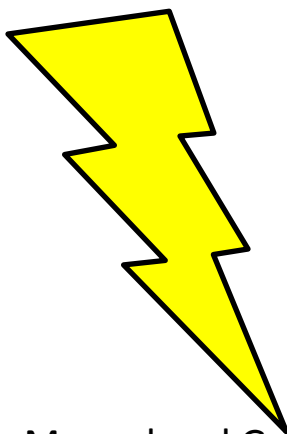


BENTLEY  
HISTORICAL  
LIBRARY

## PROCESSING HYBRID AND DIGITAL COLLECTIONS



An AutoPro User Manual and General Reference

Version 3.0

October 29, 2013

Prepared by Michael Shallcross, Division of Digital Curation

1. Introduction .....	1
Institutional Context .....	1
Digital Processing as a Concept and Approach at the Bentley Library .....	1
Overview of the Automated Processing Workflow .....	1
Notes on the Windows Command Prompt.....	4
2. Initiate a Processing Session .....	6
Launching AutoPro.....	6
Deposit Confirmation.....	7
Processor Identification .....	7
Identification of Archived Websites.....	8
Final Steps .....	8
3. Main Menu and Processing Options .....	9
4. Guide to Individual Procedures.....	11
Step 1: Virus Scan.....	11
Step 2: File Extraction .....	11
Disk Image Extraction with FTK Imager .....	12
Archive File Extraction with 7-Zip .....	14
Step 3: Personally Identifiable Information (PII) Scan .....	15
Step 4: Identify Missing File Extensions .....	17
Step 5: Format Conversion.....	19
Step 6: Appraisal, Arrangement, and Packaging for Deposit.....	20
Characterize Deposit: Understanding the Extent and Diversity of Content .....	21
Review Directory Hierarchy .....	22
View Relative Size of Directories.....	22
Compare File Extensions within Deposit .....	22
Compare Age of Files .....	23
Review Duplicate Content.....	23
Manually Review Content.....	24
Quick View Plus:.....	24
IrfanView:.....	25
Inkscape: .....	25
VLC Media Player: .....	26

Separations: Removing Superfluous Content .....	26
Resource #1: AutoPro-separations-onlySelectedItems.bat .....	27
Resource #2: AutoPro-separations-extensionRemoval.bat .....	27
Record Manual Conversions .....	28
Create a List of Resources for an HTML Access Page.....	29
Arrangement.....	30
Packaging Content for Deposit and Defining Deep Blue 'Items' .....	30
Determining How to Package Content .....	31
Example of a Packaging and Deposit Plan .....	32
Packaging Content with <i>AutoPro-zipOneFolder.bat</i> .....	33
Packaging Content with <i>AutoPro-zipMultipleFilesFolders.bat</i> .....	34
Completing the Appraisal and Arrangement Procedures .....	35
Step 7: Add Descriptive and Administrative Metadata .....	35
Add Deposit-Level Metadata .....	35
Check for Pre-Completed Metadata Spreadsheet.....	36
Creating a New Item .....	37
Item-Level Metadata.....	37
Associate Bitstreams with the Item .....	38
Create an Item for Archived Websites.....	39
Reviewing an Item.....	40
Finalize Description and Technical Metadata Extraction.....	41
Step 8: Transfer to Long-Term Storage.....	42
Transfer to Deep Blue Deposit Folder.....	42
Step 9: Clean Up.....	43
5. Version History.....	45

This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>.



# 1. Introduction

## Institutional Context

Established in 1935 by the University of Michigan Regents, the Bentley Historical Library serves as the official archives of the university and documents the history of the state of Michigan and the activities of its people, organizations and voluntary associations. The library has successfully managed and preserved digital content since the 1997 accession of former University President James J. Duderstadt's digital desktop. Given the steep increase in born digital and digitized content accessioned by the library in recent years, archivists have sought more efficient and standardized processing procedures. The Andrew W. Mellon Foundation-funded MeMail Project (2010-2011) provided the library with resources to establish a workflow and corresponding policies for the ingest and processing of archival email, but a similar solution was needed for mixed digital content (i.e. Office documents, PDFs, audio and video files, images, etc.). Archivists in the library's Digital Curation Division have advanced the work of the MeMail Project in developing the AutomatedProcessor (or AutoPro), a series of inter-dependent Windows shell scripts that automate key steps in preparing digital content for long-term preservation and access.

## Digital Processing as a Concept and Approach at the Bentley Library

Archival processing in the digital era requires traditional steps such as appraisal, arrangement, and description in addition to procedures that ensure the authenticity, integrity, and security of content. "Digital processing" therefore corresponds to the "generate AIP" function of the Open Archival Information System (OAIS) Reference Model's Ingest entity. After a Submission Information Package (SIP) has been assigned an accession record, digital processing permits archivists to assume intellectual control, establish the integrity of materials, and perform preservation events (i.e. scans for viruses and personally identifiable information, conversion to preservation formats, recording of descriptive and technical metadata, etc.) that transform the SIP into an Archival Information Package (AIP).

Bentley archivists initially developed a manual workflow with more than 40 discrete steps that required the operation of numerous stand-alone applications and saving tool output to various log files. In addition to being highly labor intensive and introducing numerous opportunities for operator error, this approach was daunting for staff without technical expertise. Given these challenges, the Division of Digital Curation developed AutoPro to:

- (1) Make digital processing more efficient by automating key workflow steps.
- (2) Reduce technical barriers and thereby permit archivists to focus their energies on the traditional archival functions of appraisal, arrangement, and description.

## Overview of the Automated Processing Workflow

AutoPro is comprised of 24 Windows CMD.EXE and Visual Basic scripts that move content through a nine step workflow and thereby simplify the operation of more than 20 applications and command line utilities. The Windows Command Prompt and Explorer windows function as the main interfaces, a feature that may be unique to staff more familiar with Graphical User Interfaces (see section 1.4 for tips

on using the CMD.EXE console). In addition to providing a framework to guide archivists through the workflow, AutoPro tracks the current processing status, generates log files for all operations, and records PREMIS preservation metadata that will be stored alongside the processed content in a preservation environment. Archivists must approve the successful completion of each step and may stop at any point in the workflow and resume their work at a later time.

After content has been accessioned and deposited in the Bentley Library's secure interim repository, a processing archivist receives a local copy and then starts AutoPro to add basic collection-level metadata and run a virus scan (the University of Michigan employs Microsoft Forefront Endpoint Protection on all work stations).<sup>1</sup> AutoPro next searches the deposit for disk image and archive files (.ISO, .AFF, .ZIP, .TAR, .RAR, etc.); if any are found, a script employs 7-Zip to extract the contents to a directory named after the archive file, with the original file paths preserved.<sup>2</sup> After verifying the extraction's success, AutoPro moves the archive file to a separations directory and records the operations in a log file. The newly extracted content is then searched for additional archive files, from which the contents are extracted, if necessary. Some disk image formats (.AFF, .AD1, .E01, etc.) will require the processor to employ FTK Imager to manually extract content.<sup>3</sup>

AutoPro then runs DROID to search for files with missing or mismatched extensions and the archivist may then append correct file extensions using information generated by the TrID File Identifier utility and collected from the PRONOM format registry.<sup>4</sup> If the archivist determines that an extension should be added or corrected, AutoPro will document the action in a log file.

In transforming the SIP to an AIP, the Bentley Library relies upon file format conversion as a primary preservation strategy. Based upon the Library of Congress's work on the "Sustainability of Digital Formats" and documentation from the Florida Center for Library Automation and other peer institutions, the library has identified a number of at-risk (i.e. proprietary or potentially obsolete) file formats and developed conversion pathways to sustainable formats with various open source and freeware tools.<sup>5</sup> AutoPro searches for these at risk formats (based upon extension) and then employs the following tools (with digital media and target format in parentheses): ImageMagick (raster images to .TIFF), Ghostscript (.PS, .EPS and .PDF to .PDF/A; JHOVE verifies if the original PDF meets PDF/A specifications), Inkscape (vector images to .SVG), ffmpeg (audio to .WAV; video to MP4 with H.264

---

<sup>1</sup> See UM ITS FAQ pages at <http://safecomputing.umich.edu/antivirus/faq.php>. Microsoft antivirus information may be found at: <http://www.microsoft.com/en-us/server-cloud/system-center/endpoint-protection-2012.aspx>.

<sup>2</sup> 7-Zip is an open source file archiving application. For more information see <http://www.7-zip.org/>.

<sup>3</sup> FTK Imager is digital forensics software produced and freely distributed by AccessData. For more information, see the FTK Imager manual release notes and manual at <http://www.accessdata.com/support/product-downloads>.

<sup>4</sup> DROID is a file identification tool developed by the National Archives (U.K.). For more information, see <http://droid.sourceforge.net/>. TrID is a freely distributed utility that identifies file types based upon a library of over 4,800 binary signatures. For more information, see <http://mark0.net/soft-trid-e.html>. PRONOM is an "on-line information system about data file formats." For more information, see <http://www.nationalarchives.gov.uk/help/PRONOM/faq.htm>.

<sup>5</sup> For more information on the Library of Congress's "Sustainability of Digital Formats" and the FCLA's format recommendations, see <http://www.digitalpreservation.gov/formats/index.shtml> and [http://fclaweb.fcla.edu/fda\\_format\\_landing\\_page](http://fclaweb.fcla.edu/fda_format_landing_page), respectively. For an overview of sustainable formats and conversion strategies at the Bentley Historical Library, see <http://deepblue.lib.umich.edu/handle/2027.42/93307>.

encoding), Aid4Mail (various email formats to .MBOX), and Microsoft Office File Converter (Office files to Open Office XML).<sup>6</sup> These preservation versions are stored alongside the original and denoted by a suffix consisting of ‘\_bhl-’ and (where possible) the CRC32 hash of the original file (i.e. oralHistoryProject\_bhl-0fbc2cc7.wav). AutoPro also creates a log of all file conversions, including the original and new filenames, timestamp, and conversion software.

In order to protect the identities of record creators and limit its exposure to risk, the Bentley Historical Library has established policies in regard to personally identifiable information (PII) such as credit card numbers and U.S. Social Security numbers. AutoPro thus employs Identity Finder DLP Endpoint to scan for PII.<sup>7</sup> Archivists then use the Identity Finder interface to verify search results and—if true positive hits are found—redact the PII (from Open Office XML and plain text files) or assign appropriate access restrictions to the content. A record of identity matches and corresponding archival intervention is maintained with the log files.

Archivists then proceed to a more in-depth appraisal and arrangement of content. AutoPro loads data visualizations (such as the distribution of file extensions, date range of content, relative size of directories, etc.) produced by TreeSize Professional to better characterize and launches Quick View Plus (a file viewing program) to rapidly review a wide range of file types for description in finding aids.<sup>8</sup> While reviewing content with Quick View Plus or the Windows Explorer, archivists use a batch file in the right-click context menu to remove superfluous files or folders to a separations directory. Every effort is made to retain the original order of materials, but archivists may group unorganized content in directories or package content in .ZIP files to simplify the management and storage (with such actions recorded in log files). Archivists also develop a plan as to how content will be deposited in Deep Blue in a manner that is both convenient to end users and in accordance with the intellectual arrangement of material in the finding aid.

Once the arrangement is established, AutoPro calls DROID to extract technical metadata and generate an MD5 checksum for all content (including files in .ZIP archives). Archivists then use the AutoPro interface to apply descriptive and administrative metadata to materials. This step produces a Dublin Core XML file and Excel spreadsheet used to deposit material in Deep Blue, the University of Michigan’s DSpace repository.<sup>9</sup> Finally, AutoPro employs BagIt to transfer a copy of all material (and metadata) to a

---

<sup>6</sup> ImageMagick (<http://www.imagemagick.org/script/index.php>) is an open source raster image editor. Ghostscript (<http://www.ghostscript.com/>) is an open source interpreter for the PostScript language and PDF documents that may be used to convert the latter documents to PDF/A. Inkscape (<http://inkscape.org/>) is an open source vector graphics editor. ffmpeg (<http://ffmpeg.org/>; for Windows builds, <http://ffmpeg.zeranoe.com/builds/>) is freely available software used for audio and video recording and conversion. Aid4Mail (<http://www.aid4mail.com/>) is a proprietary email conversion program. Microsoft File Converter (<http://www.microsoft.com/en-us/download/details.aspx?id=11454>) is part of the freely available Office Migration Planning Manager.

<sup>7</sup> Identity Finder Data Loss Prevention (DLP) Endpoint (<http://www.identityfinder.com/us/Business/IdentityFinder/EnterpriseClient>) is proprietary software that can identify potentially sensitive information.

<sup>8</sup> TreeSize Professional is a proprietary hard disk space and file manager and Quick View Plus is a file viewing utility. For more information, see <http://www.jam-software.com/treesize/> and <https://avantstar.com/>, respectively.

