step 4: You can view the files by double-clicking on the file name.

Step 5: For each file, the manifest can indicate the name of the archive it was extracted from.

Self-extracting archive files are not executed for a number of reasons. One is that the extracted files might be saved over another file of the same name. Another is that they need to be checked for viruses. Instead, they are passed to a program that ignores the executable header, and extracts the files.

You can add Archive Extraction Utilities to the APT by choosing the Assoc. Archive File Type with Extractor option on the Options pull-down menu.

Decrypt Password Encrypted Files

Some of the PC application software of the late 80's and early 90's included the built-in capability to encrypt a file using a password. During experiments in processing the contents of the Bush hard drives, password encrypted files were detected by Quick View Plus. 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. One of these products, Password Recovery Toolkit (PRTK) from Access Data, was acquired for experiments in recovering the passwords. During installation of the APT it is configured to work with the APT as follows:

When a password encrypted file is encountered (which is indicated either by Quick View Plus, or by the File Type Identifier)

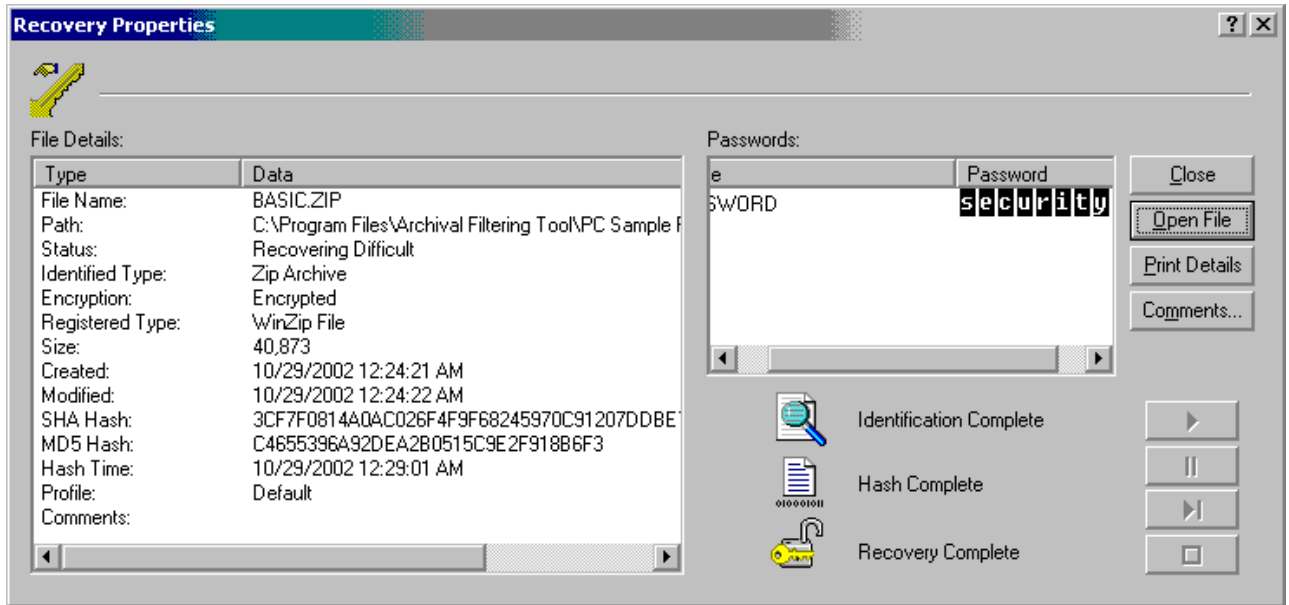
Step 1: Select Preservation from the Activity pull-down menu

Step 2: Highlight the filename of the encrypted file.

Step 3: Select Decrypt from the File pull-down menu.

If the PRTK has a procedure for recovering passwords for files of that type, it will attempt the recovery. In the cases of a weak encryption method, the password may be recovered in a few seconds to a few minutes. Stronger encryption methods may require hours to days of processing to recover a password.

The Password Recovery Toolkit was asked to recover the password of a PKZip password encrypted test file. It displays some properties of the file and the recovered password.



It recovered the password "security" used to encrypt the files internal to the zip archive in just a few seconds.

Step 4: You can open the file using a copy of the original application used to encrypt the file, and use the password to decrypt the file.

Step 5: The file can be viewed with Quick View Plus and copied back to the file system with a different file name, but associated with the original password encrypted file.

Step 6: The password is recorded in the manifest, as is a record of the fact that the file was decrypted. The copy of the encrypted file in temporary storage, not the original in the file system, is erased.

You can add password recovery utilities to the APT by selecting the "Associate File Type with Decryption Utilities" Option from the Option pull-down menu.

Repair Corrupted Files

During processing you may encounter a file that can be read and recognized by Quick View Plus but cannot be displayed properly. For example, Quick View Plus displays a message that it cannot display the file or display a blank screen. A "loop" may occur in the document so that moving the cursor down brings you to an earlier part of the document. The file header or function codes in the text of the file have been corrupted.

It may be possible to repair the file using a file repair utility.

The file system with the corrupt file should be loaded.

Step 1: Select Preservation from the Activity pull-down menu

Step 2: Find and highlight the corrupted file.

Step 3: Select Repair from the File pull-down menu.

Step 4: If the file type is known, and there is a file repair utility for the file type of the highlighted file, an attempt will be made to recover the file.

- a) If the file is recovered, the repaired file will replace the original file. You should view the file to check that it has been recovered.

- b) If you are told that the file could not be repaired, you may want to try extracting the text from the file. See Convert Files to other Formats.

Step 5: In the Manifest, "Preservation: Repaired File" will be added to the file section for this file.

In version 2.05 of the APT, there are file recovery utilities for the following file types:

File Type
dBase III database
dBase III database with dbt
dBase III memo
dBase IV Database
dBase IV database with dbt
dBase V database
FoxPro Database with fpt
PKZip 1.0 Self-extracting Archive
PKZip 2.04 Self-extracting Archive
PKZip Archive
WinZip 32-bit Self-extracting Archive
WordPerfect 5.0 Document
WordPerfect 5.1/5.2 Document
WordPerfect 6.0 Document

You can add file repair utilities to the APT by selecting the Assoc. File Type with Repair Utility option under the Options pull-down menu.

Convert Files to Other Formats

You may encounter a legacy file type for which there is not viewer. As an alternative to writing a viewer for the obsolete file type, you may want to consider converting the file to a current format for which there is a viewer. For instance, there is no viewer for Advanced Revelation. Advanced Revelation databases are stored in two files with the same filename and two different file extensions, lk and ov. There is a conversion package that runs under Advanced Revelation that will read the two database files and convert them to comma-separated values, dbase, or XML formats, which have viewers.

The file system containing the file to be converted must be loaded into the APT.

Step 1: Highlight the file that you want to convert.

Step 2: Select the Preservation Activity

Step 3: From the File pull-down menu, select Convert.

Step 4: Select the target file format.

Step 5: Select Go.

Step 6: The file will be converted and stored with the name "Converted filename"

Step 6: "Preservation: Converted File" will be added to the file section for this file in the Manifest.

Version 2.05 of the APT has conversion utilities for converting:

Source File(s)	Target File	Conversion Utility
Advanced Revelation Database (lk, ov)	CSV or XML	
Borland Reflex	CSV or XML	
Harvard Graphics 2.0 chart	Harvard Graphics 3.0 Chart	Harvard Graphics version 3.0
WordPerfect 5.x	ASCII Text	

You can add file conversion utilities to the APT by selecting the Assoc. File Type with Conversion Utility option under the Options pull-down menu.

Arranging a File System

Introduction

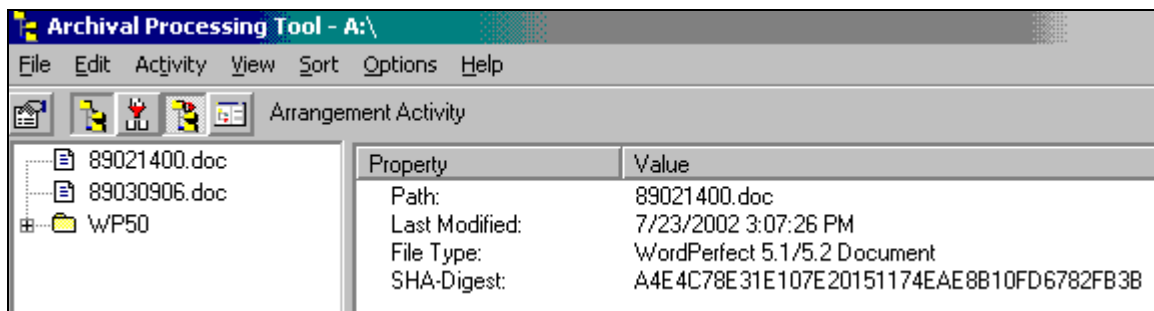
Arrangement is "the intellectual and physical processes and results of organizing documents in accordance with accepted archival principles, particularly provenance, at as many as necessary of the following levels: repository, collection, record group for Fonds, subgroups, series, subseries, file unit, and item." Intellectual/administrative arrangement by provenance provides information about records creator. Intellectual arrangement by file structure provides information about records. The physical arrangement of files refers to arranging documents within folder and folders within boxes. For electronic files, it refers to arranging files within directories and directories within a digital file system.

An archivist needs system assistance in physically arranging PC document files within folder in a file system. He may need to rearrange files that are in a DOS directory order into a more logical order, e.g., by file name or by date last modified. DOS and other operating system such as Unix and Windows NT do not provide required behavior for rearranging files into a different logical order.

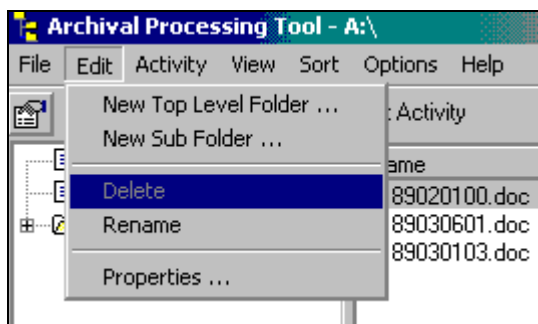
If operating system and application software files need filtered out of a file system, the filtering activity should be performed before arrangement.

The Arrangement Menu

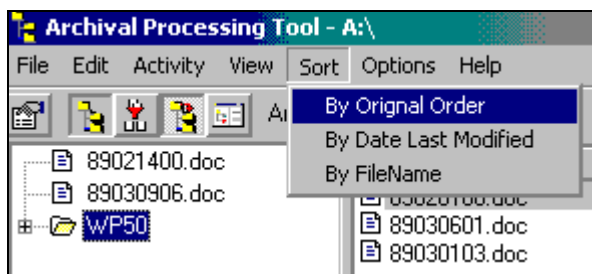
If a file system is loaded and the arrangement activity is selected, seven pull-down menus will be available. The File, Activity, View, Options and Help pull-down menus are the same as explained in the section Exploring a File System and are the same for every activity.



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.

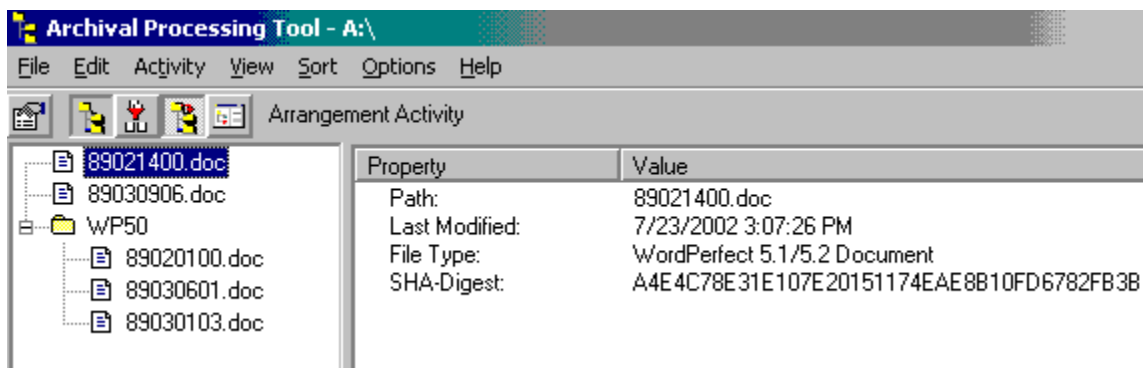


Sort: This pull-down menu allows you to sort the contents of a folder by date last modified, file name, or back into its original order.

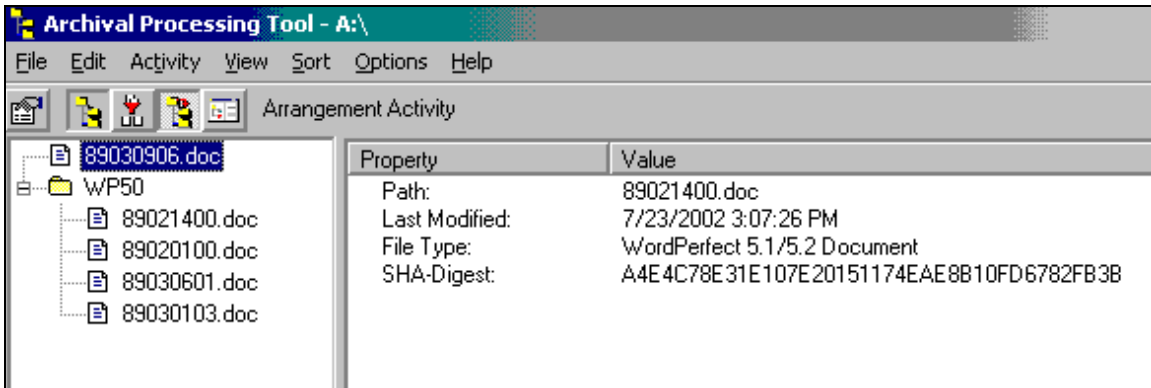


Moving a File into a Folder

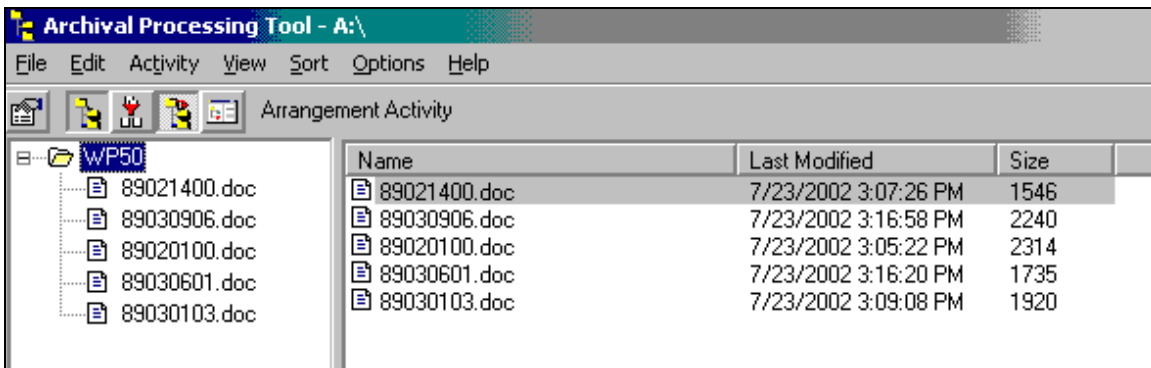
Suppose that the loaded file system appeared as follows. The files have all been viewed and are all WordPerfect 5.x documents. However, two of them are in the root directory.