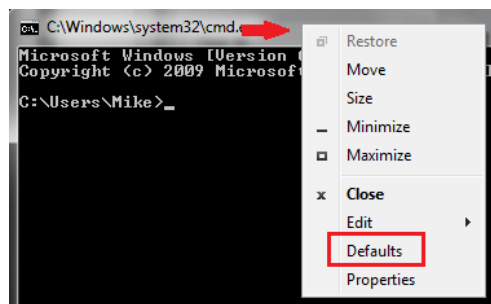
<sup>9</sup> Explore the Bentley Historical Library’s archival community in Deep Blue at: <http://deepblue.lib.umich.edu/handle/2027.42/65133>.

secure dark archives.<sup>10</sup> At the conclusion of processing, AutoPro deletes the working directory, restore point and temporary files and the archivist records the completed digital deposit in the Bentley's collections management database.

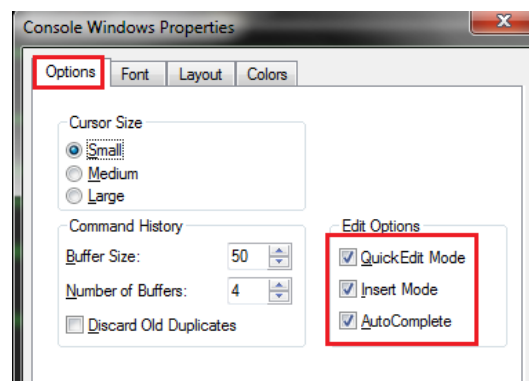
This basic workflow and the component software is subject to change, as the Division of Digital Curation actively tracks the development of standards and professional best practices. It is furthermore recognized that unique features of digital deposits and material may require additional steps to process and record metadata.

## Notes on the Windows Command Prompt

This section of the User's Manual provides additional information on features and functionality of the Windows CMD.EXE console (also referred to as the command prompt). The properties for the CMD.EXE console on your work station should have been configured when AutoPro was installed but you may want to check to make sure that the following options are set. Open a CMD.EXE console window (enter CMD.EXE into the Start Menu's search box), right-click on the border of the window, and select "Defaults" from the context menu.



When the "Console Windows Properties" window opens, make sure that the boxes for the following items are checked under the "Options" tab:



- "QuickEdit Mode" allows you to highlight text with the mouse and then hit "Enter" to copy it to the clipboard

<sup>10</sup> BagIt is part of an open source set of transfer tools developed by the Library of Congress. For more information, see <http://sourceforge.net/projects/loc-xferutils/>.

- “Insert Mode” allows you to paste text from the clipboard by right-clicking where you would like to insert text
- “AutoComplete” allows you to hit the “Tab” key to complete the entry of folder and file names when entering path information

Another great feature is the “Command History” (familiar to Linux/Mac terminal users): use the up and down arrow keys to browse through information previously entered into the CMD.EXE console. This feature will be particularly useful when compiling administrative and descriptive metadata!

Resources:

AutoPro relies on a number of CMD.EXE utilities and Windows batch file syntax to move content through its work flow. If you’d like to learn more about using the CMD.EXE console and batch files in general, the following sites can provide some basic information:

- *An A-Z Index of the Windows CMD Command Line* (<http://ss64.com/nt/>)
- *(DOS) commands and their usage in batch files*  
(<http://www.robvanderwoude.com/batchcommands.php>)
- *Command-line Reference A-Z*  
(<http://www.microsoft.com/resources/documentation/windows/xp/all/proddocs/en-us/ntcmds.mspx?mfr=true>)



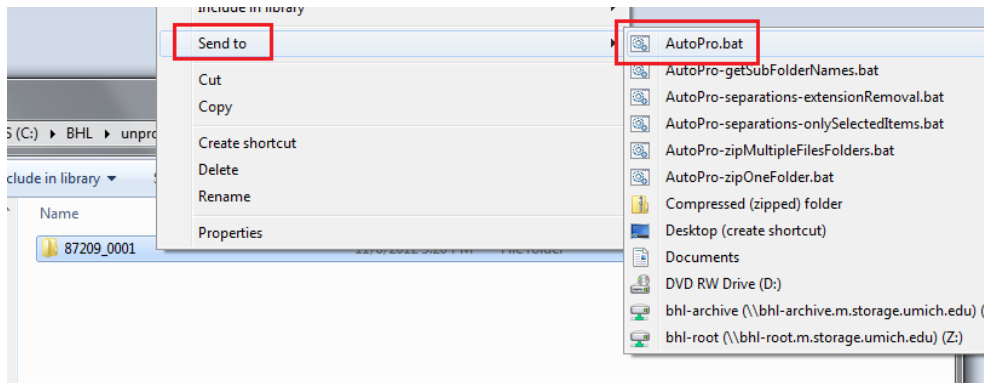
## 2. Initiate a Processing Session

While you may be able to process a deposit of digital content in one session, you may need several sessions over a series of days to completely process and package content so that it is ready for storage in a long-term repository. Follow the steps below to initiate all processing sessions for a given deposit.

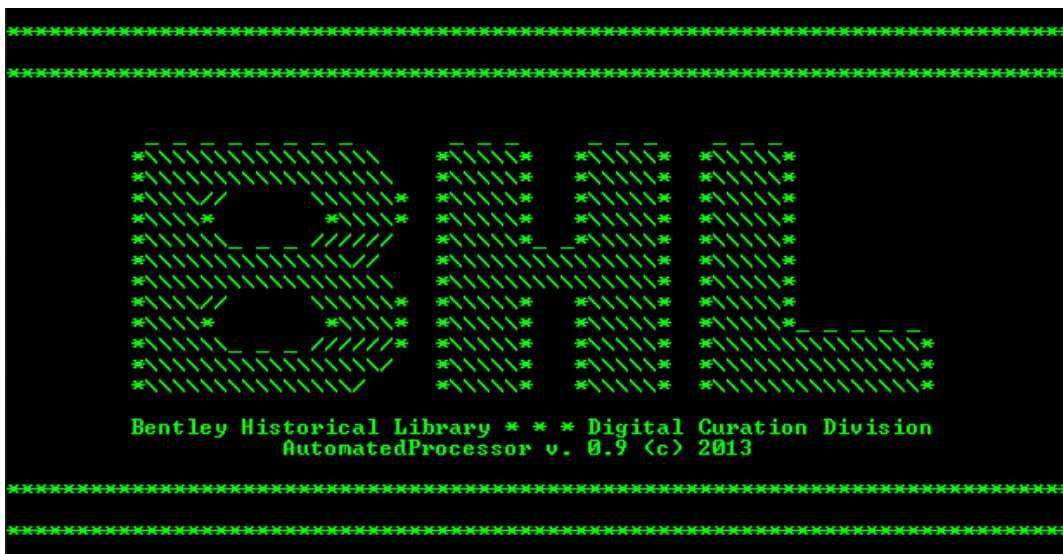
### Launching AutoPro

Go to the 'Unprocessed' directory and locate the appropriate **deposit directory** (i.e., the folder holding the unprocessed material you'll be working on); it will be identified by a unique **deposit ID** (the collection ID plus a four digit suffix).

1. Right-click on the processing directory (i.e. \87209\_0001\).
2. Select *AutoPro.bat* from the "Send to" context menu.



The AutoPro application will now open; hit any key to proceed.



## Deposit Confirmation

You will then be prompted to enter and confirm basic information for the deposit, including the collection ID, creator, and title. The collection ID will be the string of numbers prior to the underscore in the BHL digital deposit number (for example, deposit number 00132\_0002 has a Collection ID of 00132).

```
*****
      Bentley Historical Library * * * Digital Curation Division
      AutomatedProcessor v. 0.9 (c) 2013
*****

CURRENTLY PROCESSING:
--> C:\BHL\unprocessed\00132_0002
--> Creator: University of Michigan. Stephen M. Ross School of Business
--> Collection: Ross School of Business Administration (University of Michigan) Publications
Is the above information correct? (Y / N / Q to QUIT): _
```

If this is a new collection, the ID should have been correctly entered by Mike or Nancy. However, if the deposit represents an addition to an existing collection, check the online finding: the ID will be the first string of digits in the Call Number. If the collection ID is wrong, quit AutoPro, change the deposit directory name, and start the process again (being sure to inform Mike or Nancy). **This simple step can save a lot of time and energy down the line!**

Call number: 00132 Eimu C221 2  
Language: The materials are in English.  
Repository: Bentley Historical Library University of Michigan  
1150 Beal Ave.

You will also identify the creator and collection name; these can be changed later in the process, so don't worry if you're still determining the exact provenance of the collection or its title.

## Processor Identification

AutoPro will record the identity of the processing archivist each time someone logs in to work on the deposit to fully document preservation events and provide a full audit trail of digital processing activities.

```
PROCESSING ARCHIVIST
At the prompt, enter your name (Last name, first name--i.e. Doe, Jane).
Shallcross, Michael
Is your name entered correctly? (Y / N) Y_
```

On subsequent sessions, AutoPro will retrieve the name of the last processing archivist from the main log file and ask you to confirm it (see below). If you are taking over from another processor, respond with *N* and enter your own.

```
*****  
Bentley Historical Library * * * Digital Curation Division  
AutomatedProcessor v. 0.7 (c) 2012  
*****  
  
PROCESSING ARCHIVIST  
Shallcross, Michael  
Is your name entered correctly? (Y / N)
```

### Identification of Archived Websites

AutoPro will then ask if the deposit contains an archived website (Digital Curation will inform you if it does):

```
*****  
Bentley Historical Library * * * Digital Curation Division  
AutomatedProcessor v. 0.9 (c) 2013  
*****  
  
Does this deposit include an archived website? (Y / N) _
```

This only applies to complete or partial websites captured with HTTrack, Teleport Pro, or similar software. This initial identification will ensure that specialized steps are taken later in the processing workflow.

### Final Steps

Once deposit information and your identity have been verified, AutoPro will set a number of variables and, if this is the first processing session for a deposit, create directories for log files, temporary files, a backup restore point, and separations in addition to generating an initial manifest (with MD5 checksums) of the content.

### 3. Main Menu and Processing Options

After verifying the processing directory and processor name, AutoPro will open its main menu. This screen is divided into four parts:

1. Main Menu of Procedures: This section lists all eleven steps in the AutoPro workflow. Steps should be completed in the order provided, but some content types (such as archived websites) may require variations.
2. Statistics: This section provides up-to-date information on the number of files (and size thereof, in bytes) in the processing directory.
3. Processing Status: This section lists all the procedures completed within the processing workflow. Upon first opening the application, it reports that a “new batch processing sequence” has been initiated.
4. Option Entry: This section permits the user to enter the number of the procedure to run next. All newly initiated workflows must begin with “1. Virus scan.” If you are returning to a previously started workflow, you should enter the next procedure (in numerical order) after the most recently completed step, as listed in the Processing Status section. After selecting a procedure, AutoPro will immediately take you to the selected procedure.

```
*****
Bentley Historical Library * * * Digital Curation Division
AutomatedProcessor v. 0.9 (c) 2013
*****

MAIN MENU OF PROCEDURES: 1
1. VIRUS SCAN
2. FILE EXTRACTION
3. IDENTIFY MISSING FILE EXTENSIONS
4. FORMAT CONVERSION
5. PII SCAN
6. APPRAISAL, ARRANGEMENT, and PACKAGING FOR DEPOSIT
7. ADD DESCRIPTIVE and ADMINISTRATIVE METADATA
8. TRANSFER to LONG-TERM STORAGE
9. CLEAN UP

Q: Quit the BHL Automated Batch Processing Sequence
*****

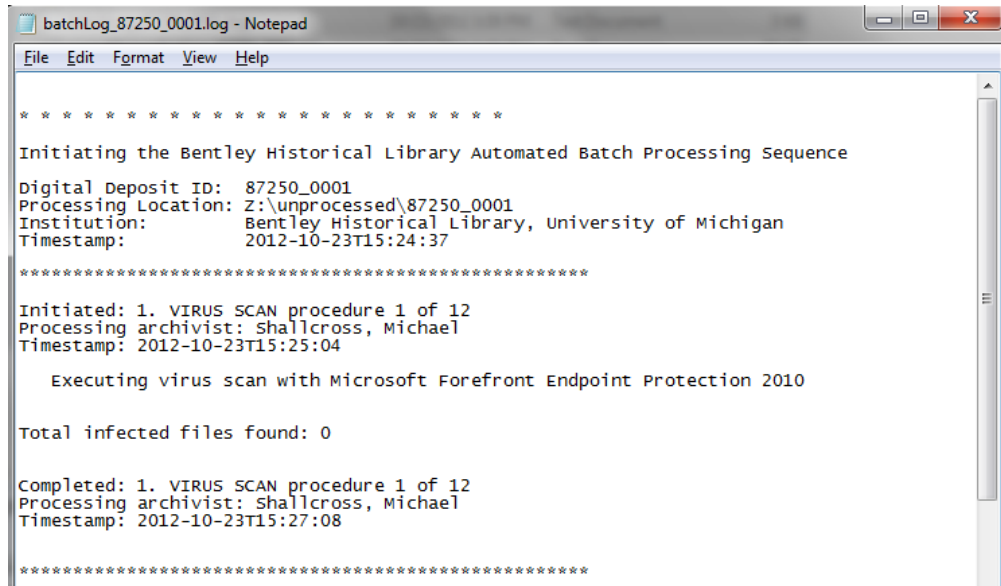
AutoPro may be completed in one session or may be stopped at any point.
Quitting an automated procedure mid-operation may result in loss of data.
Please complete a procedure before exiting the sequence.
*****

2
CURRENT STATISTICS--DEPOSIT 00132_0002:
FILES: 12
VOLUME: 31670657 bytes
*****

3
PROCESSING STATUS--DEPOSIT 00132_0002 (as of Thu 10/10/2013):
Completed: 1. VIRUS SCAN procedure 1 of 9
Completed: 4. FORMAT CONVERSION procedure 4 of 9
Completed: 5. PII SCAN procedure 5 of 9
Completed: 6. APPRAISAL, ARRANGEMENT, and PACKAGING FOR DEPOSIT procedure 6 of 9
*****

Enter an option: 4
```

At each stage of the workflow, AutoPro will record the procedure name, processing archivist, initiation and completion timestamps, and additional information in the main batch processing log file (see example below).



In addition, all ‘preservation events’ (that is, activities performed by the archivist that impact the provenance, authenticity, and/or integrity of the content) are recorded in a PREMIS (PREservation Metadata: Implementation Strategies) spreadsheet.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	eventType	eventId	eventIdentifierValue	eventDateTime	eventDetail	eventOutcome	linkingAgentId	linkingAgentValue	linkingAgentType	linkingAgentRole	linkingAgentSoftware		
2	virus scan	UUID	d76cfe99-9706-4f20-1212-07-27T15:50:2	2012-07-27T15:50:2	identification action complete	MARC21 Code	MiU-H	Executor	tool	Microsoft Ant software			
3	decompression of file	UUID	84c36d87-d675-4e30-1212-07-27T15:54:3	2012-07-27T15:54:3	decompression action complete	MARC21 Code	MiU-H	Executor	tool	7-Zip 9.20 software			
4	personally identifiable	UUID	f119bf41-1b56-4720-1212-07-27T15:58:2	2012-07-27T15:58:2	identify and re-act	MARC21 Code	MiU-H	Executor	tool	Identity Finder software			
5	message digest calculation	UUID	82262b07-9155-4c20-1212-07-27T16:09:3	2012-07-27T16:09:3	calculation of N	MARC21 Code	MiU-H	Executor	tool	md5deep 4.0 software			
6													
7													

From left to right, this document records (A) the event type, (B) the event identifier type (here, a Universal Unique Identifier, or UUID), (C) an event identifier value, (D) a time stamp, (E) a description of the event, (F) the outcome, (G) the agent identifier type (here, the MARC21 institution code), (H) the value or identifier for agent (here, MiU-H, the Bentley’s MARC21 identifier), (I) the type of agent, and then information on the software ‘agent’ employed (J-L).

At the conclusion of this (and subsequent) procedures, you will have the opportunity to enter A to ADVANCE to the procedure, M to return to the MAIN MENU or Q to QUIT AutoPro.

```
Options:
-A: ADVANCE to next step (3. FILE EXTRACTION)
-M: return to AutoPro MAIN MENU
-Q: QUIT
<A / M / Q>
```

If you choose to quit, you will resume processing at the point you left off.

## 4. Guide to Individual Procedures

The following pages will provide an explanation of and guidance for running the individual procedures in the AutoPro workflow. **Be sure to take your time!** While processing should be conducted in a timely and efficient manner, there is no race. Do not try to get through the workflow as fast as you can— instead pay careful attention to detail, especially in those steps that require user interaction, review of content, and metadata entry.

Also, despite extensive testing and troubleshooting, unique files or deposit conditions may result in an unrecoverable error or application failure. If you have any issues with AutoPro, please immediately share them with Mike so that a workaround can be developed in an efficient and timely manner.

### Step 1: Virus Scan

AutoPro runs Microsoft Antimalware Service Command Line Utility (a component of Microsoft's Forefront Endpoint Protection 2010) on each file in the deposit. If a virus or malware is detected, the antivirus software will delete the file and record the deletion in a log file.

If one or more infected file cannot be removed, you will be alerted by AutoPro and be permitted to view a listing of such files.

```
*****
Bentley Historical Library * * * Digital Curation Division
AutomatedProcessor v. 0.6 (c) 2012
*****

ANTIVIRUS SCAN
Scan complete; checking log file for issues...

  Total files scanned: 17
  Total infected files found: 4

User must address outstanding issues; see Digital Curation for assistance.
Hit "Enter" to see a list of files that could not be cleaned or removed.
You may wish to consult the antivirus scan log for more details:
--> C:\BHL\logs\87209_0001\virusScan_87209_0001.log

Press any key to continue . . . _
```

If this occurs, please consult with Mike; upon review of the antivirus log, it may be necessary to manually delete the remaining infected files.

### Step 2: File Extraction

This procedure consists of the identification of disk image (such as .E01, .IMG, .AFF, etc.) and archive files (such as .ZIP, .TAR, .RAR, etc.) extraction of content (with its original directory structure) with FTK Imager and 7-Zip, respectively.

## Disk Image Extraction with FTK Imager

Upon initiation of the procedure, AutoPro will search for common disk image file format. If none are found, the resource will continue on to search for common archive files (see below). Please note that given the wide variety of disk image file formats this step may not detect all formats. If additional disk images are detected in the Appraisal and Arrangement workflow step, they may be extracted at that time.

If disk image files are detected, AutoPro will open a text file with the full paths to each file; save locally or refer to it in its location in the 'decompress' folder in the deposit's /tmp/ directory.

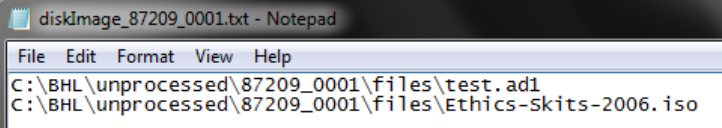
```
This procedure will search for disk image files (.ISO, .AFF, .DMG, etc.)
and archive files (.ZIP, .GZIP, .TAR, etc.) so that content may be
extracted.

Notepad has opened a list of common disk images found in the deposit.
Coordinate with Digital Curation to extract content from the Disk Images
with FTK Imager.

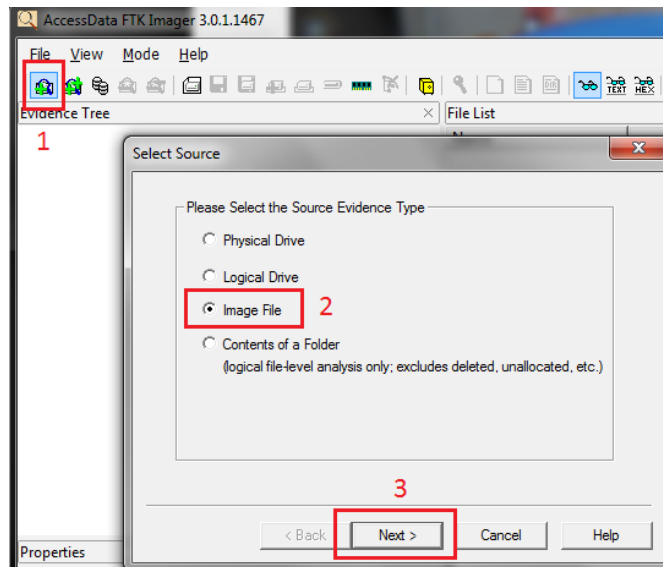
Please note that the contents of the disk images must be scanned for any
viruses or malware.

Options:
  -C: CONTINUE processing; AutoPro will conduct a virus scan again. You may
      then return to the main menu and initiate this procedure again.
  -Q: QUIT and resume later

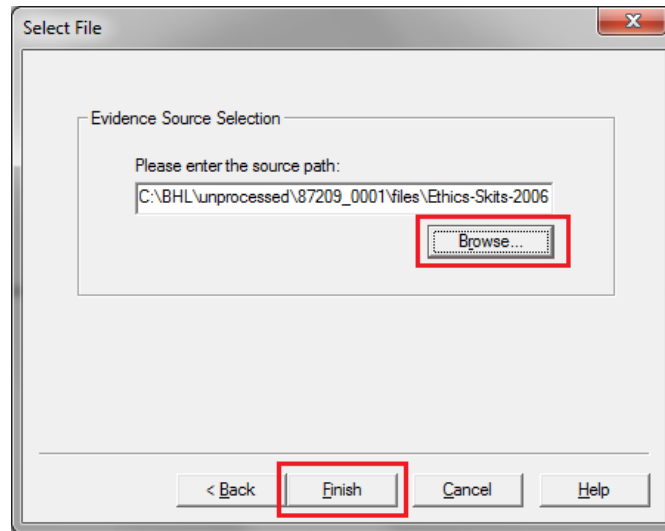
<C / Q>
```



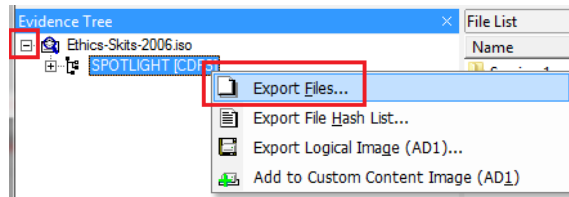
After opening FTK Imager, (1) click the “Add Evidence” icon (or go to File > Add Evidence Item), (2) check the radio dial next to “Image File,” and (3) click “Next.”



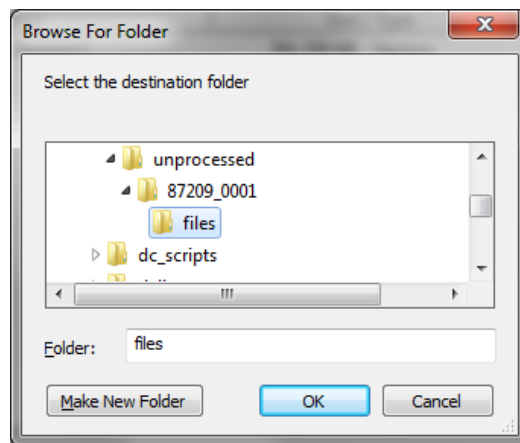
You may then copy and paste the path to a disk image file in the new window's text box (or browse to its location) before clicking “Finish.”



FTK Imager will now show the disk image in the upper-left “Evidence Tree” panel. Expand the disk image file (by clicking on the ‘+’ next to its name), right click on directory immediately within, and select the “Export Files...” option.

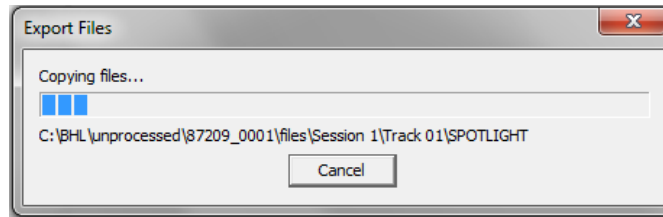


You will then be prompted to select a destination for the extracted files; to maintain original order, choose the same directory that holds the original disk image file.

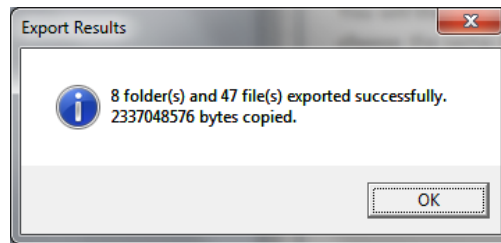


FTK Imager will immediately begin to extract files from the disk image:

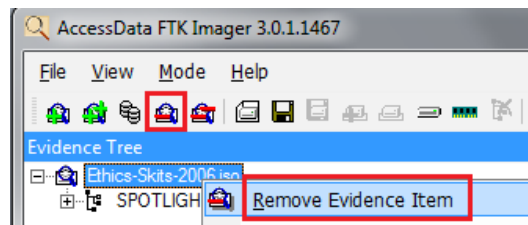




When the operation is successfully completed, you will receive a notice:



You may now remove the disk image (or “evidence item”) from FTK Imager; select the image file and either click the “remove evidence item” from the toolbar or right-click and select the option from the menu. You are now ready to proceed to the next disk image!



Once you are done extracting content from the disk image file(s), you may enter *C* into the AutoPro interface to COMPLETE this step. If you need to quit and resume file extraction at a later date, enter *Q* and recommence this procedure the next time you log into AutoPro.

```
Options:  
-C: CONTINUE processing; AutoPro will conduct a virus scan again. You may  
then return to the main menu and initiate this procedure again.  
-Q: QUIT and resume later  
<C / Q>
```

Upon completing the step, AutoPro will rerun a virus scan to verify that the newly extracted contents do not contain a virus or malware. The program will then return to the Extraction procedure to search for other archive files.