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.

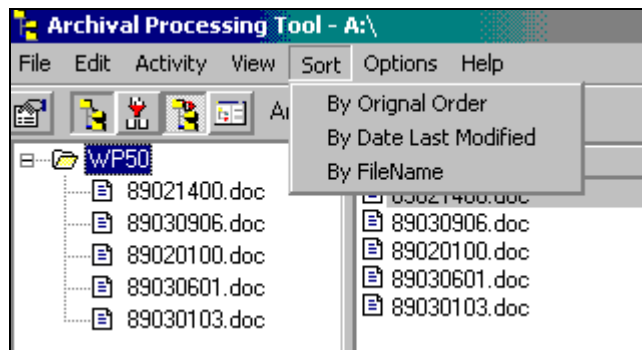


You can do the same for the other file in the root directory.

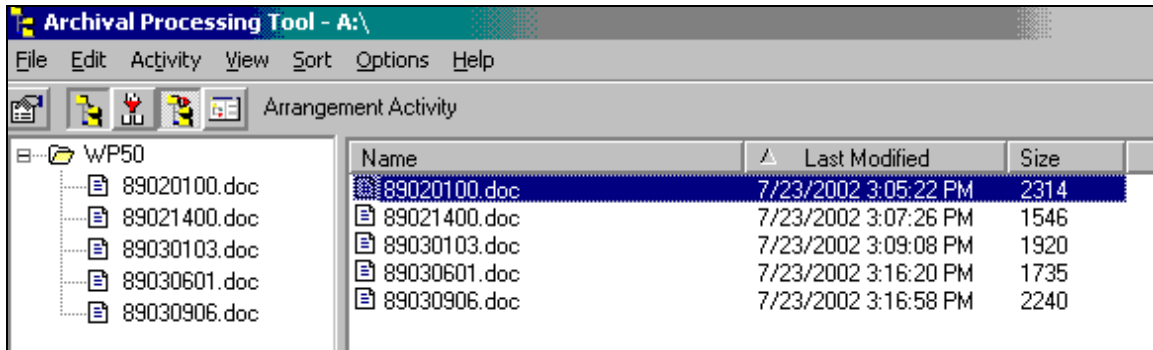


Reordering the Files in a Directory

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 data last modified. Suppose that you believe a more logical order for these files is by ascending 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.

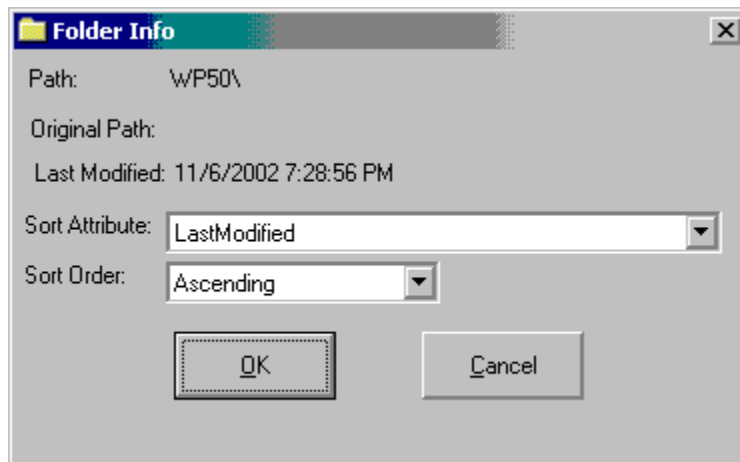


Alternatively, you could have clicked on the column header Last Modified, and the files would have been sorted in ascending order of date last modified. If you click on that header now, the files would be sorted in descending order of date last modified.

The only way to get the files back into their original order is by selecting that option from the Sort pull-down menu.

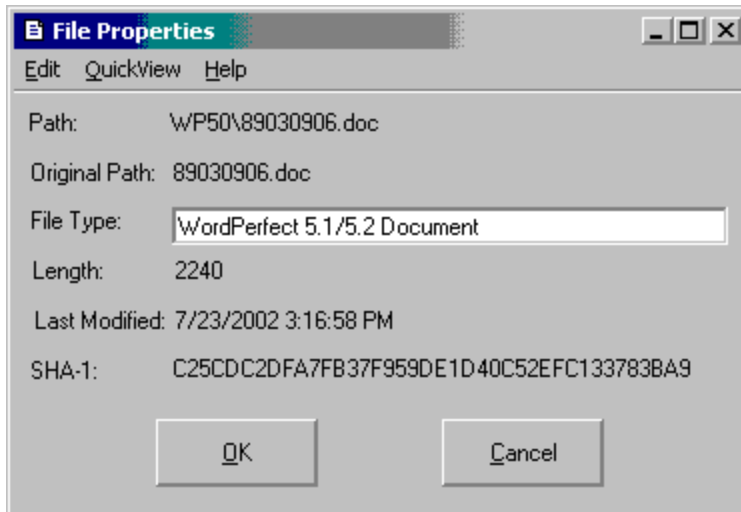
Viewing the Properties of a Folder or File

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



Here we see that 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.

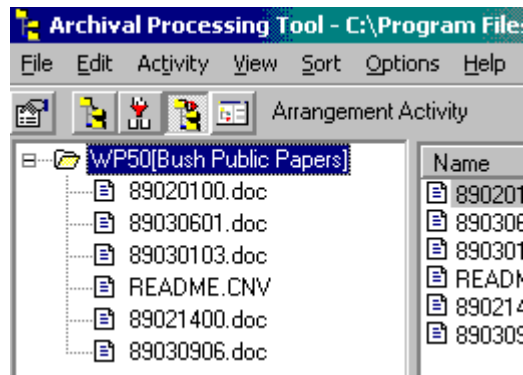


In this case the filename of file that was moved was highlighted. The current path to the file is shown and it's original path.

When a file system that has been arranged is saved as a TAR archive file, these attributes will be saved in the manifest of the archive file.

Renaming a Directory (Folder)

DOS directory names could only be up to 8 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 editing the folder title. Select **Edit > Rename**. The files in this folder were taken from the Bush Public Papers, so let's use that as the name. Rather than delete the WP50, we will follow the convention used at the Bush Presidential Library of enclosing additions to a folder name in square brackets.



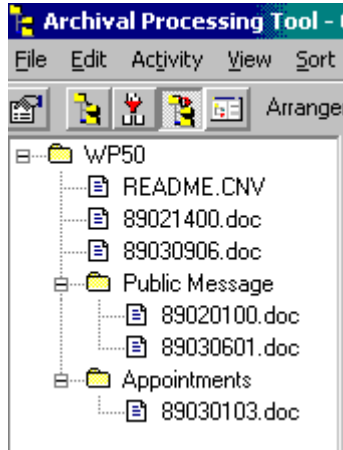
Creating New Folders

Suppose that upon examination of the files one discovers that some were Public Statements and 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.

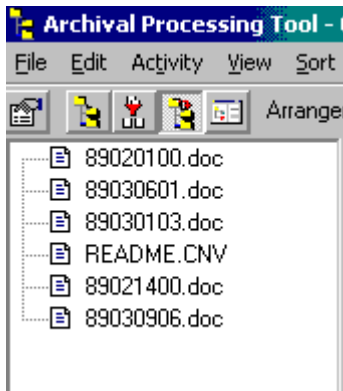
Step 2: Select **Edit > New Subfolder**. A pop-up dialog box will appear.

Step 3: 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.

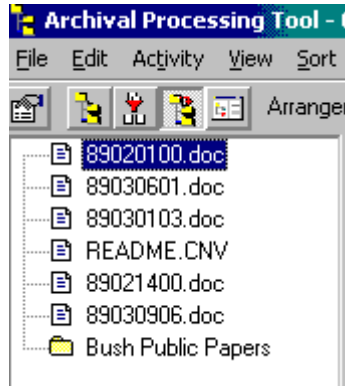
If a file system does not have a top-level folder but is just a root directory of files, you can create a top-level folder.



Step 1: Highlight any one of the files in the root directory.

Step 2: Select **Edit > New Top-Level Folder**. 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.

Reviewing a File System

Introduction

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

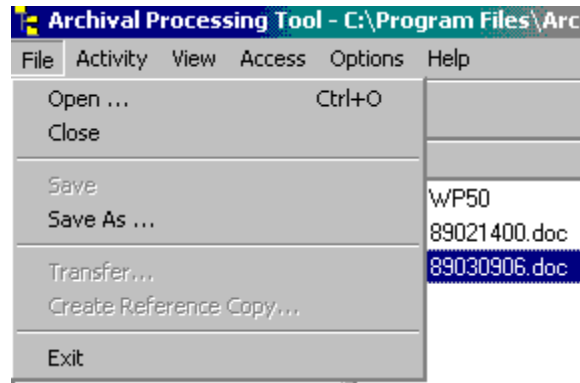
Archivists must review Presidential records and Donor documents for these access restrictions. Archivists need assistance in reviewing Personal Computer files and in recording the results of their review.

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.

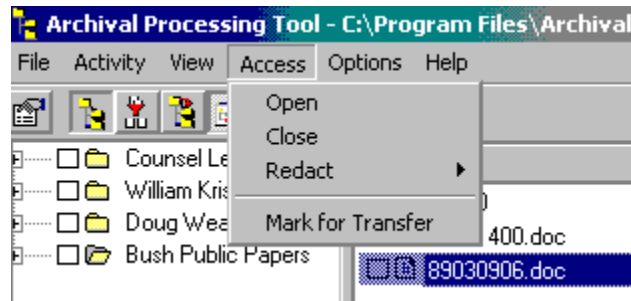
The Review Menu

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

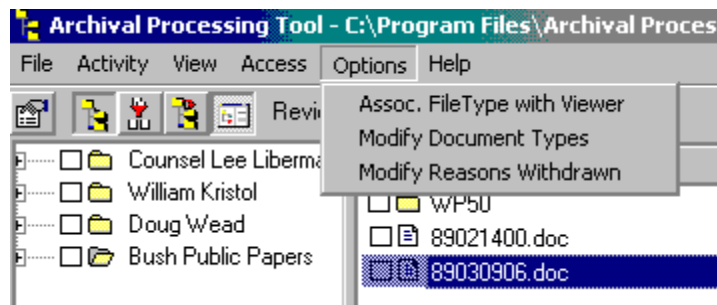
File: In addition to supporting opening of a file system to be reviewed and closing or saving the file system reviewed, this menu supports transferring files out of the file system that are not records, and creating a reference copy of opened and redacted files in the reviewed file system.



Action: There are three actions that this menu supports: Opening a file because there are no access restrictions, Closing a file because there are access restrictions, and Redacting a file because some of the information can be opened but parts have access restrictions. From this menu, you also mark a record for transfer out of the file system.



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



Reviewing a Record

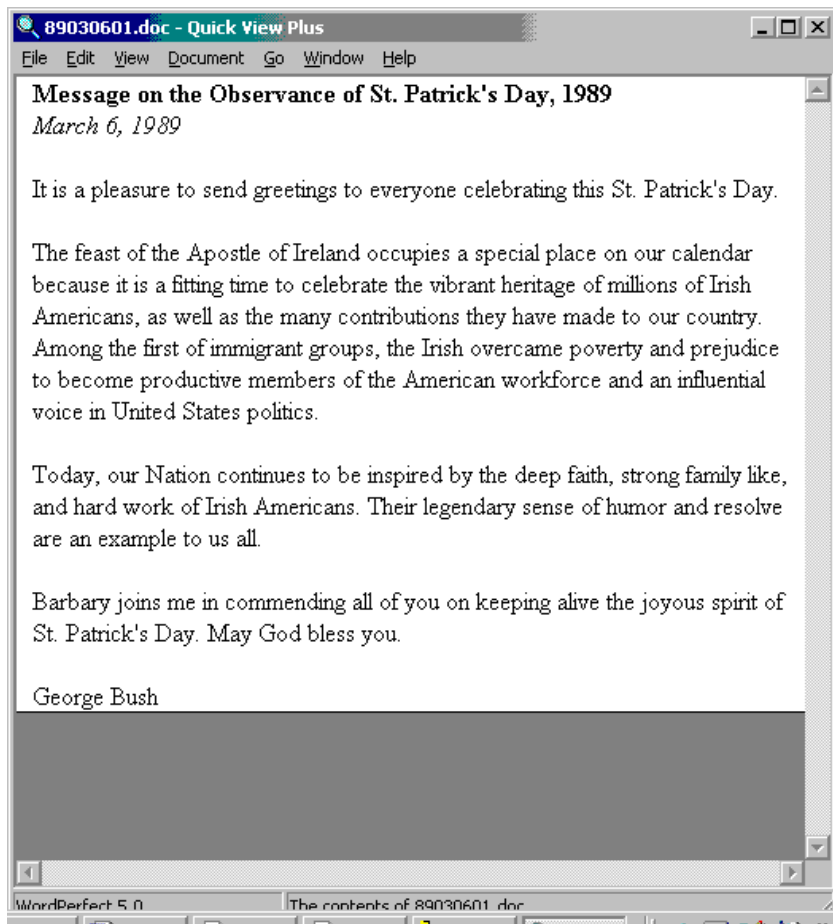
You can review the documents in a file system by viewing each one, reading it to determine whether there are any access restrictions, and then using the Action pull-down menu to open, close, or redact the document, or to transfer the document to another location, e.g., Library.