### ***Archive File Extraction with 7-Zip***

AutoPro will search for a variety of common archive file formats and, if any are found, extract the contents of each into a newly created folder that bears the name of the original archive file. To avoid name collisions and identify content extracted by the Bentley Historical Library, a suffix (consisting of “-bhl\_” and the eight-character CRC32 hash of the archive file) will be applied to the name of this new directory.

If an archive file cannot be extracted with 7-Zip, you will have an opportunity to manually extract content using the Windows “Compressed (zipped) folder” utility. If the archive file is corrupted, it may be impossible to retrieve content; consult with Mike if issues arise.

### Step 3: Personally Identifiable Information (PII) Scan

Personally identifiable information (PII) includes Social Security numbers (SSN), credit card numbers, bank account information, and passwords, any of which may be used to steal an individual’s identity or perpetrate fraud in some manner. To ensure that the Bentley Historical Library is aware of the presence of PII, AutoPro uses Identity Finder DLP Endpoint to search for such material. Identity Finder uses regular expressions as well as validation of information (via Luhn algorithms for credit card numbers and dictionaries of known SSN number patterns) to reduce the occurrence of false positive results.

AutoPro immediately initiates Identity Finder; the scan may take a long period of time, especially if the deposit includes large PDFs or email accounts. (AutoPro is configured to skip the scans of image, audio, and video files to reduce processing time). At the scan’s conclusion, you will be asked if Identity Finder produced any search results:

```
*****
Bentley Historical Library * * * Digital Curation Division
AutomatedProcessor v. 0.6 (c) 2012
*****

Identity Finder QUICK START GUIDE

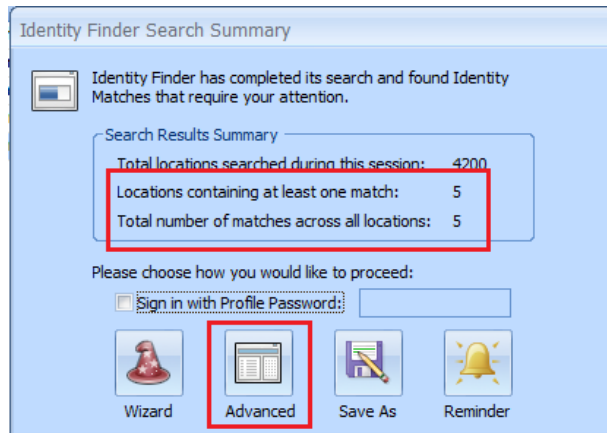
1. Identity Finder is now searching each file in the processing directory
   for Personally Identifiable Information (PII).

2. When the search is complete, an Identity Finder "Search Summary" window
   will appear with details on the number of files (i.e. "locations")
   searched and the number of identity "matches" found, if any.

Options:
-N: there are NO identity matches.
-Y: YES, there are identity matches.

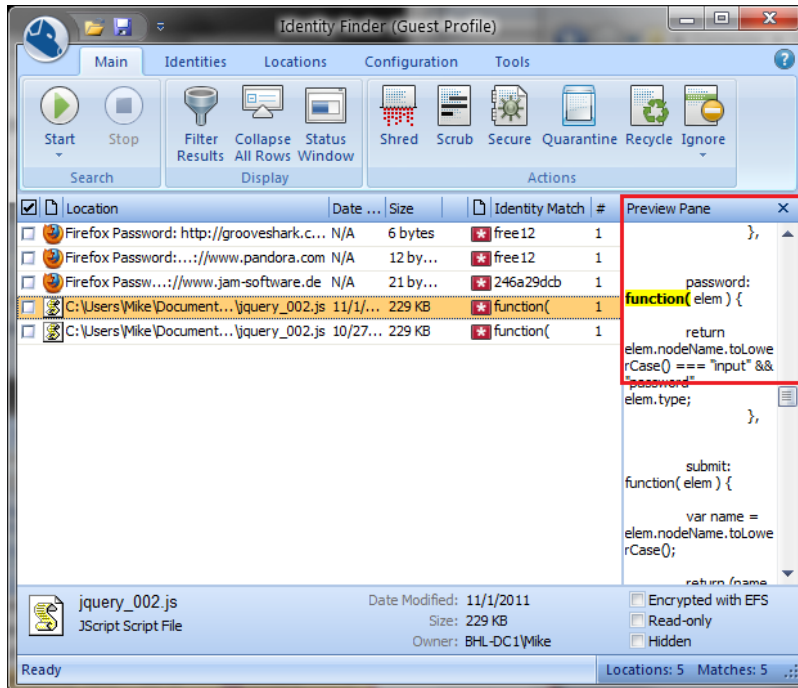
(Y / N)
```

The “identity matches” will be displayed in an Identity Finder Search Summary window. The following example has five matches; the number would read ‘0’ if no matches had been identified.

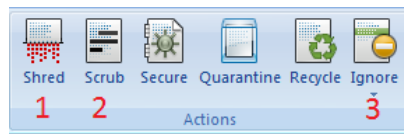


For no matches, you will simply enter *N* at the AutoPro prompt; if there are matches, enter *Y*, click the “Advanced” option and then follow the step-by-step instructions in AutoPro.

If there are identity matches, you will need to review each one to determine if the match is legitimate or merely represents a “false positive.” All files containing identity matches will appear in the *Location Pane* on the left-hand side of the Identity Finder interface. Please note that some of these files may contain multiple matches. If you click on a file (or an identity match), the potentially sensitive information will be displayed in the right-hand *Preview Pane*.



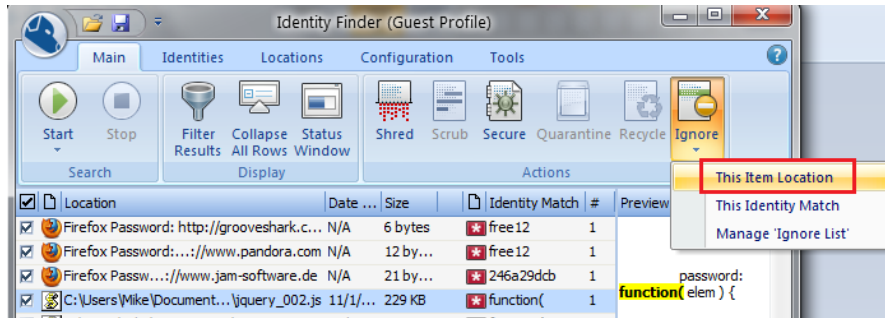
There are three possible actions you can take with content that has been found to contain an identity match:



1. If the content is in non-unique or routine business documents (such as business expense reports or P-Card logs), you may use the “SHRED” option to securely delete content. Check the boxes next to the target filenames, click the “Shred” action, and then confirm your choice. Please note that this option will permanently delete the files—please consult with Nancy, Mike, or division heads if you have any uncertainty!
2. If the identity match is in a plain text file or MS Office Open XML document (i.e. .DOCX, .PPTX, or .XLSX), you may use the “SCRUB” action to redact the PII from file. This option replaces the PII with a string of X’s. Please note that if a preservation copy of a MS Office 1997-2003 file has

been created, you may redact the PII from the preservation copy (in OOXML) and then SHRED the original. Please consult with a division head before taking this step.

3. For false positives, check the boxes next to the affected files, click on the “IGNORE” action, and then select “This Item Location” from the options.



Finally, if the identity match is in a significant record that must be preserved but cannot be redacted, it may be necessary to impose an access restriction. Please consult with Nancy, Mike or a division head to determine the appropriate restriction.

#### Step 4: Identify Missing File Extensions

To aid in the creation of preservation copies and facilitate the eventual use of content by patrons, AutoPro will help the processor identify missing or incorrect file extensions. To run this procedure you must make sure that DROID is configured so that it will NOT calculate MD5 checksums. (If checksums are calculated, this procedure may take a very long time to complete and the resulting spreadsheet of information will not be parsed correctly by AutoPro.)

```
*****
Bentley Historical Library * * * Digital Curation Division
AutomatedProcessor v. 0.9 (c) 2013
*****

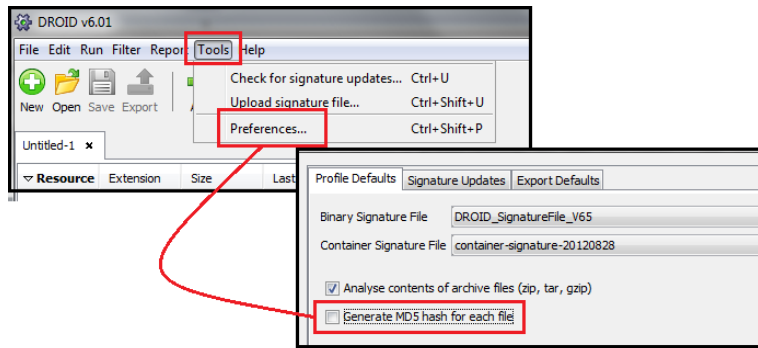
Initiated: 3. IDENTIFY MISSING FILE EXTENSIONS
*****

This procedure will permit you to add file format extensions to files for
which the original extension is either missing or is user-defined.

NOTE: Before continuing this procedure, you must make sure that DROID will
NOT generate MD5 checksums.

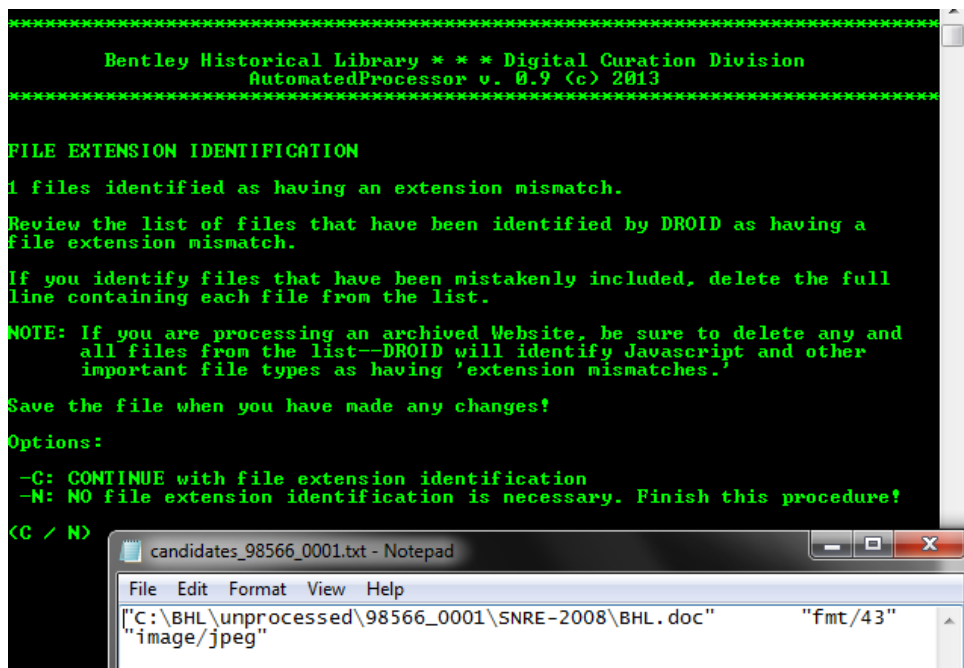
Is DROID currently configured to generate MD5 hashes? <Y / N> _
```

If you don't know or are unsure, assume that DROID is set to calculate checksums and answer Y. After DROID opens, it will check for any updates; if any are available, approve their installation. Next, click on the *Tools* item in the top navigation menu and select *Preferences*. Make sure that the box next to “Generate MD5 hash for each file” is **NOT** checked. Click OK and then close the DROID window.



AutoPro will then ask you to confirm that these settings have been saved and DROID is closed. AutoPro will launch DROID to produce a spreadsheet that identifies files with missing or mismatched file extensions.

Once this operation is complete, AutoPro will generate a DROID report and inform you of the number of files (if any) identified as having 'mismatched' extensions. A text file of these files will then open; each line will contain the full path, PRONOM unique identifier (PUID), and mime type (if the latter two were identified).



**Advanced user note:** File extension identification is not required; if there is an extremely large number of files with mismatched extensions or you have reason to believe the mismatches were incorrectly identified, you may finish the procedure immediately. You may also delete lines from the file if you do not wish the associated files to have extensions identified (for instance, Javascript and .CSS files from archived web pages will be incorrectly identified as having an extension mismatch).

If you continue, AutoPro will loop through this list of files, using the TrID file identification utility and the PRONOM file format registry to propose appropriate file extensions. Based upon these results (and, if necessary, your review of the file) you may choose to adopt one of the proposed extensions. Enter the three-digit extension without a period; AutoPro will apply it to the file and record the operation in a log file.

```

*****
                Bentley Historical Library * * * Digital Curation Division
                AutomatedProcessor v. 0.9 (c) 2013
*****

MANUAL FILE EXTENSION IDENTIFICATION
Filename: C:\BHL\unprocessed\98566_0001\SNRE-2008\BHL.doc
TrID output: probability of file-type association based on binary signature:
 38.4% .JPG  JFIF-EXIF JPEG Bitmap
 30.7% .JPG  JFIF JPEG Bitmap
 23.0% .JPG  JPEG Bitmap
  7.6% .MP3  MP3 audio

PRONOM format registry information:
  Mime type: image/jpeg
  Suggested extension(s) from the PRONOM format registry:
  --> .jpg
  --> .jpe
  --> .jpeg
*****

OPTIONS:
 - Enter an appropriate 3 character file extension based upon the TrID and
   PRONOM output
 - Hit "ENTER" to do nothing and leave the file as-is (recommended if file
   cannot be identified by TrID/PRONOM or by archivist's review)
 - Enter "Q" to QUIT and exit AutoPro
Extension (or "ENTER" to do nothing):

```

If you are unsure of what the extension should be, simply hit “Enter”—AutoPro will leave the file unchanged and proceed to the next one on its list.

AutoPro will continue until all the files on the list have been reviewed, with the addition of any extensions recorded in a log file.

### Step 5: Format Conversion

The creation of preservation copies for content in at-risk formats is one of the Bentley Historical Library’s primary preservation strategies. AutoPro searches for at-risk formats (using the matrix developed by the Division of Digital Curation) and creates temporary files listing all files associated with the following formats and media types (conversion software noted in italics):

- Raster images (.BMP, .PSD, .PCD, .PCT, and .TGA to .TIF): *Image Magick*
- Raw digital camera image files (.3FR, .ARW, .CR2, .DCR, .MRW, .NEF, .ORF, .PEF, .RAF, .RAW, .X3F to .JPG): *Image Magick*

- Vector images (.AI, .WMF, .EMF to .SVG): *Inkscape*
- PostScript files (.PS to .PDF/A): *Ghostscript*
- Encapsulated PostScript files (.EPS to .PDF/A) : *Ghostscript*
- Portable Document Format (.PDF to PDF/A) : *Ghostscript*
- Audio files (.WMA, .RA, .AU, .SND to .WAV): *ffmpeg*
- Video files (.FLV, .WMV, .RV, .RM, .MTS to .MP4 with H.264 encoding): *ffmpeg*
- DVD-encoded video (VIDEO\_TS folders to .MP4 with H.264 encoding): *Handbrake*
- Office Documents (.DOC, .PPT, .XLS to Open Office XML): *MS Office File Convertor*
- Email (.PST, .MSG, .EML, etc. to .MBOX): *Aid4Mail*

After compiling listing of each format/media type, AutoPro calls subroutines that kick off the various conversion applications, using preset parameters to create preservation copies. Each conversion involves the following steps:

- Check to see that a preservation copy has not already been created
- Generation of the preservation copy.
- Validation of conversion success
- If successful, the conversion is recorded in a log file.

These sustainable versions of content are stored alongside the original bitstream and are differentiated by the addition of a suffix to the filename (“-bhl\_” plus the 8 character CRC32 hash of the original file). Thus, a standard PDF file named “BHL-LogoBlack.pdf” will yield a preservation copy named “BHL-LogoBlack.\_bhl-f14fe8e6.pdf.” For the conversion of PDF files to PDF/A, AutoPro employs JHOVE to verify if the original file meets the PDF/A 1-a or 1-b specifications.

The conversion process is largely automated, with the exception of the email and DVD-encoded video routines. In each case, AutoPro will guide you through operations with detailed instructions. Please consult with Mike if any issues arise!

## **Step 6: Appraisal, Arrangement, and Packaging for Deposit**

This procedure provides you with an opportunity to gain full intellectual control of the content so that it can be meaningfully described in a finding aid and packaged in a manner that will facilitate its long-term preservation, management, and access. **Take time to review the files!** Familiarity with the content is essential for the production of rich metadata and informative finding aids; it is therefore *highly recommended* that you (1) *take notes* throughout this procedure to help with separations and gaining intellectual control of content and (2) *establish the intellectual arrangement* before completing any arrangement or packaging of files for deposit.

While digital processing should be done as efficiently as possible, it is very important that you do not rush through this step. If your collection lacks an appropriately detailed finding aid and descriptive metadata, researchers may not be able to discover valuable information therein. You may quit this procedure at any time and return to complete it at a later date.

AutoPro will display an overview of goals for this procedure; some of these actions may be inapplicable (i.e., no content needs to be separated or no physical arrangement of files is necessary). Note that if you have an archived website in the deposit, you will create a list of one or more resources that will be converted into hyperlinks to content for a BHL access page that AutoPro will automatically create in the next procedure.

```
*****
Bentley Historical Library * * * Digital Curation Division
AutomatedProcessor v. 0.9 (c) 2013
*****

APPRAISAL, ARRANGEMENT, and PACKAGING FOR DEPOSIT

This procedure will permit you to:

  1. Conduct an appraisal and review of content
  2. Record manual conversions of files (if necessary)
  3. Separate unneeded materials via right-click menu options--see manual
     for more information.
  4. Arrange content via Windows Explorer--only if necessary
  5. Package content in .ZIP files via right-click menu options--see manual
     for more information.

NOTE: For each archived Website you MUST create a list of one or more
      resources that will be linked to from a BHL access page.

Press any key to continue . . . _
```

After reading through the instructions, you may hit any key and then choose from the following options:

```
*****
Bentley Historical Library * * * Digital Curation Division
AutomatedProcessor v. 0.9 (c) 2013
*****

APPRAISAL, ARRANGEMENT, and PACKAGING FOR DEPOSIT

Please select from the following options:

CHARACTERIZE DEPOSIT                                REVIEW CONTENT
  1: Directory hierarchy                            Q: Quick View Plus
  2: Relative size of directories                    U: ULC Media Player (audio/video)
  3: File extensions / formats                       IRF: IrfanView (raster images)
  4: Age of files                                    INK: Inkscape (vector images)
  5: Duplicate content

OTHER:
M: Record MANUAL conversions
L: create a LIST of resources for HTML access page

C: COMPLETE procedure
QUIT: QUIT AutoPro and resume appraisal and arrangement at a later time

Enter an appraisal option: _
```

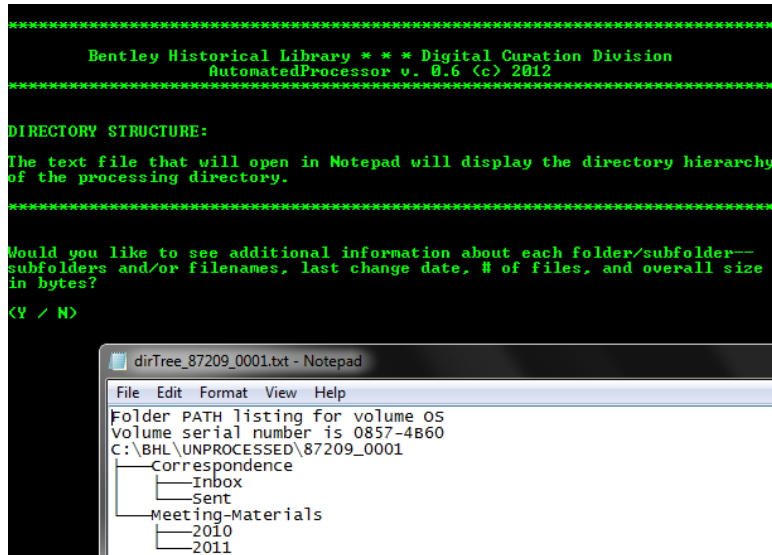
### ***Characterize Deposit: Understanding the Extent and Diversity of Content***

Items within the “Characterize Deposit” employ TreeSize Professional and Windows utilities to provide visualizations of various aspects of the deposit. While these procedures may not be necessary if you have a clear understanding of the nature and extent of the deposit, they can be useful in reviewing a large collection of unstructured data.



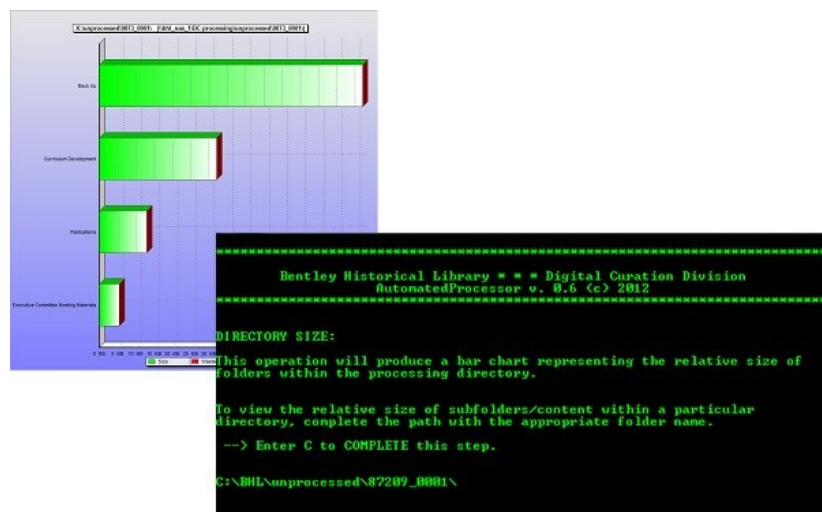
## Review Directory Hierarchy

Reviewing the directory hierarchy may aid in the appraisal of the deposit and help you to understand the breadth and scope of content transferred to the library. Folder naming conventions and directory structure may also reveal organizational principles (if any) and areas of particular interest. The “Directory Hierarchy” displays the structure of folders within the deposit:



## View Relative Size of Directories

The “Relative size of directories” will produce a bar graph comparing the sizes of the folders in the main processing directory. To examine the relative size of the contents of subfolders, simply complete the folder path in the AutoPro interface or enter C to complete the step.



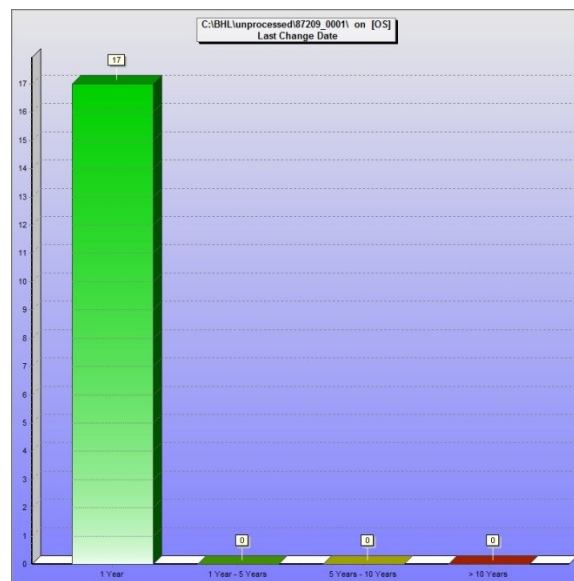
## Compare File Extensions within Deposit

The “File extensions” option produces a list of file extensions grouped by format type (images, video, Office files, etc.), with the number of files for each extension included. This information can be useful in understanding the dominant types of materials as well as for noting unique file types.

	A	B	C	D	E	F	G
1	<b>Extension</b>	<b>Size</b>	<b>Allocated</b>	<b>Percent</b>	<b>Files</b>	<b>Description</b>	
2	<b>Container Files</b>	<b>241.4 MB</b>	<b>241.4 MB</b>	<b>51.2%</b>	<b>3</b>	<b>Compressed Archives and disk images</b>	
3	.zip	169.3 MB	169.3 MB	35.9%	2	Compressed (zipped) Folder	
4	.tar	72.1 MB	72.1 MB	15.3%	1	TAR File	
5	<b>Mail Files</b>	<b>153.0 MB</b>	<b>153.0 MB</b>	<b>32.5%</b>	<b>18</b>	<b>Email messages and files of email clients</b>	
6	.pst	74.6 MB	74.6 MB	15.8%	1	Outlook Data File	
7	.mbox	73.1 MB	73.1 MB	15.5%	8	MBOX File	
8	.eml	5.3 MB	5.3 MB	1.1%	7	E-mail Message	
9	.msg	0.0 MB	0.0 MB	0.0%	2	Outlook Item	
10	<b>Graphic Files</b>	<b>39.2 MB</b>	<b>39.2 MB</b>	<b>8.3%</b>	<b>8</b>	<b>Files containing pictures, images or mouse cursors</b>	
11	.arw	14.3 MB	14.3 MB	3.0%	1	ARW File	
12	.orf	11.6 MB	11.6 MB	2.5%	1	ORF File	
13	.psd	5.0 MB	5.0 MB	1.1%	1	IrfanView PSD File	
14	.bmp	4.1 MB	4.1 MB	0.9%	1	IrfanView BMP File	

### Compare Age of Files

The “Age of files” option generates a bar graph representation of the age range for content, identifying material produced within the year, from 1 to 5 years ago, from 5 to 10 years, and then older than 10.