It should not be possible to open, close, redact or transfer a document 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 hasn't, it is not possible to select the actions from the Action pull-down menu. If the file has been viewed,

each of the actions are possible. [Note: a more restrictive enforcement of this rule would require that the file be currently viewed in order to take a review action.

Open File

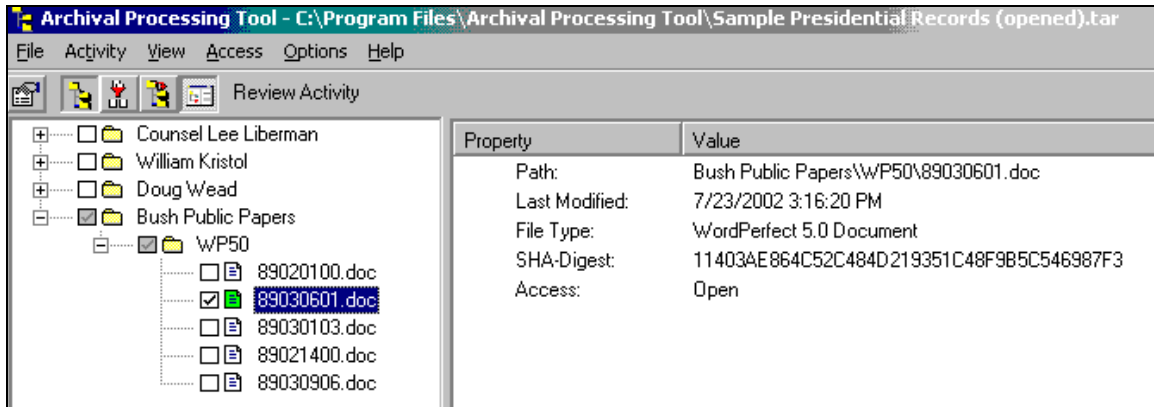
Suppose you had viewed file 89030601.doc by double clicking on the file name. You have decided that this is Presidential message released by the White House and previously published by the Office of the Federal Register, NARA. There are no restrictions on the release of this record, it is a copy of a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



To open it:

Step 1: Highlight the file name by clicking on it once.

Step 2: From the main review menu, select **Action>Open**.

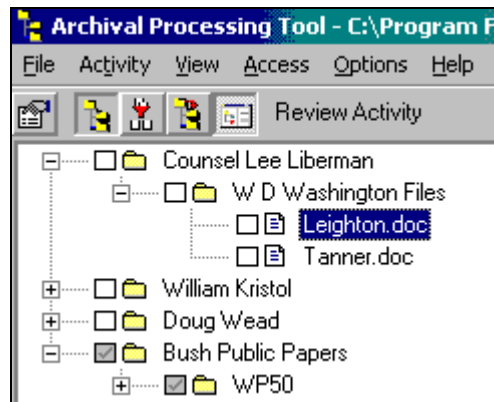


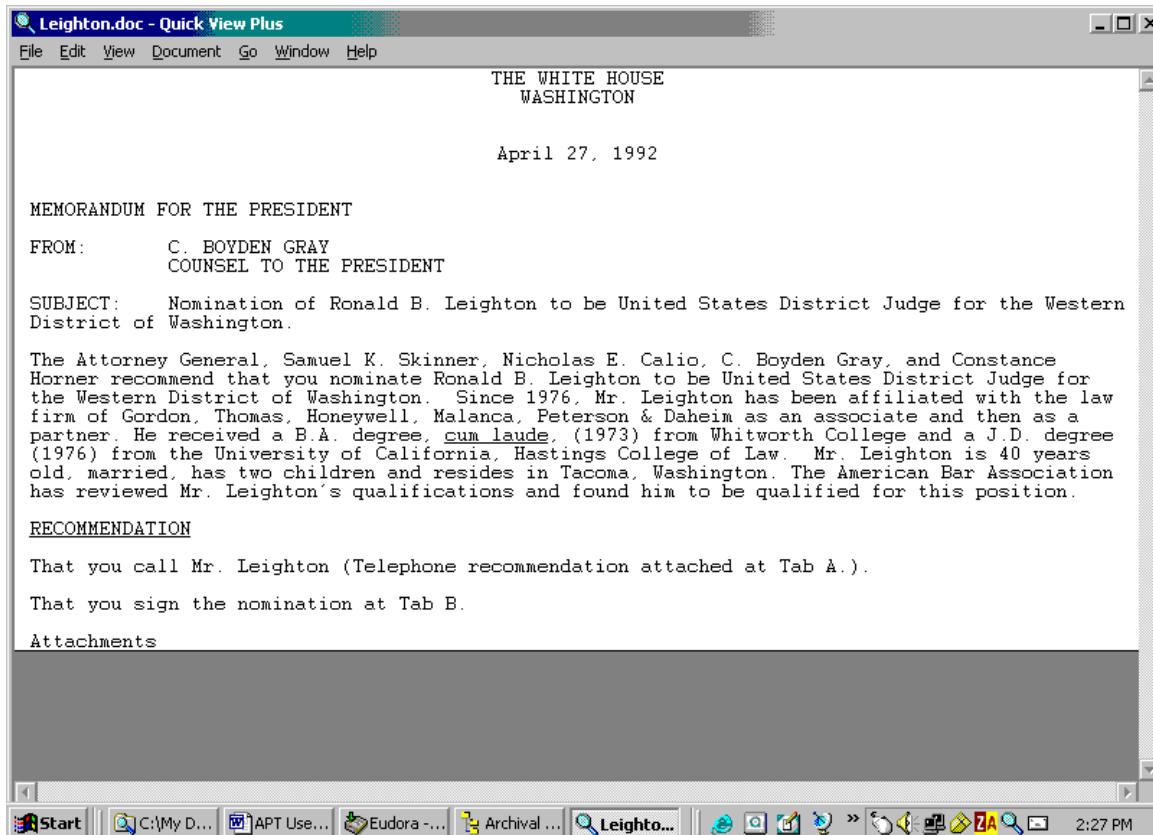
The check in the box preceding the file name indicates that the file has been reviewed. The green document icon indicates that the file has been opened, i.e., there are no access restrictions.

Close File

If there are FOIA exemptions, PRA restrictions or donor restrictions on the release 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 Leighton.doc.





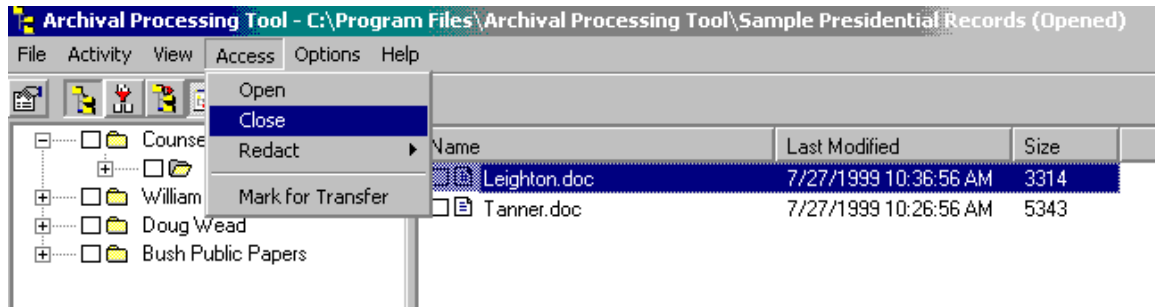
After reading the record you conclude that access to the entire document must be denied because of the Presidential Record Act restriction on release of documents involving confidential advice to the President, in this case a recommendation regarding a judicial nomination. To close this document, you must return to the main window, but don't close the window that the document is displayed in, because you will need information from this document in carrying out the withdrawal or closure process. 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

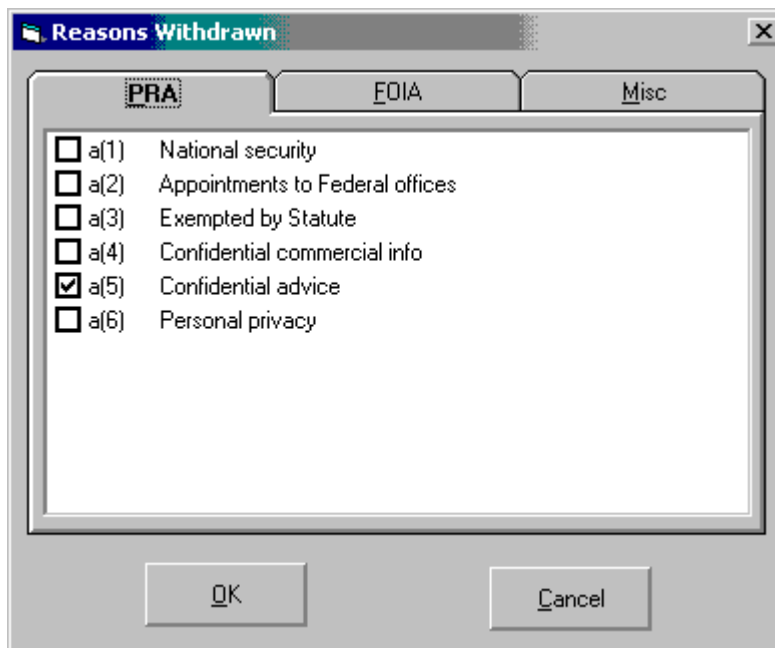
To close a record:

Step 1: Highlight the filename of the record by clicking on it once.

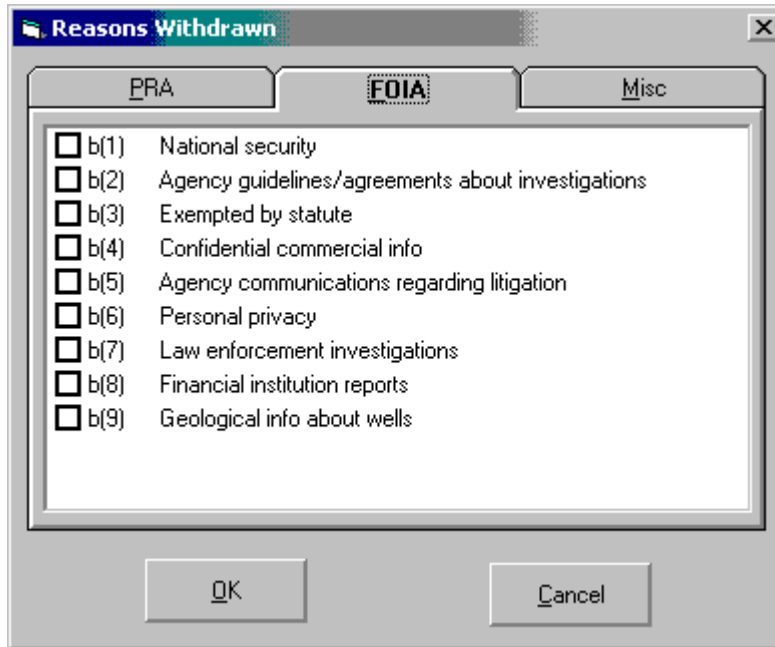
Step 2: From the main review menu, select **Action>Close**.



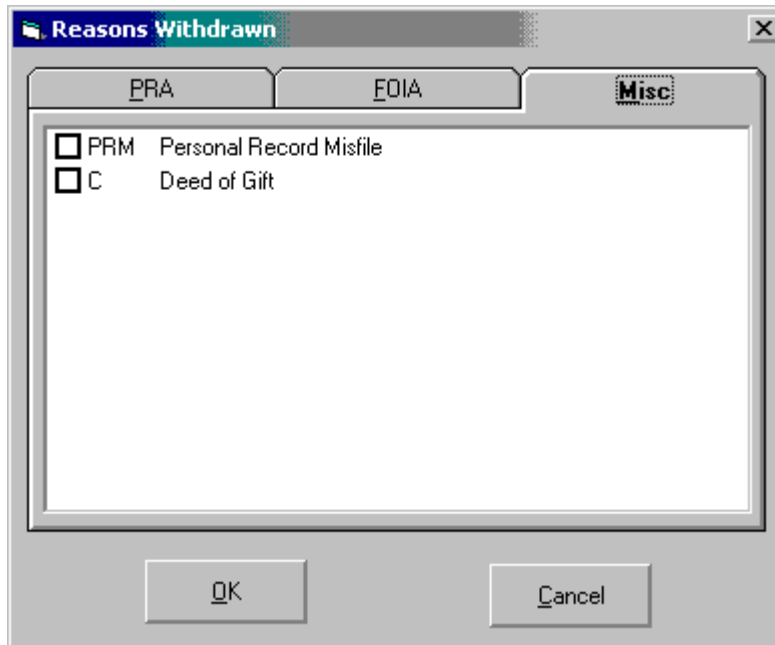
Step 3: The “Reasons Withdrawn” dialogue box will appear. Select the reason(s) for withdrawal and select OK, which will record these reasons for withdrawal.



There 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 other than PRA and FOIA restrictions.

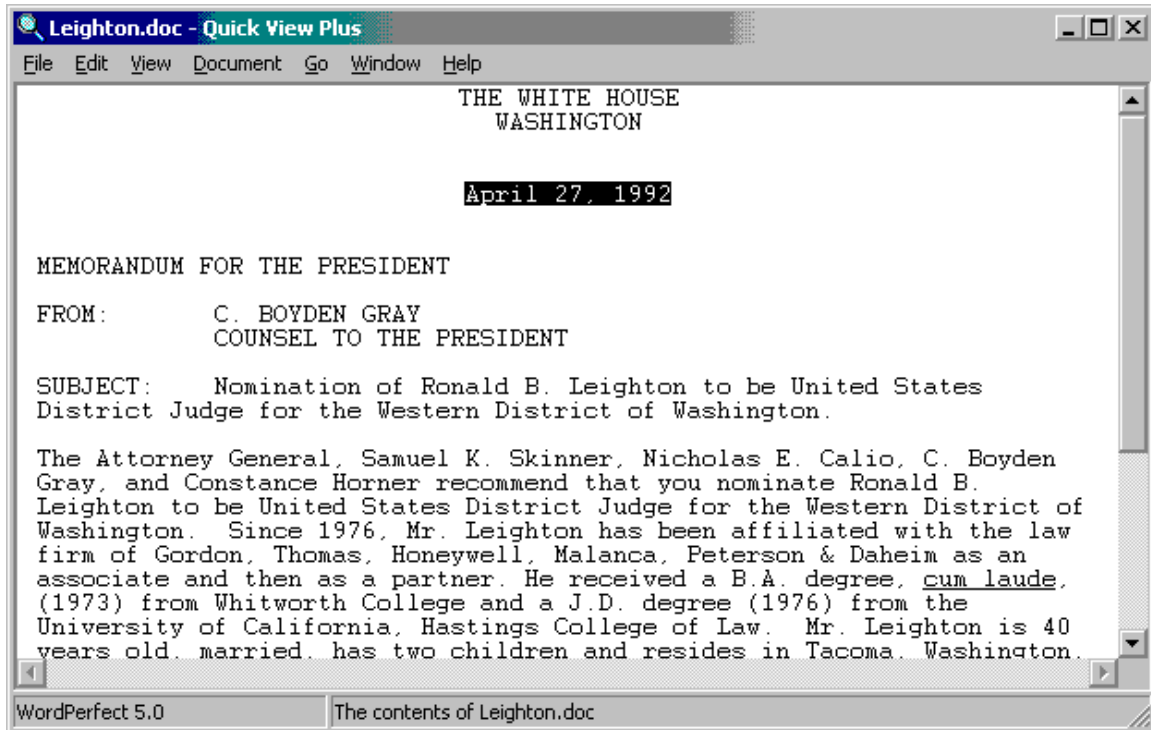


Step 4: The "Closure Info" dialogue box will then appear. Complete the required fields.

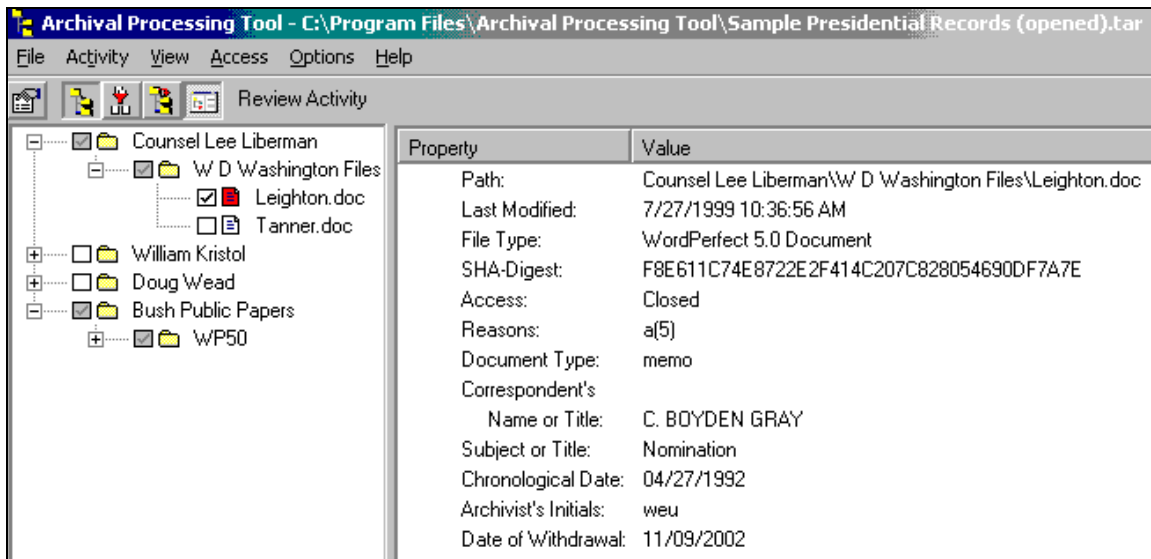
You should indicate the Document Type by selecting the downward arrowhead to the right of the Document Type field, and then selecting the document type from that list. If document type of this document is not in the list, cancel the closure information box, and select **Options>Edit Document Type List** and add the needed document type to the list. Then restart the process of closing the document.

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 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 window pane. If any of this information seems incorrect, it can be edited by highlighting the filename and selecting **View>Properties**



Redaction

Suppose that in reviewing a file displayed with the Quick View Plus Viewer, you discover some information that is exempt from release under FOIA or restricted from release under provision

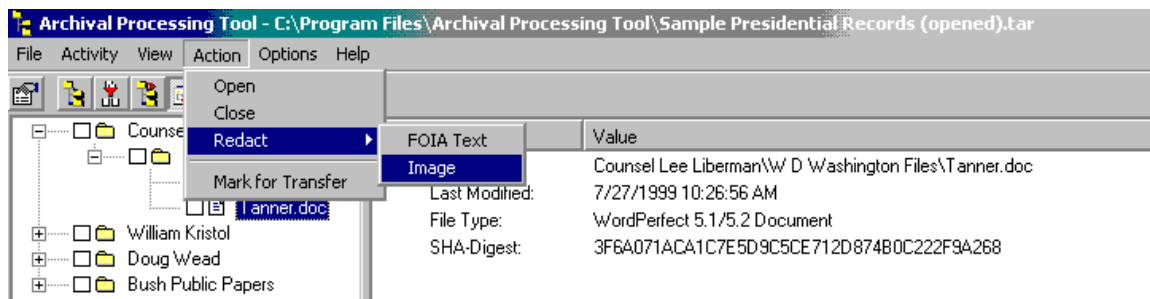
of the PRA. However, the document has significant content that is not subject to access restrictions. You might decide to redact those portions that are subject to access restrictions, close the original document and open the redacted document. The APT supports redaction of document images using Kodak's Imaging for Windows, which comes with Windows NT and Windows 2000.

Redacting Document Images

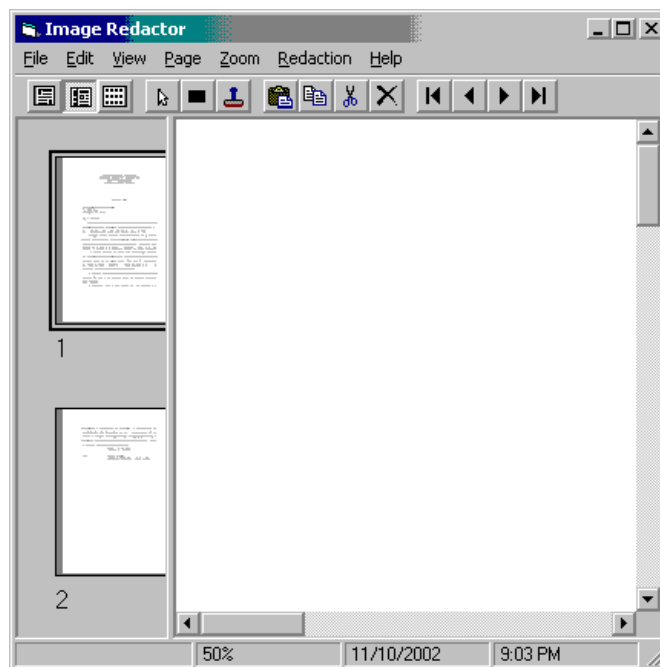
Suppose that we were viewing the document with filename Leighton.doc and deciding that it had an access restriction that we would like to redact.

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

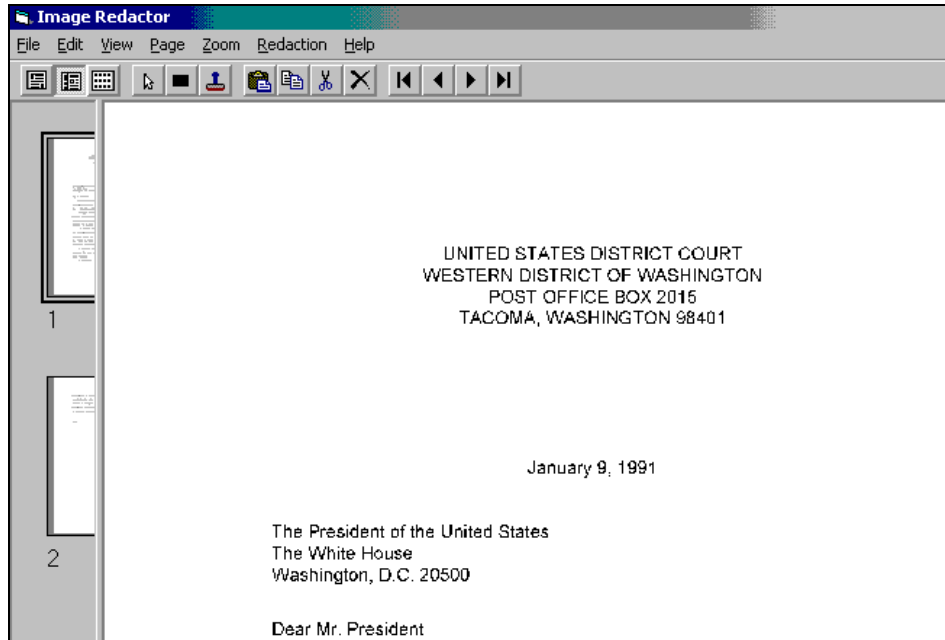
Step 2: The file previously viewed in Quick View Plus should still be highlighted. From the APT menu bar select **Action>Redact>Image**.



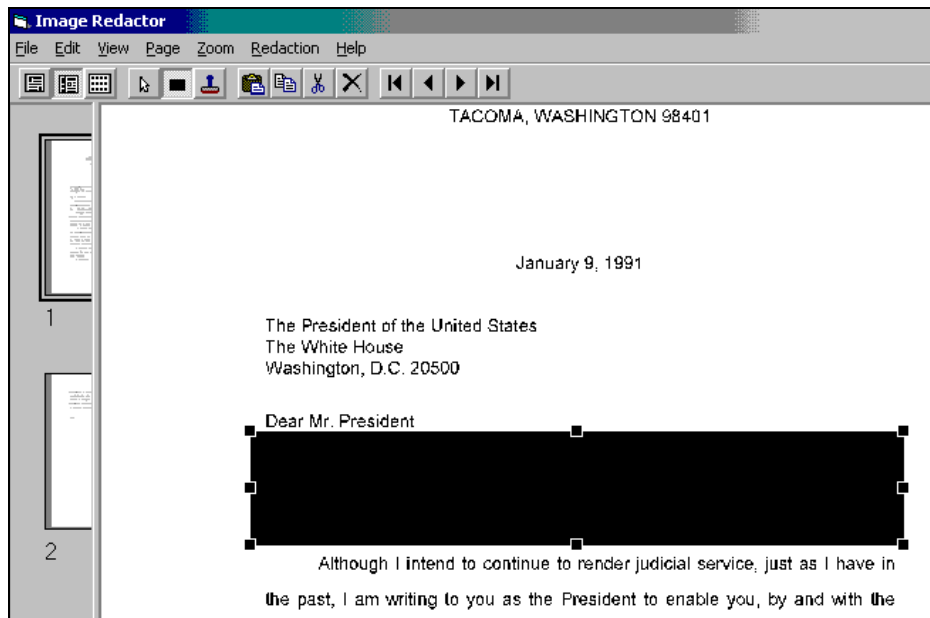
The contents of the highlighted file will be converted to a TIFF (Group 4) multi-page format file and stored in a Windows temporary directory. The file will then automatically be loaded into a customized Imaging window. By default, thumbnails are shown for each page in a multi-page document.



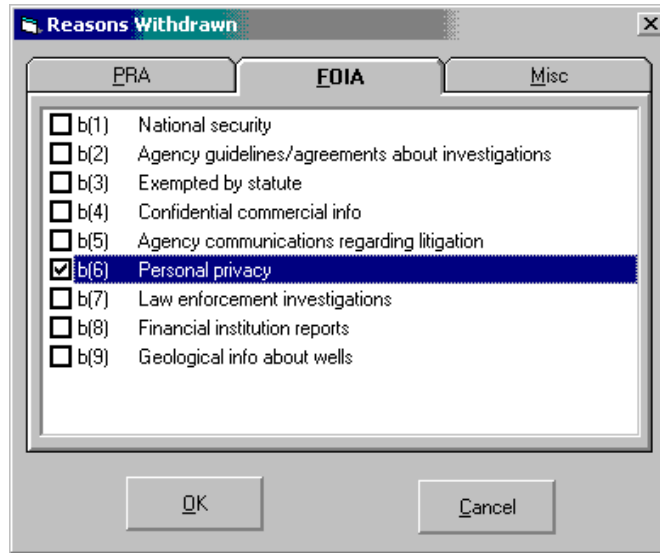
Step 3: Select Zoom from the pull-down menu to size the displayed page. "Actual size" or "Fit to width" seem to be the best zoom option. Selecting "Scale to gray" from the View pull down menu sometimes gives a sharper image for black and white text.



Step 4: To redact text, select "block rectangle" from the Redaction pull-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.

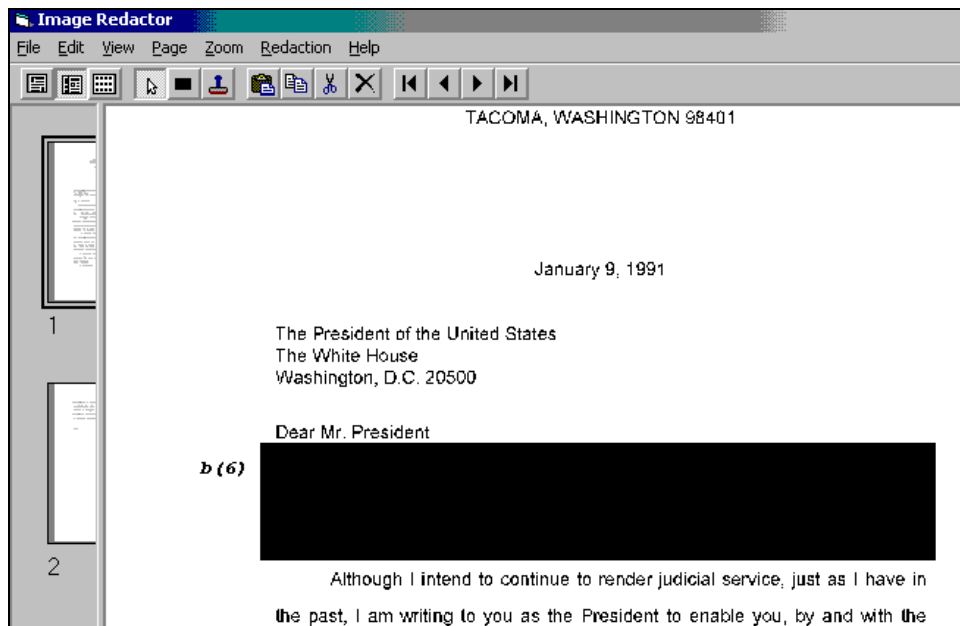


Step 5: The reason(s) for redaction is indicated with a rubber stamp. Stamps can be selected from the Redaction pull down menu or by selecting the “rubber stamp” icon. The paragraphs of the PRA restriction and FOIA exemptions are shown in a Reasons Withdrawn dialog box.