### Review Duplicate Content

The “Duplicate content” option uses MD5 checksum comparison to identify duplicate material and then produce a .CSV report. The Bentley Library is not going to do mass de-duplication due to the difficult and labor-intensive analysis required to identify the ‘record’ version of content. To maintain original order and prevent the potential loss of record versions of content, you may allow duplicate content to remain in the deposit.

	A	B	C	D	E
1	Name	Path	Size	Last Change	Last Access
2	todo.txt	(2 Files)	22.4 KB	8/27/2012 5:58 AM	8/27/2012 5:58 AM
3	todo.txt	C:\Users\Mike\Desktop\test\FITS\openplanets-fido-ae19806\	11.2 KB	8/27/2012 5:58 AM	8/27/2012 5:58 AM
4	todo.txt	C:\Users\Mike\Desktop\test\FITS\vido\openplanets-fido-ae19806\	11.2 KB	8/27/2012 5:58 AM	8/27/2012 5:58 AM
5	ARL SPEC Kit.zip	(2 Files)	30.0 KB	10/3/2012 1:06 PM	[multiple]
6	ARL SPEC Kit.zip	C:\Users\Mike\Desktop\test\9999_0001\Back Up\Executive Committee Meeting Materials\Correspondence\	15.0 KB	10/3/2012 1:06 PM	11/1/2012 12:46 PM
7	ARL SPEC Kit.zip	C:\Users\Mike\Desktop\test\9999_0001\Executive Committee Meeting Materials\Correspondence\	15.0 KB	10/3/2012 1:06 PM	2/6/2012 9:53 AM
8	Updated_Agenda.eml	(2 Files)	36.6 KB	[multiple]	[multiple]
9	Updated_Agenda.eml	C:\Users\Mike\Desktop\test\9999_0001\Back Up\Executive Committee Meeting Materials\Correspondence\L	18.3 KB	#####	11/1/2012 12:46 PM
10	Updated_Agenda.eml	C:\Users\Mike\Desktop\test\9999_0001\Executive Committee Meeting Materials\Correspondence\Literature	18.3 KB	10/3/2012 1:06 PM	10/28/2011 1:37 PM
11	Agenda_for_News_and	(2 Files)	37.2 KB	[multiple]	[multiple]
12	Agenda_for_News_and	C:\Users\Mike\Desktop\test\9999_0001\Back Up\Executive Committee Meeting Materials\Correspondence\L	18.6 KB	#####	11/1/2012 12:46 PM
13	Agenda_for_News_and	C:\Users\Mike\Desktop\test\9999_0001\Executive Committee Meeting Materials\Correspondence\Literature	18.6 KB	10/3/2012 1:06 PM	10/28/2011 1:36 PM
14	ccn.doc	(2 Files)	38.0 KB	10/3/2012 1:06 PM	[multiple]
15	ccn.doc	C:\Users\Mike\Desktop\test\9999_0001\Executive Committee Meeting Materials\Agendas\	19.0 KB	10/3/2012 1:06 PM	11/1/2011 1:02 PM
16	ccn.doc	C:\Users\Mike\Desktop\test\9999_0001\Back Up\Executive Committee Meeting Materials\Agendas\	19.0 KB	10/3/2012 1:06 PM	11/1/2012 12:46 PM
17	CCE_assignment.doc	(2 Files)	49.0 KB	10/3/2012 1:06 PM	[multiple]
18	CCE_assignment.doc	C:\Users\Mike\Desktop\test\9999_0001\Back Up\Executive Committee Meeting Materials\Agendas\	24.5 KB	10/3/2012 1:06 PM	11/1/2012 12:46 PM
19	CCE_assignment.doc	C:\Users\Mike\Desktop\test\9999_0001\Executive Committee Meeting Materials\Agendas\	24.5 KB	10/3/2012 1:06 PM	11/1/2011 1:02 PM

At the same time, if the duplicate content report reveals a high concentration of duplicate content—or entire folders that were used to backup material—you may separate this content with the methods described below.

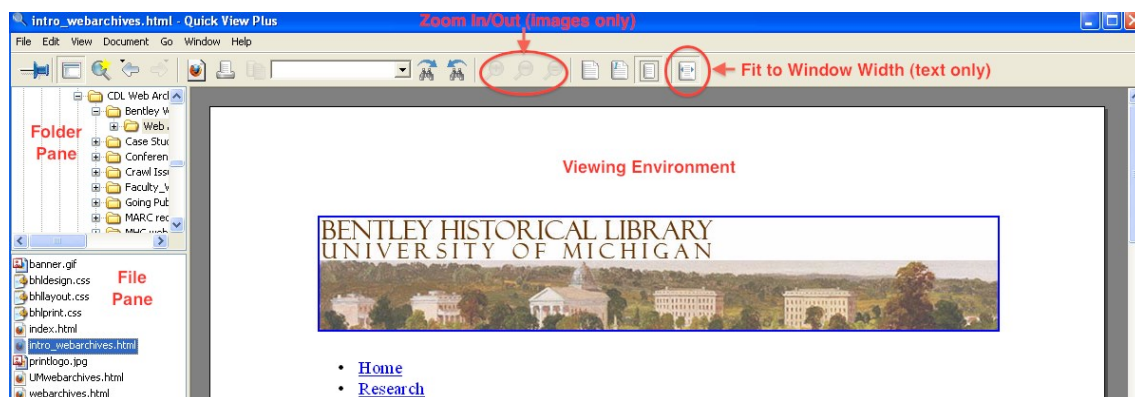
### Manually Review Content

AutoPro also brings together various tools that can be used to review materials; of these, Quick View Plus (QVP) is ideal for browsing through a wide variety of Office files, images, PDFs, some email formats (including .PST and .MBOX), and other common file types. IrfanView’s thumbnail view allows for quick browsing of large image galleries (although QVP is able to view many of the same files) while Inkscape should be used for vector images and VLC Media Player for audio and video content. The following will provide a brief overview of how to operate these various applications after opening them via AutoPro.

#### Quick View Plus:

The QVP interface is divided into three main parts in addition to the navigation menu and ribbon at the top of the application window. The right portion of the interface holds the *Viewing Environment* while the left-hand side is divided between the *Folder Pane* on the top and the *File Pane* on the bottom.

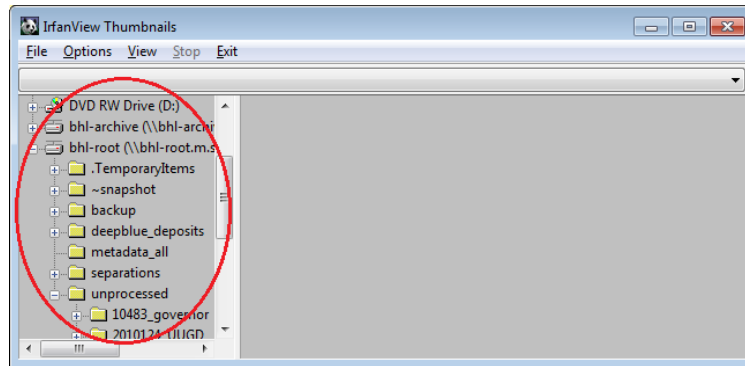
After QVP opens, use the mouse or arrow keys (right and left arrows may be used to expand/collapse subfolders) to navigate to the appropriate directory in the *Folder Pane*. Once the appropriate folder has been selected, a list of its contents (both subfolders and files) will be displayed in the *File Pane*. You may use the mouse or the tab key to move to the *File Pane*; then, whatever file is highlighted will appear in the *Viewing Environment*. Please note that very large files (especially email) may take longer to open.



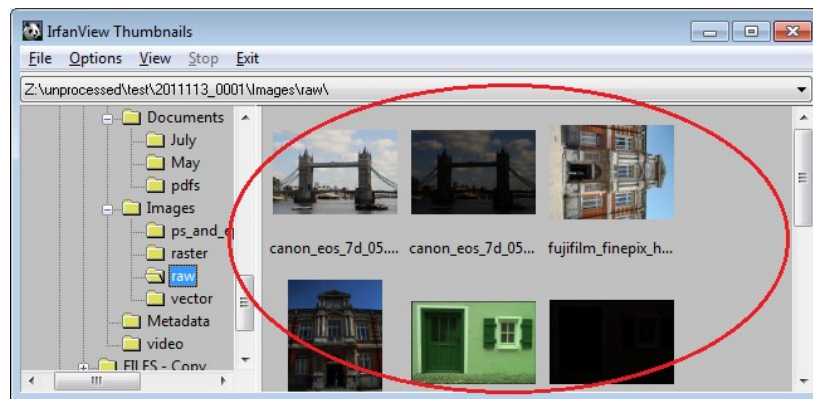
While viewing a file, you may search for text in the ribbon's text box, zoom in/out (for images only) and adjust the document size so that it fits entirely in the view area or is fit to the area's width.

### IrfanView:

When IrfanView is selected, AutoPro will open the application in "thumbnail" view, with the main processing directory selected in the left-hand folder pane. It will be necessary for you to navigate down into the directory hierarchy to the folder(s) that contain image galleries.

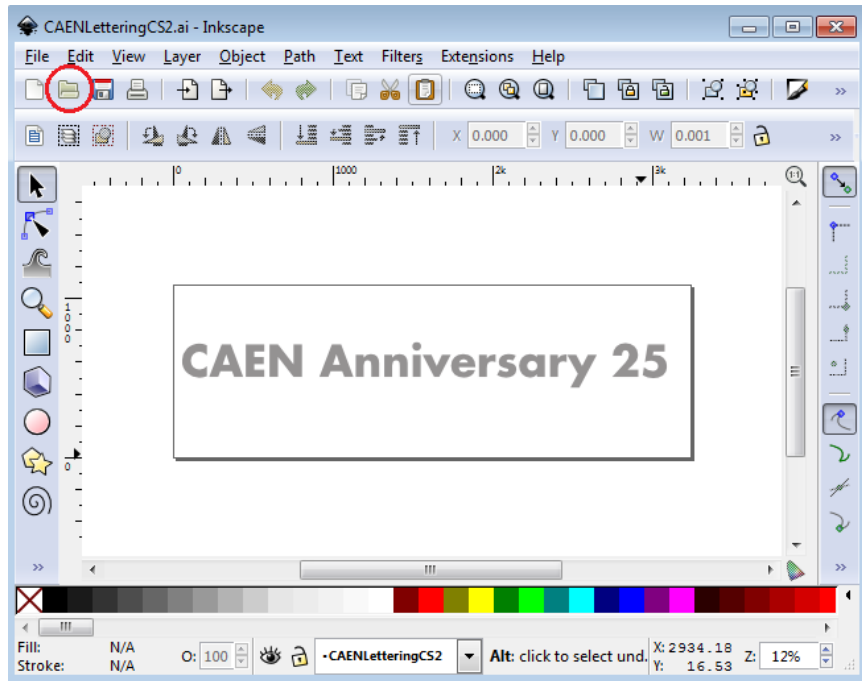


Once you have selected a folder with images, IrfanView will automatically load thumbnails of all image files into the viewing pane. You may then click on a thumbnail to take a closer look at the full-scale or use the navigation pane to browse to other folders.



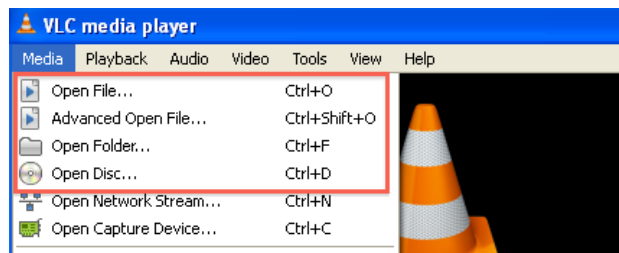
### Inkscape:

Inkscape may be used to view vector image files (such as .SVG, .AI, .WMF, etc.) that cannot be opened in IrfanView or Quick View Plus. AutoPro will open Inkscape and you will then need to click on the open file icon (or use *File > Open* in the navigation menu) to access vector image files.



VLC Media Player:

AutoPro will open VLC Media Player after which you may use the *Media* menu item to open a single file or an entire folder (in which case, all audio/video files within the folder will be added to a playlist).



Play controls are located at the bottom of the Media Player window; in addition to *Play*, *Pause*, and *Stop* buttons, the archivist may fast forward or reverse progress by adjusting the slider on the progress bar.



If there are multiple files in a playlist, you may click the arrow keys to move to the next/previous item in a playlist.