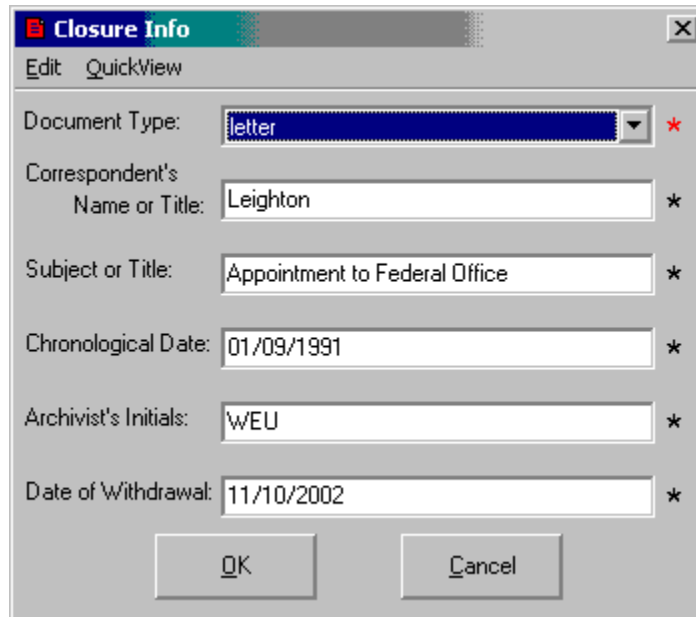
Step 6: Check the restrictions that apply to the redacted text, tables or figures, and select OK.

Step 7: Place the rubber stamp near the redacted text to which they apply (in the margin or in a white area above the redacted text) and click the mouse button. The identifier(s) for the restriction or exemption will be inserted at that point. [Be sure to select the pointer, or the identifier will be inserted at each point you click the 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 pull-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 pull-down menu. A dialog box will be displayed for entering information about the closed original of the redacted document



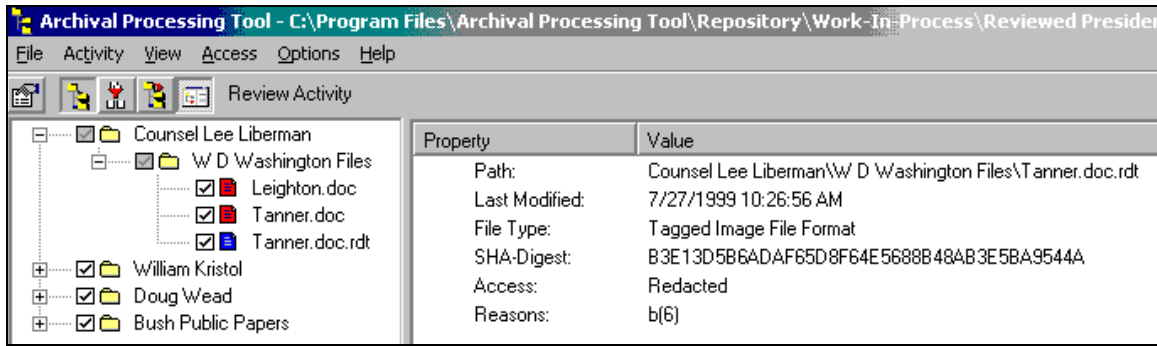
Step 10: Enter the information. You should indicate the Document Type by selecting the downward arrowhead to the right of the Document Type field, and then selecting the document type from that list. If document type of this document is not in the list, cancel the closure information box, and select **Options>Edit Document Type List** and add the needed document type to the list. Then restart the process of closing the document.

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).

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.

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 window pane. If any of this information seems incorrect, it can be edited by highlighting the filename and selecting **View>Properties**



Redacting Text

A second method of redacting text is included in the Review Activity. It supports redacting the text of the document rather than an image of the document. This tool was included primarily to indicate that it was technologically possible to do so. It is recommended that you not use this method of redaction. The tool currently supports only indication of FOIA exemptions, not 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, pull-down menu select **File>Exit Quick View Plus**

Step 4: In the APT window, select **Access>Redact>Text** from the pull-down menu. This opens Word and pastes the text you copied into a document. Note: 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 your left mouse button, place the cursor to the left of where you want to begin highlighting. Important: While highlighting, you must keep the left mouse button depressed. When you reach the end of the text to be redacted, lift your finger. 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.

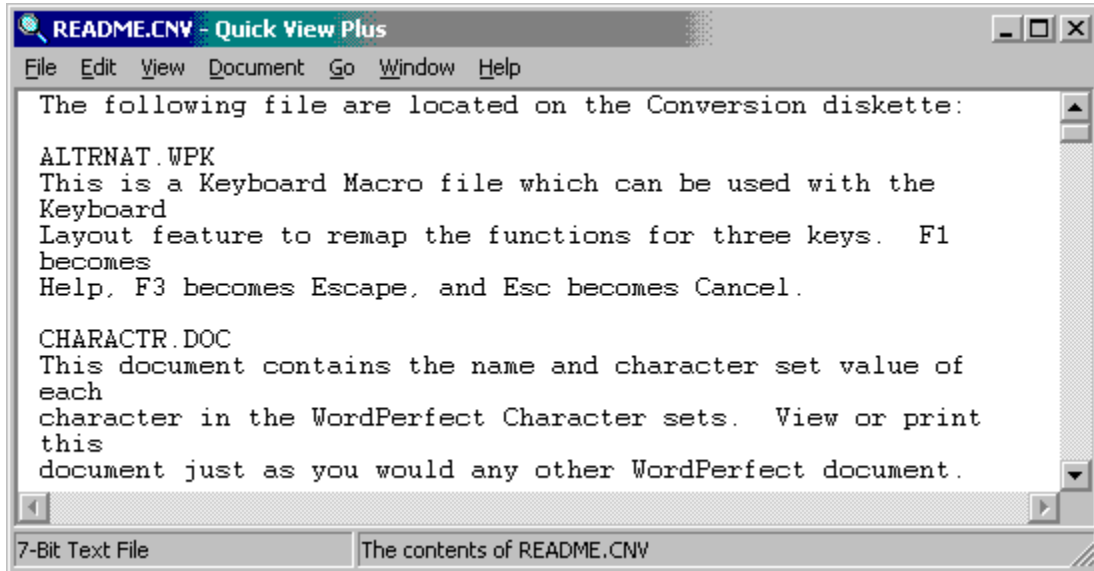
Step 7: To add additional exemptions [such as (b)(3) and (b)(6)], select some of the dots behind the first exemption, and select the next exemption button.

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.

Marking for Transfer

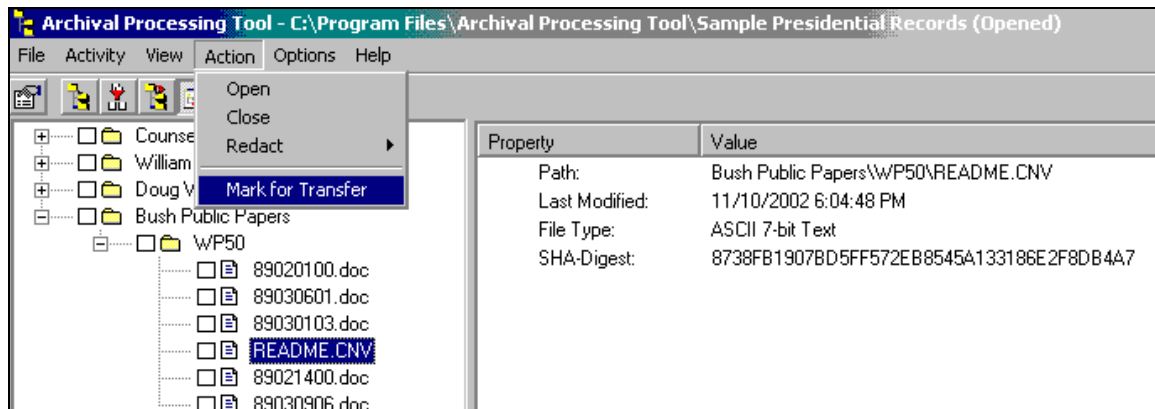
Upon reviewing a file, an archivist might determine that the file needs to be transferred to some other collection. For instance, a copyrighted software application might need to be transferred to the library, or a system or software application file not created by an individual or office might need 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.



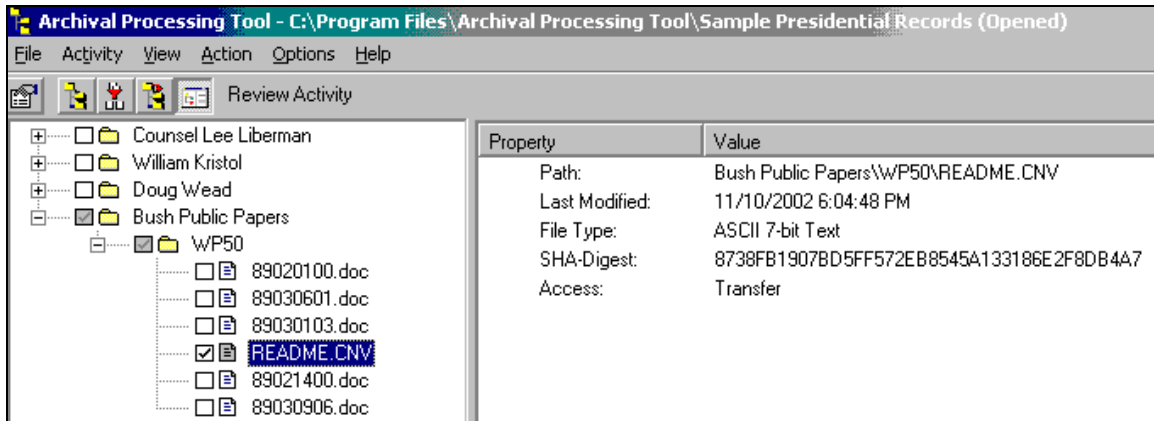
It is a document that is included with a WordPerfect Installation. It was not caught during filtering, and needs to be transferred somewhere else, perhaps to a library of software applications.

To mark a file for transfer

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



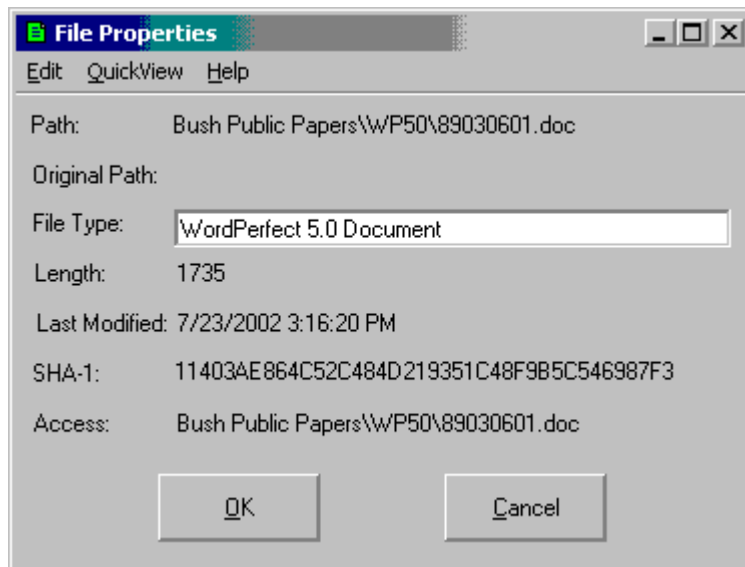
Step 2: From the review menu bar, select **Action > Transfer**.



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.

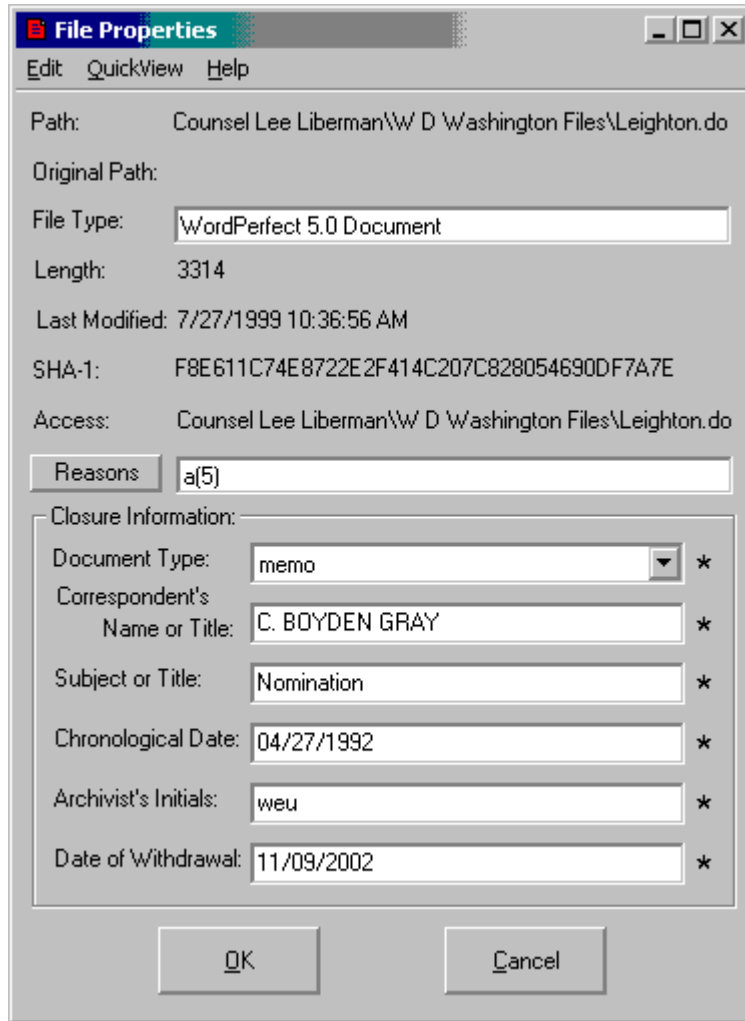
Viewing File Properties

If you highlight a filename and select from the Menu bar **View > Properties**, you will see whether there are access restrictions in the file's properties. 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.