***Separations: Removing Superfluous Content***

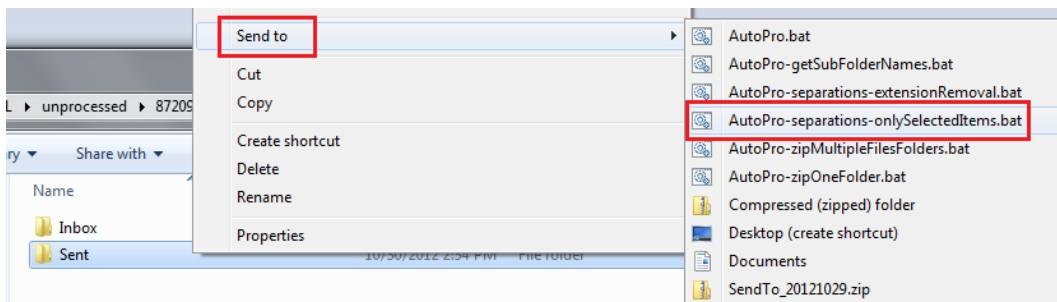
The appraisal and review process may reveal content that should be separated from the collection prior to its deposit in a long-term repository. This may include certain file types and content deemed to be

superfluous or outside the collecting scope of the library. When content is moved to the separations directory, AutoPro will recreate the folder structure of the deposit so that separated materials retain their original position and context.

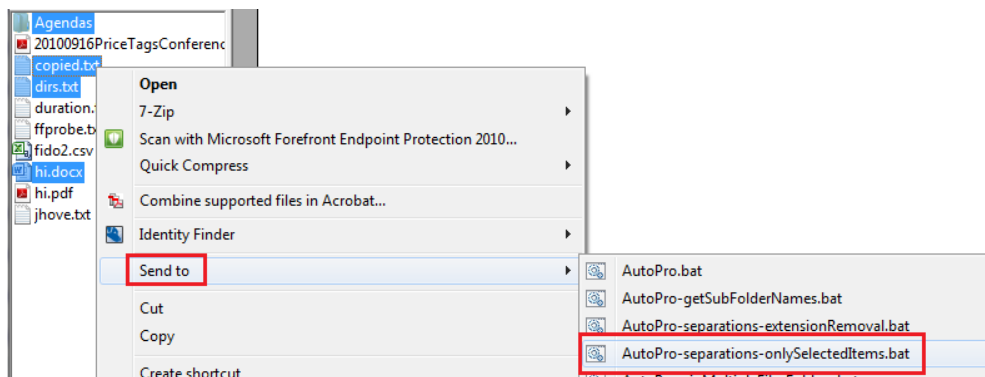
Upon initiation of the Appraisal and Arrangement procedure, AutoPro will search for and separate a number of common files generated by operating systems. These include thumbs.db and .LNK file shortcuts on Windows and .DS\_STORE and resource fork (.\_\* ) files produced by Macs. If you would like to move additional files to the deposit’s separations directory, you have two options, both of which are batch files found in the “Send to” section of the right-click context menu. Please note that each option is available when you are reviewing content with Quick View Plus or a Windows Explorer window.

### Resource #1: AutoPro-separations-onlySelectedItems.bat

Use the batch file “AutoPro-separations-onlySelectedItems.bat” to remove only those items you have selected to the separations directory. As the following example illustrates, this option may be used on a single file or folder:



This option may also be used on multiple files and/or folders, selected by clicking on the left mouse button while holding down the “Shift” key (to separate an entire range of files) or the “Ctrl” key (to choose a select number of files and/or folders for separation):



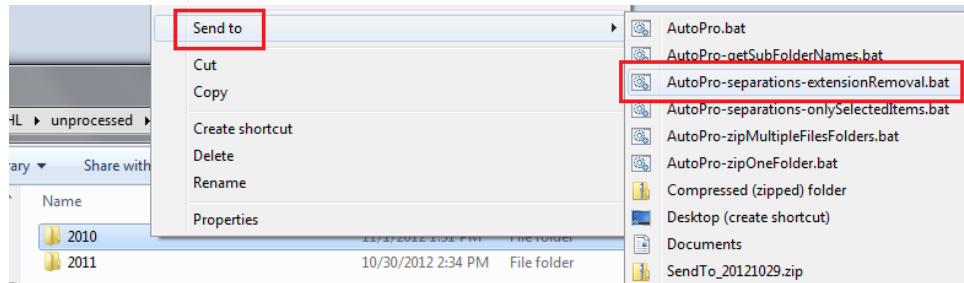
Please note that when selecting multiple files and/or folders for separation, all content must reside in the same parent directory.

### Resource #2: AutoPro-separations-extensionRemoval.bat

In the second option, the archivist may remove all files of a selected extension from a given folder—and all subfolders within that directory. **Use this option with care!** If used at an upper level of the

processing directory hierarchy, all files of the chosen extension will be removed from each subfolder in that branch of the deposit.

To use, this option, right-click on a folder and then click on the batch file “AutoPro-separations-extensionRemoval.bat” from the "Send to" context menu.



A new CMD.EXE window will open and prompt you to enter the extension for the files you wish to remove. Be sure to enter the appropriate extension (as used in Windows) and to precede it with a period (i.e. “.ini”).

```
*****
Bentley Historical Library * * * Digital Curation Division
AutomatedProcessor v. 0.7 (c) 2012
*****

SEPARATE ALL FILES OF A SPECIFIC FILE EXTENSION

At the prompt, enter the file extension (with leading period) you wish to
remove from this directory. USE CAUTION! The program will remove every file
with this extension to the separations directory.

Example: .ini

Enter file extension with a leading period: .txt
All files with this extension will be removed from this directory: .txt

Is this extension entered correctly? (Y/N)
```

AutoPro will ask you to verify your choice; enter *N* to identify a different extension or *Y* to proceed with the separation. Once AutoPro has an affirmative, all files with that extension will be removed to the separations directory.

At the conclusion of the Appraisal and Arrangement procedure, AutoPro will generate a manifest of all files that have been separated from the deposit (with figures on the number of files and volume) and store a copy of this file with other log files.

### ***Record Manual Conversions***

When this option is selected, a new screen will open with instructions for recording the original filename, the filename of the new preservation/access version, and the software used to perform the conversion.

```

*****
Bentley Historical Library * * * Digital Curation Division
AutomatedProcessor v. 0.9 (c) 2013
*****

MANUAL FORMAT CONVERSION

This step will help record the manual conversion of files. At the prompt,
hit "Enter" to open a spreadsheet or "Q" to QUIT and return to main menu.

On each row of the spreadsheet, record:
1. The original filename
2. The filename of the new preservation version
3. The software (with version number) used to perform the conversion

Options:
Q: QUIT and resume later
S: spreadsheet is completed and SAVED

<Q / S>: _

```

You will record this information in an Excel spreadsheet that will open automatically, using one row for each file. Consult with Digital Curation if you need help automating this step. If there are a large number of files, you may quit this action and finish at a later date.

***Create a List of Resources for an HTML Access Page***

Archived websites captured by the library with HTTrack or Teleport Pro will be placed in Deep Blue with an HTML document page that describes the content and provides access points to significant information. Once this option is selected, you will be asked to provide the full path to the folder that holds the archived website. If the material is in a domain folder (i.e., \www.soe.umich.edu\ ) you should create a folder to hold the site (i.e. \SOE-Website-2008\).

```

*****
Bentley Historical Library * * * Digital Curation Division
AutomatedProcessor v. 0.9 (c) 2013
*****

IDENTIFY RESOURCES TO BE LINKED TO IN THE BHL ACCESS PAGE

At the prompt, copy and paste the full path to the folder that contains the
archived Website.

--> NOTE: Each archived version should be in its own folder; this folder
should be located immediately within the deposit directory.

Be sure to make any changes or corrections to the folder name before
entering it here.

Folder: C:\BHL\unprocessed\98564_0001\SNRE-2008

```

You will then have an opportunity to record the date of any prior processing and the name of the processor; if not applicable, hit enter.



```

At the prompt, enter information about any earlier processing of this
archived website in the following format:

--> Smith, Joe; 2003

Leave blank and hit "Enter" if there were not previous processors.
Deromedi, Nancy; 2007

```

Next, enter the full path to a single resource (HTML or other file) you wish to include as a hyperlink in the BHL access page that AutoPro will create. After you verify that the path is correct, you will enter a title that identifies the resource for users:

```

*****
          Bentley Historical Library * * * Digital Curation Division
          AutomatedProcessor v. 0.9 (c) 2013
*****

IDENTIFY RESOURCES TO BE LINKED TO IN THE BHL ACCESS PAGE
FOLDER: SNRE-2008

At the prompt, enter the full path to a single resource you wish to include
as a link in the BHL access page:
Resource: "C:\BHL\unprocessed\98564_0001\SNRE-2008\www.umich.edu\~wctf\index.
htm"

Is the above resource correctly entered? (Y / N) y

Enter a descriptive title for this hyperlink: Home page for the 2005 Women of
Color Task Force website

```

You will then be permitted to add additional resources if multiple links to content are required.

### ***Arrangement***

The Bentley Historical Library strives to respect provenance and maintain the original order of content in order to preserve important contextual information found in the structure of directories and the associations of different files. Given the structure that many record creators impose on their files and the importance of original order, **additional arrangement will be unnecessary in most cases.**

At the same time, a basic assumption in our digital processing workflow is that folders in the top level of a processing directory represent key functions or activities of the record creator. You may therefore need to create a top-level folder in order to organize files or subfolders into series (avoid using spaces and/or non-alphanumeric characters in the names). It may also be necessary to impose some organizational principle if the files were copied pell-mell from their source location with no structure. While manual arrangement may be performed via Windows Explorer, you can also ‘arrange’ content during the packaging process described below. **Be sure to complete the intellectual arrangement before you arrange or package any files!**

### ***Packaging Content for Deposit and Defining Deep Blue ‘Items’***

To ensure that content is packaged correctly the first time, you **MUST**:

- Complete your intellectual arrangement.

- Prepare a draft of your finding aid and submit it for review.
- Discuss with Digital Curation how the content will be deposited and presented in Deep Blue. Develop a deposit plan to guide your packaging.
- **Use Windows Task Manager to kill Quick View Plus if you used it to review content.** Even after you close the program, QVP continues to run in the background and 'use' the files and/or folders you accessed with it.

Finally, be sure to complete all packaging before proceeding to the next procedure. There are two options for packaging content, both of which are Windows batch files available via the right-click "Send to" menu.

### Determining How to Package Content

The processing archivist (in consultation with Digital Curation) must determine when it is appropriate to deposit individual files or to package multiple files (or subfolders) into a .ZIP file. This decision will depend upon a number of factors:

- The intellectual arrangement of the material in the finding aid. Ideally, we would like to package content at a level where we can provide clear descriptive metadata. We also want to provide access (via links in EAD finding aids) to a discrete and well-defined chunk of content. Depending on the collection and material at hand, we may package content at the series, subseries, folder, or item level.
- The size of files—individually or as bundled in one or more .ZIP files. To avoid excessive download times for users, .ZIP packages should not exceed 2 GB, if possible. (Of course, very large audio or video files will be deposited 'as is.')

There is no 'one way' to ready materials for deposit into Deep Blue, but the following scenarios will suggest the possible ways in which content may be packaged:

- A deposit with a large number of small Office files related to a single function might be packaged into a single .ZIP file (which might represent an entire series, subseries, or folder).
- If the deposit contains records in separate folders that are related to diverse functions, each of these subfolders may be packaged as a separate .ZIP file.
- In some cases (i.e., a large number of uncompressed .TIFF images), it may be necessary to divide a single folder into multiple .ZIP file packages so that the materials can be uploaded and downloaded efficiently. In this case, the .ZIP file names should contain the original folder name as well as an indication of the breakdown (alphabetical or chronological): *meeting\_minutes-January-April.zip* or *committee\_reports\_A-M.zip*
- Very large files (video or audio, for example) or high-value content that needs to be described or linked to at the item level will be deposited individually.

*\*\*Please note that a single deposit may include any or all of the above packaging options\*\**

### Example of a Packaging and Deposit Plan

To better understand how content may be deposited, consider the following example from the *Toward A Fair Michigan Records*. The former program director of Toward A Fair Michigan (TAFM) transferred her work files to the library on a single optical disk in 2010. The files were arranged in a single directory with multiple subfolders that reflected different aspects of her work as program director. In processing the collection, the archivist determined that this group of files, taken as a whole, represented a distinct series: **Program Director's Files**. The subfolders in the main directory were determined to represent a 'level 2' hierarchy, as they dealt with specific functions of her role (i.e., Board Materials, Correspondence, Litigation, Press, Programs, etc.).