The File properties of a file that has been closed include the reasons for withdrawal and closure information.



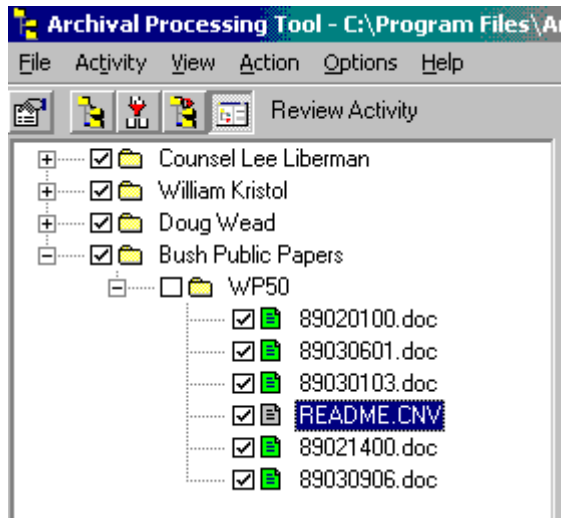
Saving a Reviewed File System

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. In particular, the APT is a prototype archival tool, not a software product. It is likely to continue to undergo revisions, and debugging and software quality is not that of a robust product.

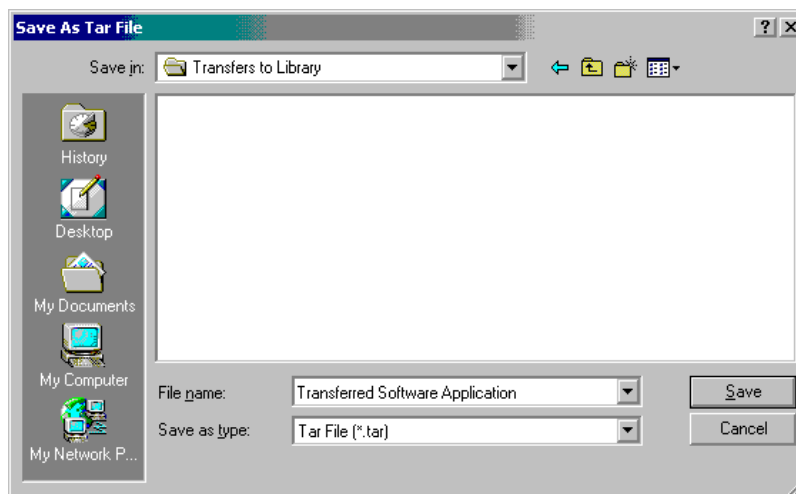
Transferring Files Marked For Transfer

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.

When All Files have been reviewed, if there are any files marked for Transfer, they should be transferred. The file system displayed below has one file Marked for Transfer, README.CNV.

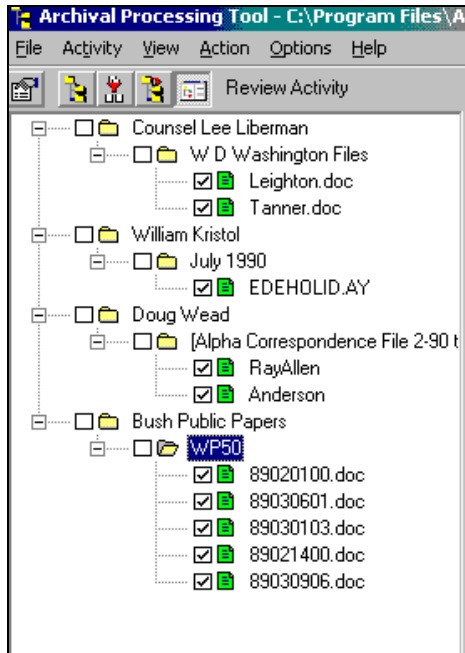


To Transfer the Files select File > Transfer from the review menu bar. A dialog box for saving the files marked for transfer will be displayed.



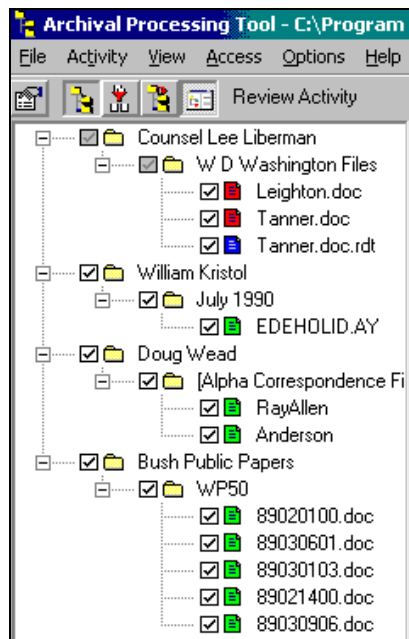
In this dialog box move to a folder that should contain transferred files, in this case Transfers to Library. Enter a name for the TAR file to contain the transferred files and select Save. A TAR file containing those files will be written to the indicated path.

The Review Activity Window will now show the File System with the files marked for transfer removed.



Creating a Reference Copy

The Bush Presidential Library separates closed and opened files in a record series 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. The APT keeps the closed opened and redacted records in the same TAR archive file. The following screen shows the contents of a reviewed file system.



When archival processing (arrangement, preservation, review and description) are complete, you can select "Access Holdings" from the activity pull down menu and create a reference copy of the processed record series that will only contain the opened and redacted copies of files.

Opening a Closed File

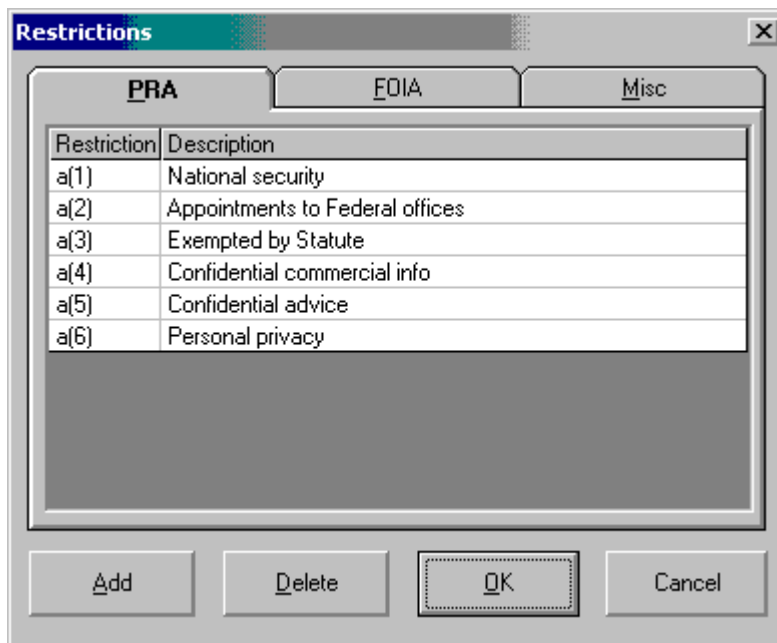
[To be written]

Re-redacting a Closed File

[To be written]

Edit Reasons Withdrawn

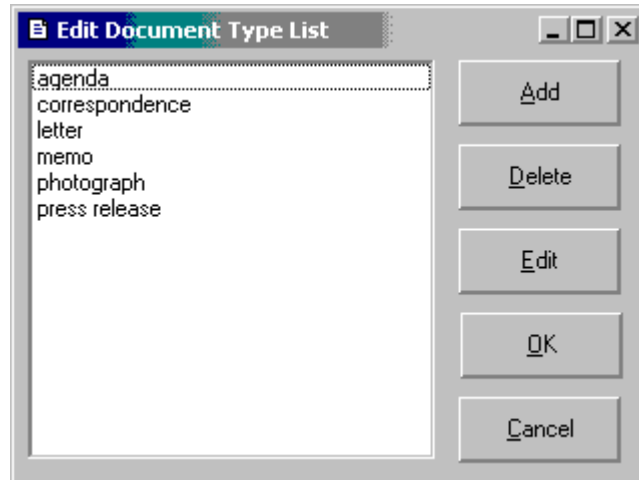
For a variety of reasons, e.g., change in access legislation, archival coding conventions, new donor restrictions, it may be necessary to edit the codes and descriptions of reasons for withdrawal or redaction. The reasons for withdrawal can be edited by selecting Options > Edit Reasons Withdrawn.



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.

Edit Document Type List

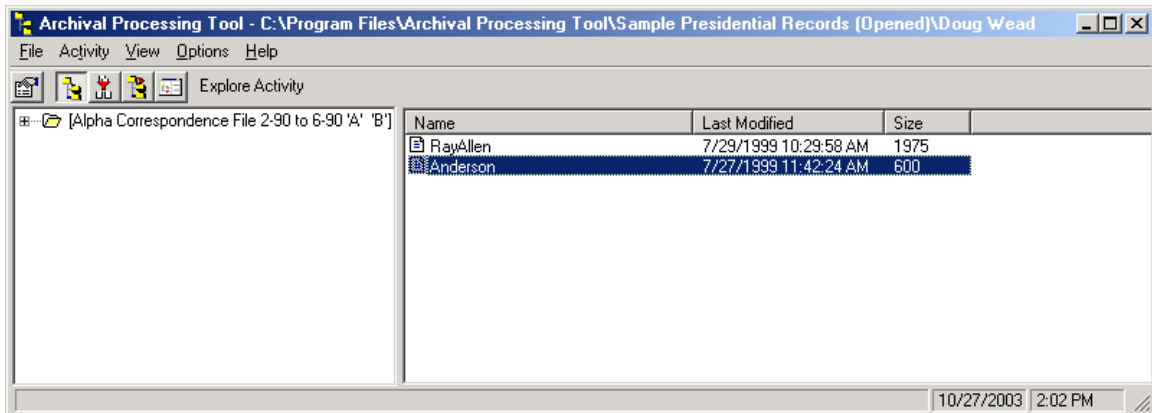
You need to indicate the document type of a closed document. A short list of document types is included with the initial installation. You can see this list, edit it and add to it by selecting **Options>Edit Document Type List**.



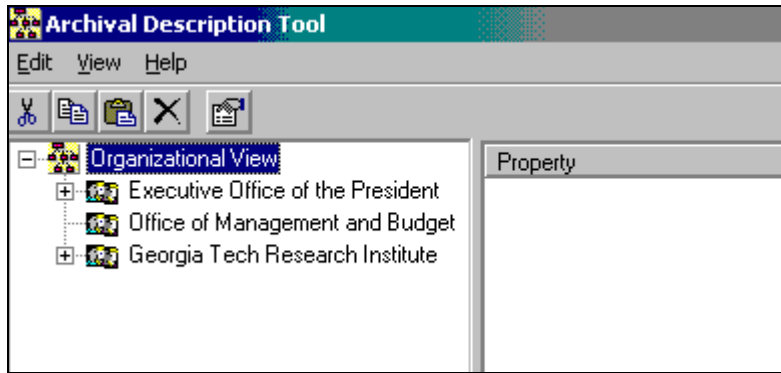
Describing a File System

Description is the archival activity of organizing and recording information that serves to identify, locate, and explain the record series of archives and the context from which those record series were selected.

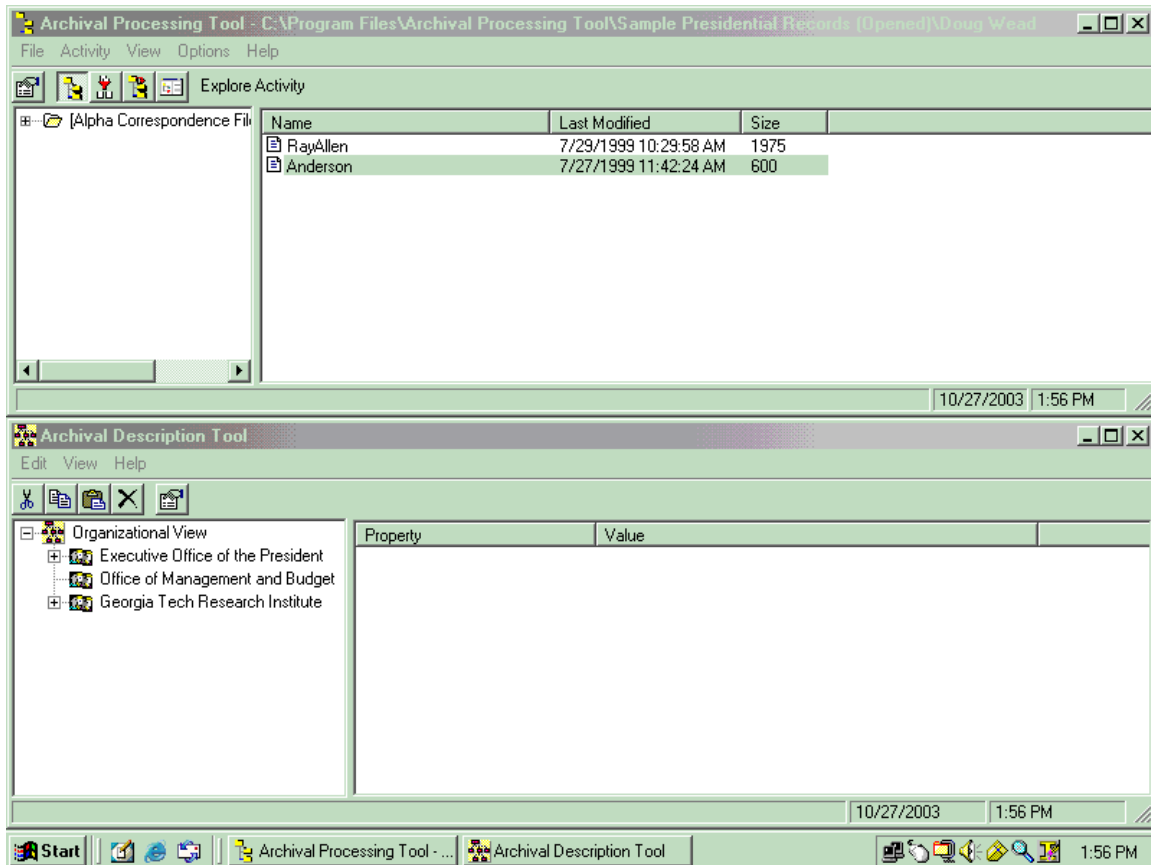
1. Use APT to move container(s) from Archivist's Work Area. Select Description from the Activity pull-down Menu. From the File pull-down menu select open, and find one of the containers you have processed, and open it.



2. Select the Archival Description Tool icon from the Desktop. A window similar to the following will appear.



- 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 proper title and description for the record series. You can then switch to the Archival Description Tool to arrange the record series according to its provenance and to describe the record series. The provenance of a record system is defined by the organization name (and possibly the organizational unit and person/title).

- In the Description window, find the name of the organization that created the records in the file system displayed in the APT window. In this example, there are three organizations—*Executive Office of the President*, *Office of Management and Budget* and *Georgia Tech Research Institute*. Select the organization that created these records. For this example, it is *Georgia Tech Research Institute*.

If the organization is not in the list, highlight *Organizational View* and click the right mouse button. The following dialog box will be displayed. The *Id* entered should be the Record Group number. For instance, the record group for the *Executive Office of the President* is 429, for *Office of Management and Budget* it is 51. Enter the *Name* of the Organization, and the *Start* and *End* date for the organization. For the George H. W. Bush Presidential Administration, we entered the first and last date of the Administration. Select *Save* to retain this organizational information.

The 'Organization' dialog box contains the following fields and controls:

- Id:** A text box containing the value '0' followed by an asterisk (*).
- Name:** An empty text box followed by a red asterisk (*).
- Start Date:** An empty date input field.
- End Date:** An empty date input field.
- Buttons:** 'Save' and 'Cancel' buttons at the bottom.

- To see the organizational units (offices) in an organization, highlight the organization name and click the left mouse button. For instance, for *Executive Office of the President*, a window similar to the following would be displayed.

The 'Archival Description Tool' window displays the following information:

Organizational View (Tree Structure):

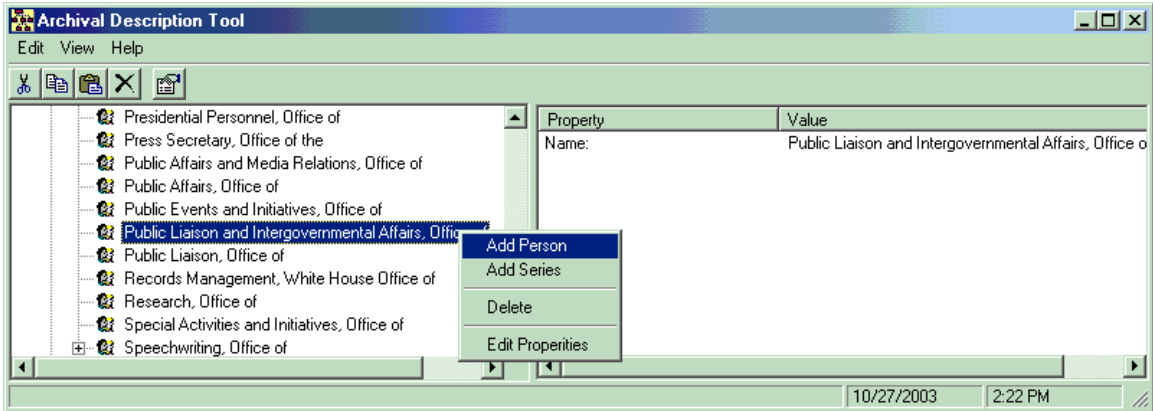
- Executive Office of the President
 - Administration, Office of
 - Administrative Office, White House
 - Advance and Special Initiatives, Office of
 - Agricultural Trade and Food Assistance, Office of
 - Appointments and Scheduling, Office of
 - Cabinet Affairs, Office of
 - Cabinet Secretary, Office of the
 - Chief of Staff to the President, Office of the
 - Communications, Office of
 - Correspondence Office, White House
 - Counsel to the President, Office of the
 - Domestic Policy, Office of
 - Economic and Domestic Policy, Office of
 - Executive Clerk, White House Office of the
 - Executive Residence, White House Office of the
 - First Lady, Office of the
 - Intergovernmental Affairs, Office of
 - Issues Analysis, Office of

Property Value Table:

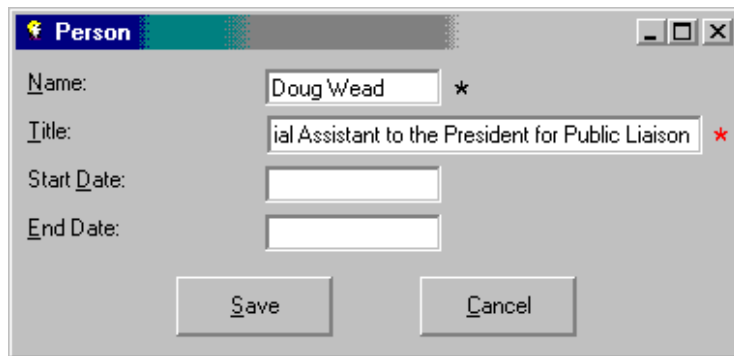
Property	Value
Id:	429
Name:	Executive Office of the President
Start Date:	1/20/1989
End Date:	1/19/1993

The status bar at the bottom shows the date 10/27/2003 and time 12:12 PM.

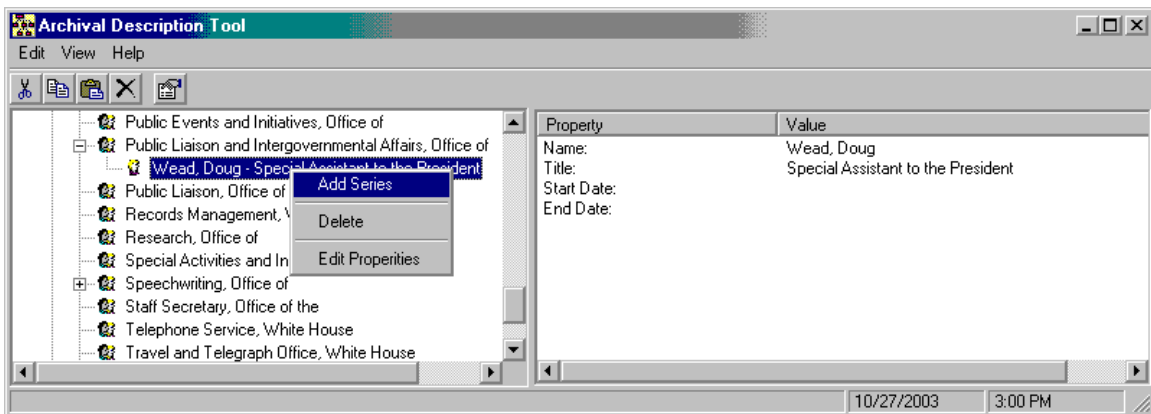
- Highlight the organizational unit (office) name and click the left mouse button. You may see a list of staff members within the office. If not you can add a staff member's name and job title by highlighting the office name and clicking the right mouse button. A dialog box will be displayed in which one can enter the name of the person and their title. If their tenure in the office is known, those dates can also be entered.



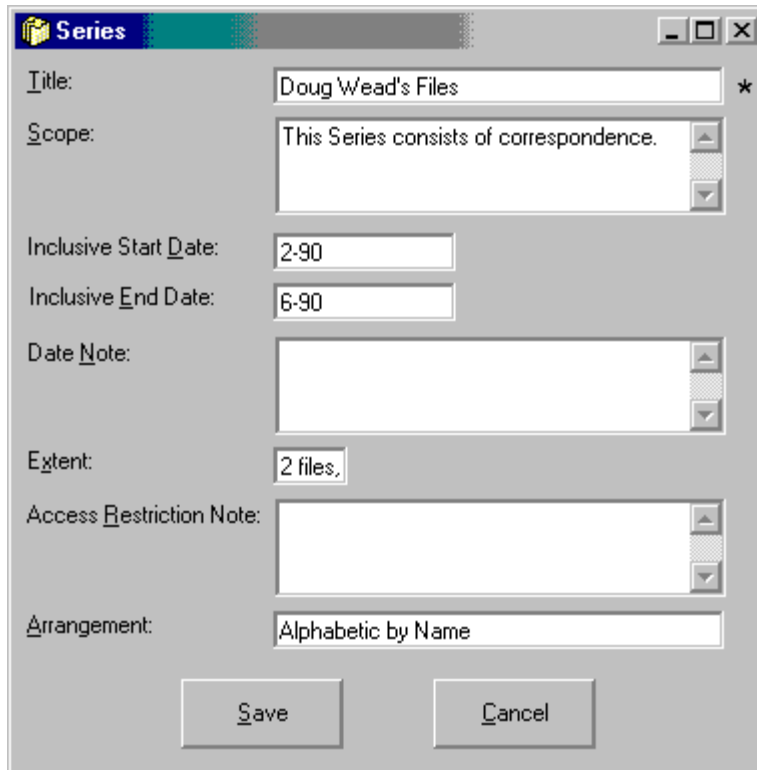
- For instance for the loaded file system, the name *Doug Wead, Special Assistant to the President for Public Liaison* is entered.



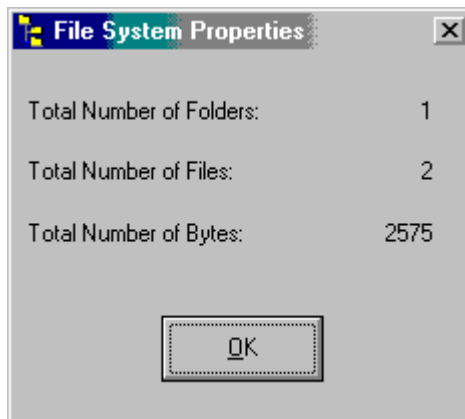
- To associate a record series with an organization, an office, or a staff member, highlight the relevant name, right click the mouse button and select *Add Series*.



- A dialog box similar to the following will be displayed. Enter the Series Title and Scope

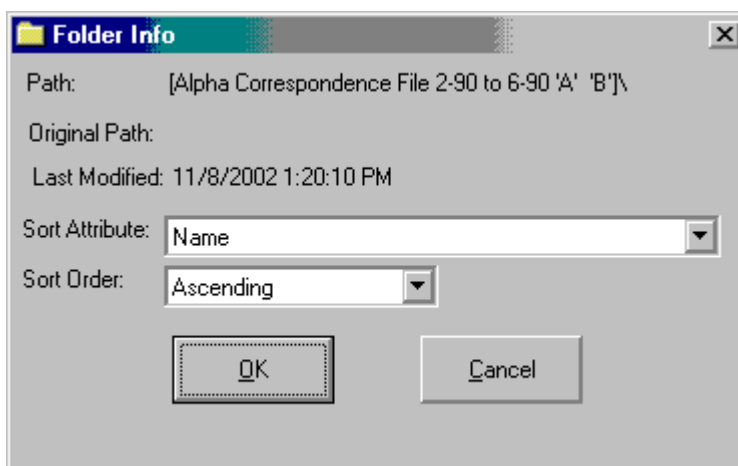


10. To determine the *Inclusive Dates* of files in the file system and the *Extent* of electronic records in the Record Series, select the System Properties option from the View pull-down menu in the Archival Processing Tool. You will see a display similar to the following.



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 in this file is 2575, 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 should enter 2 KB. One can also enter the number of files.

11. To determine the arrangement of files in a folder, click on the folder icon, and select Properties from the Edit pull-down menu. You will see a dialog box similar to the following. It indicates that the folder labeled "Alpha Correspondence" is arranged in ascending order of file name. It could have been in descending order. It could also be in ascending or descending order of file date. For arrangement, we enter *alphabetic by filename*.



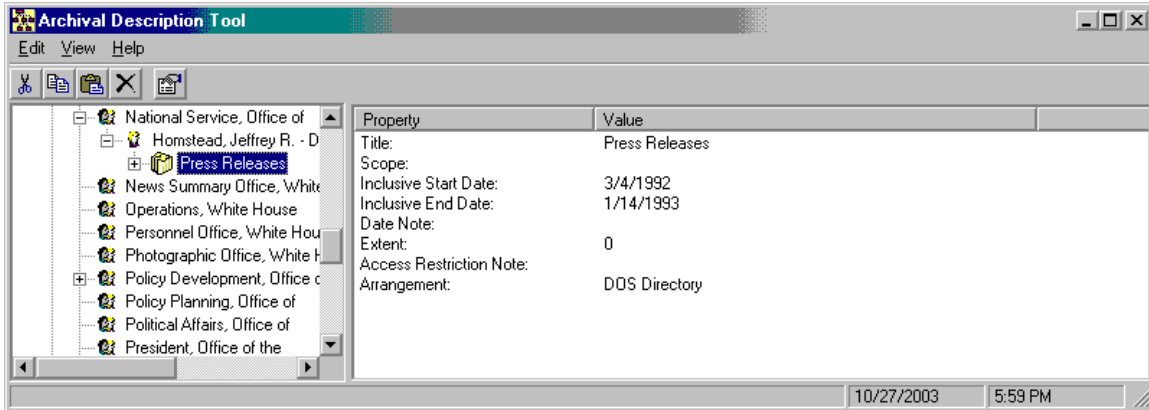
12. If one highlights a series name, one can add file unit titles.
13. To associate container id's with record series, you highlight the name of record series. Associated with the file system loaded in the APT. When the right mouse button is clicked a pop up menu gives the option to "Assoc with container." Select this option and enter the name of the container, and select save.
14. To store the processed and described file system in the Archival Holdings of processed record series, from the APT file pull-down menu you select Store in Holdings. The file system will be stored into a pre-designated storage area for processed record series. You will be prompted as to whether you want to delete the container from your work area.
15. You should load WIP and indicate that the File System has been described. You should also load the accession register, and in the container table indicate that the status of processing the container is complete, i.e., enter a "C"
16. You can then exit the APT and the ADT.

Access Archival Holdings

Archival Holdings are containers of processed record series. They include copies of files open to public access, redacted copies, originals of redacted copies and closed files. Archivists need to access them in order to create reference copies of record series that include only those records of the record series that are open to public access. They also need to access them to reprocess them due to expiration of FOIA exemptions and PRA restrictions. Finally, they need to access them in order to perform preservation operations such as conversion to current file formats.

Accessing Record Series in Archival Holdings

1. By selecting the option *Access Holdings* from the APT pull-down menu, you can access the archival holdings of processed record series. The following window, similar to that of the Description Activity will be displayed.



2. Browse the organizations, organizational units (and possibly Person's Names) until you find the relevant record series. Highlight the name of the series. Select open from the File pull-down menu. The file system structure of the record series will be displayed.
3. To display a different record series, select close from the File pull-down menu, and repeat the previous step.

Copy Record Series to Work Area for Preservation or Review

If preservation operations are needed on the record series or closed or redacted documents need to be reviewed due to expiration of PRA restrictions or FOIA exceptions, you make a copy of the record series.

With the record series is opened as described in the previous section, select Copy to Work Area from the File Pull Down menu. The following dialog box will appear in which you indicate the archivist's folder in which to save the record series container.

Move Record Series from Work Area Back to Archival Holdings

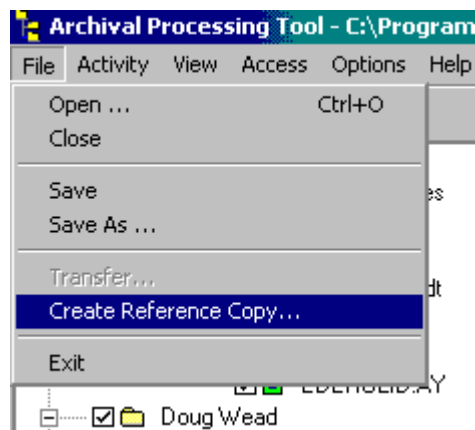
When preservation and/or review activities are complete, the processed record series can be moved back to archival storage by having the file system loaded in the APT and selecting Access Holdings. Then select Replace Record Series from the File Pull Down menu. The record series will be replaced. If the container which is being replaced is not in the archival holdings, you will receive the error message that "The container is not in Archival Holdings."

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 Archival Holdings will not only contain open records, but closed records, original copies of redacted records, and redacted copies. You must create a Reference copy of record series for public access that does not include closed records, or originals of redacted records.

1. From the File Menu of the APT, select "Access Holdings" from the activity pull down menu.
2. Browse the tree of organizations, organizational units, person's names and series until you find the record series for which a reference copy is needed. Highlight the name of the record series and select open from the file pull-down menu. The file system structure of the processed record series will be displayed
3. A reference copy of the file system that includes only the opened and redacted copies can be created by selecting **File > Create Reference Copy**.



Installing the APT and Configuring Resources

This section of the User's Guide also appears as a Readme.txt file in the APT.zip file.

Windows 2000

When The Windows 2000 operating system is loaded, the disk should be formatted as FAT32, not NTFS.

In Explorer or My Computer, in **Tools > Folder Options** and under the Tab View, be sure that

"Hide File extensions for known file types" is not checked.

"Display the Full Path in Address Bar" is checked.

Microsoft Office

The APT uses the Access Database Management System included in Microsoft Office. Hence, Microsoft Office should be installed for the APT to properly operate.

Antivirus Software

NARA uses the McAfee Virus scanner. It needs to be configured for on-access scanning, that is, the scanner looks for viruses in files that you open, copy, save or otherwise modify and files that you read from or write to floppy disks, CD-ROMS or network drives.

WinZip and Other Archivers

Step 1: Install WinZip

Step 2: From the PERPOS portal, copy the files PKUNPAK.exe, ARJ271.exe, and laha213.exe.

Step 3: In C:\Program Files, create a new folder named Archivers. In his folder, create three new folders named ARC, ARJ and LHA.

Step 4: Copy PKUNPAK.exe into ARC directory.

Step 5: Copy ARJ271.exe into ARJ directory. ARJ271.exe is a self-extracting archive. Select this program in windows Explorer. It will ask if it's OK to extract into C:\ARJ. Say yes. After it has extracted the programs into c:\ARJ, copy them into c:\Program Files\archives, and delete the C:\ARJ directory.

Step 6: Copy LHA213.exe into the LHA directory. LHA213.exe is a self-extracting archive.

Step 7: In WinZip, select Options > Configuration and select the Program Locations tab.

Step 8: For ARJ, browse to find c:\Program Files\archivers\ARJ\arj.exe

Step 9: For lha, browse to find c:\Program Files\archivers\ARJ\lha.exe

Step 10: For ARC, browse to find c:\Program Files\archivers\ARJ\PKUNPAK.EXE

Step 11: For the location of the Virus Scanner, browse to the location

Step 12: For the location of the viewer, browse to c:\program Files\Quick View Plus\Program\qvp32.exe

Installing the APT

The Archival Processing Tool is provided in a zip file. The readme.txt file explains the contents of the zip file and the installation.

Step 1: Open the APT.zip in WinZip.

Step 2: Select Install from the WinZip menu bar. (Alternatively, Extract the zip archive into a temporary directory and select setup.)

During the install process, the following message is likely to appear:

Version Conflict

A file being copied is not newer than the file currently on your system. It is recommended that you keep your existing file.

Filename 'c:\WINNT\system32\imgthumb.ocx'

Description 'Thumbnail control'

Your Version 5.02134.1

Do you want to keep this file

Yes No No to All

You should answer Yes.

Step 3: Create a Desktop icon for APT by right clicking the mouse anywhere o the desktop. Select New > Shortcut. For the filed for the Location of the Item, Browse to find c:\program Files\Archival Processing Tool\APT.EXE. Select OK. Select a title for the Program. Replace APT.EXE with Archival Processing Tool. Select Finish.

Storage Management

Record series are stored/accessed from different storage devices or different areas of the same storage device depending on their stage of processing. Similarly, decisions must be made as to where databases such as the Accession Register and Archival Catalog will be stored.

It is assumed that the APT, ADT, Accession Register, and Archival Catalog, and WIP are stored on the C: drive.

In Windows Explorer, create five subdirectories on the C-Drive.

- Accessioned
- Work Area
- Holdings
- Reference
- Library

Copy the table "storage management.mdb" into the directory C:\Microsoft Office Tools\Office Highlight that file name and double click on it. You will be in Microsoft office Access and will be the table *storage management*. Double click on it and you will see a data table similar to the following

Metadata and Record Series Storage	Storage Device
Containers Transferred to Presidential Library	A;; D:
Accessioned Containers	C:\Accessioned
Working Storage	C:\Work Area
Containers of Processed Record Series	C:\Holdings
Containers of Files Transferred out of File Systems	C:\Library
Reference Copies	C:\Reference

You may change the Storage Devices and folder titles at this time or at a future date.

Quick View Plus

The APT uses the Quick View Plus (version 7.0) set of Viewers. Follow the instructions for installing Quick View Plus

Password Recovery Software

Access Data

Software to Decrypt Password Encrypted Files

Software to Erase Temporary and Deleted Files

File slack space. In FAT and FAT32 file systems, files are stored in clusters. If a file does not entirely take up a cluster there is slack space in the cluster at the end of the file.

One of Windows largest security problems is it's swapping of memory to a swap or paging file on the hard drive.

In Windows 2000, temporary files are stored in C:\TEMP

The disk cleaning process should overwrite the data to the criteria in the DoD 5220.22-M disk sanitizing standard.

From Explorer, highlight local disk (C:), Right click mouse, select Properties, select Disk Cleanup.

Eraser 5.3 (Free), Secure Erase

Informatik Image Driver

In Windows Explorer:

Step 1: Find gtech_106092.zip in c:\Program Files\Archival Processing Tool.

Step 2: Create directory c:\Image Driver

Step 3: Open that file with WinZip and extract the files into C:\Image Driver

Step 4: Run Installer.exe in Image driver directory. When the installer asks you if you want to make the image driver the default, say no. Otherwise, every time you want to print it will assume that you want to print to the image driver, not to your default printer. If you accidentally say yes, so that the Informatik image driver is the default printer, select Printers from the Control Panel. Select the printer you want to be the default printer. Select File > default printer.

Step 4: If you have a backup of the file gtech_106092.zip on a diskette or CD, you can delete it from the directory c:\Program Files\Archival Processing Tool.

Filters

NSRL Hash Code Filter.ftl

File Type Filter.ftl

OS-App Hash Code Filter.ftl

OS-App and File Type Filter.ftl

Tutorial Files

PC Files Test Suite

Sample Presidential Records

Sample Corrupted Files

Sample Encrypted Files

Sample Files Needing Conversion

File Repair Tools

WPMD 3.0 - The File Doctor for WPerf 5.x, Shareware from Software by Seidman

Wp5md.exe

Doctor 6, v. 2.2 - The File Doctor for WPerf 6.x, Shareware from Software by Seidman

Wp6doct.exe

Recover - DBF recovery, ABRI Technologies, <http://www.abri.com/recover.html>

AZR - Advanced Zip Repair <http://www.repairfile.com/> \$29.95 repairs zip and zip sfx

Licenses

Written as a part of this government sponsored project

Wpcal.vwr

Wpnb.vwr

Freeware

LHA 2.13 Haruyasu Yoshizaki

PKARC

Fcrackzip

Zoo.exe

Pxunsec.exe Password Recovery for Paradox 3.5 and 4.0 Databases

PxDAZZ.exe Password Recovery for more recent Paradox Databases

Winwu.exe

Wfwcd.exe Recovers Word for Windows Passwords

Wpccrack.c Recovers passwords for WordPerfect 5.x files

Wpuncryp.c Given the correct password, decrypts WordPerfect 5.x files

Shareware

ARJ.exe Robert Jung

ARC.EXE System Enhancement Associates, Inc.

COTS

QuickView Plus

DbfView

WinZip32

Operating System Software

DOS 5

Help.com

Windows 3.1

Pifedit.exe

Calendar.exe

Windows

WinHelp.exe

Windows Media Player wmplayer.exe

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) To copy files to a storage medium for purpose of backup.

(n) A disk, tape, or file that contains files that have been backed up.

(adj.) A file attribute that marks files that have been modified since the last backup.

closed file

a file unit or series to which access is restricted or denied.

compressed file

a file in which redundant information has been eliminated and thus is shorter and requires less storage space than the original file.

file blocking filter

A file filter in which files that match the pattern are blocked from passing through the filter. Those files that don't match the filter are said to be passed.

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 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.

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 that maps a bit-string of arbitrary finite length to a bit-string of fixed length referred to as a hash-code, hash-value or simply hash.

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.

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.

self-extracting archive

An executable program file that includes both an archive file and software to extract or "unarchive" the contents of the archive file. Users can extract the contents of a self-extracting archive file by simply running it.

Appendix A: APT Error Messages

Error Message	Description	Corrective Action
<p>APT Run-time Error 429 Active X Component Can't Create Object</p>	<p>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 VBATT.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.</p>	<p>It is necessary to unregister the DLL and then reregister it.</p> <p>Select 'Start' and then 'Run'.</p> <p>Enter:</p> <pre>Regsvr32 -u "C:\Program Files\Archival Processing Tool\VBATT.dll"</pre> <p>and select OK.</p> <p>You should see the message:</p> <pre>DllUnregisterServer in C:\Program Files\Archival Processing Tool\VBATT.dll succeeded</pre> <p>Remove the -u parameter from Regsvr32 command but leaving the same quoted path to the DLL, and select OK.</p> <p>You should see the message:</p> <pre>DllregisterServer in C:\Program Files\Archival Processing Tool\VBATT.dll succeeded</pre> <p>Retry selecting the Archival Processing Tool.</p>
<p>Archival Processing Tool\APTT is not accessible Access is denied</p>	<p>This message may occur after exiting the APT.</p> <p>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.</p>	<p>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.</p> <p>Alternatively, ignore the error message. The APT erases the files when it reenters when Quick View Plus no longer has control of the files.</p>
<p>"Bad Argument" APT aborts</p>	<p>Error in File Type Identifier</p>	<p>Make a copy of the file system that was being filtered. Break the file system</p>

		into two roughly equal parts and run the file type Filter on each part. Typically, the error will occur in one of the parts. Break the part that has the error into two parts and run the file type filter on each part, etc. until you find the file that is causing the error. Zip the file and email it to the PERPOS Research Team who will eliminate the bug in the file identifier.
"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 APT.zip file contains a sample file type filter that indicates file types that are operating system or software application files. These are file types that are not created by users of the operating system and software applications.

Ami Professional Macro
Ami Professional Template
DOS Batch Command Text
DOS Code Page Information
DOS Device Driver
DOS Hardware Configuration Text
DOS Help
DOS System Data Country
DOS System Data Keyboard
Empty File
OS/2 16-bit Dynamic Link Library
OS/2 Virtual Device Driver
Unix Shell Script Text
Windows 16-bit Dynamic Link Library
Windows 3.1x Installation Information Text
Windows 3.x Help
Windows 32-bit Dynamic Link Library
Windows 95 Installation Information Text
Windows Collect of Virtual Device Drivers
Windows Cursor Resource
Windows Icon Resource
Windows NT Installation Information Text
Windows Program Information File
Windows Program Initialization File Text
Windows Program Manager Group File
Windows Virtual Device Driver
WordPerfect Application Resource Library
WordPerfect Block
WordPerfect Column Block
WordPerfect Device Driver
WordPerfect Dictionary
WordPerfect Display Resource
WordPerfect Equation Resource
WordPerfect External Dictionary
WordPerfect Help
WordPerfect Hyphenation Code
WordPerfect Hyphenation Data
WordPerfect Hyphenation Lex
WordPerfect Keyboard Definition
WordPerfect Macro
WordPerfect Macro Resource
WordPerfect Mouse Driver
WordPerfect Overlay
WordPerfect Printer Resource ALL
WordPerfect Printer Resource PRS
WordPerfect Rectangular Block

WordPerfect Setup
WordPerfect Spell Code Rules
WordPerfect Spell Code Word List
WordPerfect Thesaurus
WordPerfect Vax Keyboard Definition
WordPerfect Video Resource

Needed Modifications to APT User Manual

1. Need to discuss files that cannot be automatically filtered out e.g., FRECOVER, Batch files
2. In section "Filter on File Type," categories of file types are discussed. In that section Appendix B should be references. Three File types of Operating system and application software are listed. Perhaps in Appendix B we need to discuss why each type is not a user-created file. Reference or give examples from PC Files Test Suite.
3. Explain the Authority for the files included in the OS-APP Hash Code Filter
4. A user manual is not going to be enough. We will need a separate on-line tutorial.
5. Explain how to create subaccounts (logon names) in addition to the administrator, and how this can automatically be inserted into the withdrawal information.
6. Consider adding a PDF redactor.
7. Explain the difference of redacted copies and the original copy. The Best QVplus settings for a redacted document (Image) may be size = fit to Window width
8. Need to install the program to completely absolutely erase deleted files.