With these considerations in mind, the TAFM packaging/deposit strategy led to the following representation in EAD:

<u>Container / Location</u>	<u>Title</u>
	<b>Program Director's Files</b> [series]:
	The Program Director's Files series is comprised of born-digital records created and maintained by TAFM Program Director Carol Allan and includes incoming and outgoing correspondence, press releases, organizational bylaws, and logistical information for public events. The series also features audio and video recordings of public forums on the Michigan Civil Rights Initiative. This series contains materials related to the other two series in the collection.
Online	Board, 2006 (ZIP archive file) [ <a href="#">download item</a> ]
Online	Correspondence, 2006 (ZIP archive file) [ <a href="#">download item</a> ]
Online	Leadership Assembly, 2006 (ZIP archive file) [ <a href="#">download item</a> ]
Online	Litigation, 2006 (ZIP archive file) [ <a href="#">download item</a> ]
Online	Organizational Documents, 2005 (ZIP archive file) [ <a href="#">download item</a> ]
Online	Press, 2005-2007 (ZIP archive file) [ <a href="#">download item</a> ]
	Programs
	Audio files
Online	Benton Harbor debate, 2006 (ZIP archive file) [ <a href="#">download item</a> ]
Online	Ellis Cose Colloquium, 2006 (ZIP archive file) [ <a href="#">download item</a> ]
Online	Grosse Pointe Unitarian debate, 2006 (ZIP archive file) [ <a href="#">download item</a> ]
Online	Mackinac Center-Lansing debate, 2005 (ZIP archive file) [ <a href="#">download item</a> ]
Online	Michigan Black Republican Council talk, 2005 (ZIP archive file) [ <a href="#">download item</a> ]
Online	Union Missionary Baptist Church debate, 2006 (ZIP archive file) [ <a href="#">download item</a> ]
Online	Wayne State University debate and summation, 2006 (ZIP archive file) [ <a href="#">download item</a> ]

The large size of the deposit (11.5 GB) required the processing archivist to package the content into multiple zip files. Rather than create arbitrary divisions within the content, each subfolder was packaged in a single .ZIP file, with the exception of the 'Programs' folder. This directory included logistical information about TAFM's programs as well as some fairly large audio recordings of public events. A decision was made to package each of the audio recordings (i.e., both the original file and preservation copy created by the library) into separate .ZIP files so that users could access a single event without having to download a very large file that contained recordings of all the programs. The materials were placed in Deep Blue in accordance with the intellectual arrangement and deposit plan:

[Program Director's Files - Leadership Assembly](#)

Toward a Fair Michigan (2013)

 (2.8MB)

[Program Director's Files - Litigation](#)

Toward a Fair Michigan (2013)

 (1.5MB)

[Program Director's Files - Organizational Documents](#)

Toward a Fair Michigan (2013)

 (270.9KB)

[Program Director's Files - Press](#)

Toward a Fair Michigan (2013)

 (752.2KB)

[Program Director's Files - Programs - Audio Files - Benton Harbor Debate](#)

Toward a Fair Michigan (2013)

 (902.8MB)

[Program Director's Files - Programs - Audio Files - Ellis Cose Colloquium](#)

Toward a Fair Michigan; Cose, Ellis (2013)

 (763.6MB)

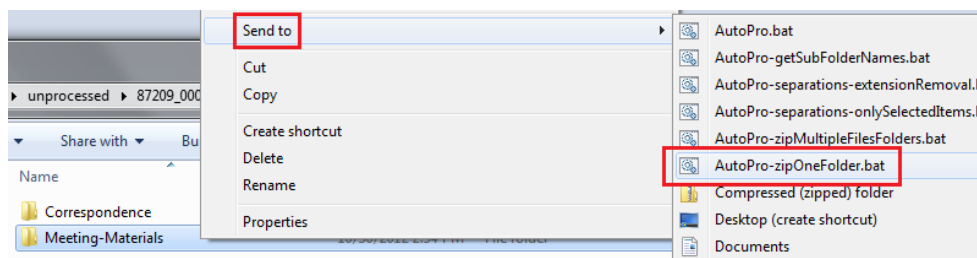
[Program Director's Files - Programs - Audio Files - Grosse Pointe Unitarian Debate](#)

Toward a Fair Michigan (2013)

 (735.5MB)

### Packaging Content with *AutoPro-zipOneFolder.bat*

To zip a single folder and all files and subfolders therein into a single uncompressed .ZIP file, right-click on the folder and select the batch file “AutoPro-zipOneFolder.bat” from the “Send to” context menu.



This operation may be performed on a top-level folder within the processing directory (i.e. \87209\_0001\ ) or a subfolder thereof. After selecting the batch file, a new CMD.EXE window will open and you will be asked to supply a descriptive filename (without the .ZIP extension) or simply retain the folder name for the .ZIP file. (The folder itself will be retained in the .ZIP file—do whatever makes the most sense to you!)

```

*****
Bentley Historical Library * * * Digital Curation Division
AutomatedProcessor v. 0.8 (c) 2013
*****
--> PACKAGING FOLDER: C:\BHL\unprocessed\98564_0001\finding-aids
At the prompt, enter a filename (without .zip extension) or Q to QUIT. The
ZIP file will be saved to C:\BHL\unprocessed\98564_0001
NOTE: Hit "Enter" to use the folder name as the filename.
<Filename / Q / "Enter"> _

```

After you verify that the filename is correct, AutoPro will package the entire folder in a .ZIP file (preserving the relative path to the folder and generating a manifest of all files therein), place it at the top-level of the deposit directory, and delete the original folder.

If your deposit includes an archived website, AutoPro will check to see if the folder in question holds a website:

```

*****
Bentley Historical Library * * * Digital Curation Division
AutomatedProcessor v. 0.8 (c) 2013
*****
--> PACKAGING FOLDER: C:\BHL\unprocessed\98567_0001\www.greatday.com
Is this an archived website? (Y / N)

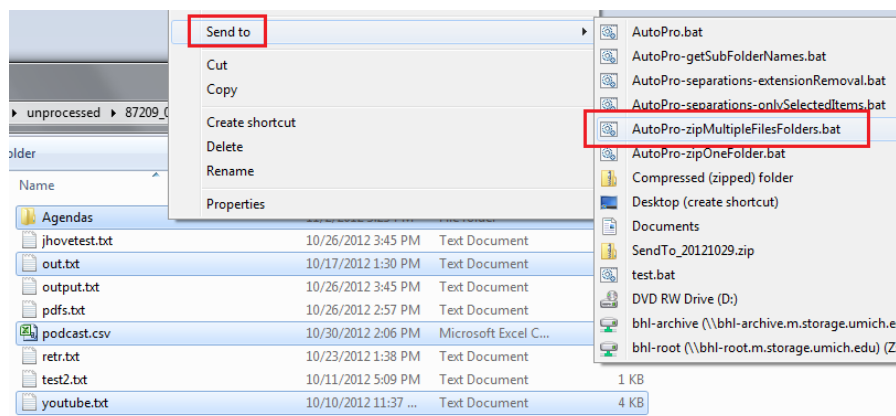
```

If so, AutoPro will automatically use the folder name as a title for the .ZIP file.

### Packaging Content with *AutoPro-zipMultipleFilesFolders.bat*

The other option to package content is to place multiple files and/or folders into a single uncompressed .ZIP file. **Please note that you will be limited to about 110 items per packaging attempt**—it may therefore require numerous operations to package a large number of individual files/folders.

To initiate this procedure, select the desired files and/or folders, using the “Shift” or “Ctrl” keys and the left mouse button as needed. Once the appropriate content is selected, right-click on one of the items and choose the batch file “AutoPro-zipMultipleFilesFolders.bat” from the “Send to” context menu.



If creating a new file, you must supply a descriptive filename (without the .ZIP extension); AutoPro will package the items in a .ZIP file (preserving the relative path to the folder and generating a manifest of all

files therein), place it at the top-level of the deposit directory, and delete the original content. If adding to an existing file, copy and paste the full path to the file in the window; AutoPro will add all the material to the specified file (preserving relative paths), update the .ZIP manifest file, and delete the original content.

### ***Completing the Appraisal and Arrangement Procedures***

Once you have completed all the procedures in this workflow step—appraisal and review of content, separations, arrangement, and packaging—select option C from the Appraisal and Arrangement main menu to COMPLETE the procedure.

## **Step 7: Add Descriptive and Administrative Metadata**

This step in the processing workflow requires you to add descriptive and administrative metadata for ‘items’ that will be deposited into Deep Blue. You will be guided by the deposit plan you developed in Step 6. The procedure will result in a spreadsheet (to batch upload content to Deep Blue), a Dublin Core XML manifest of materials, and a modified EAD record of administrative metadata for the deposit as a whole.

Please note that you may quit at any point (when given the option) and resume at a later date.

Upon initiation of this step, AutoPro will remind you to finish the intellectual arrangement and description that should have begun with the Appraisal, Arrangement, and Packaging of content (see Step 6).

```
This procedure will help archivists prepare content for deposit and long-term preservation in the University of Michigan Library's DSpace repository (Deep Blue).  
NOTE: Before starting this step, archivists MUST already have:  
1. established the intellectual arrangement of the deposit  
2. described content in a new or updated finding aid  
3. shared the finding aid with a division head or Digital Curation  
4. spoken with Nancy or Mike about how content will be deposited and presented in Deep Blue.  
Options:  
-C: CONTINUE; the above requirements have been met!  
-Q: QUIT procedure and return later  
<C / Q> _
```

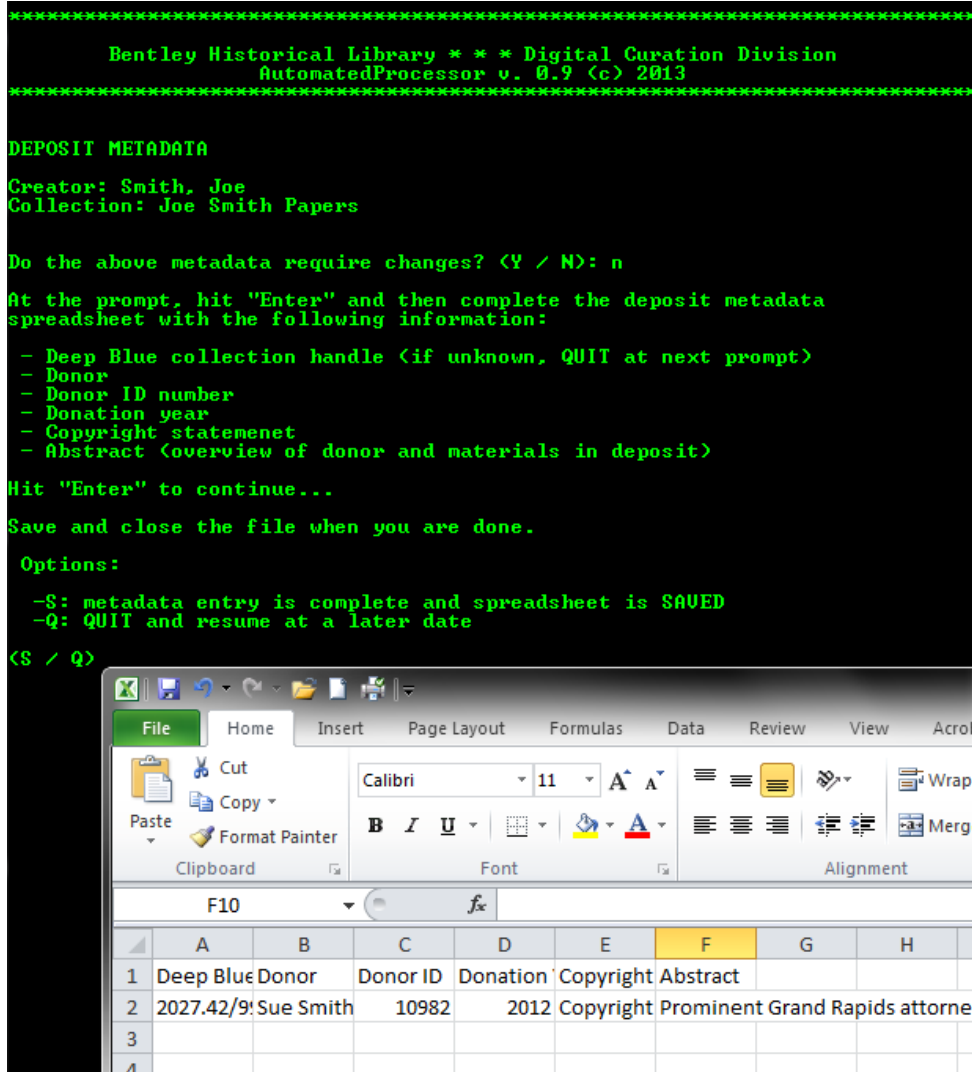
By completing an initial draft of the finding aid or catalog record, you have a thorough understanding of the relationships between different parts of the deposit and, if present, previous accessions to the collection.

### ***Add Deposit-Level Metadata***

AutoPro will require you to enter descriptive and administrative metadata about the deposit in an Excel spreadsheet. This information includes:

- Deep Blue collection handle
- Donor

- Donor ID number
- Year of donation
- Copyright statement
- Abstract describing the creator and the contents of the deposit (not the collection as a whole).



If you return to this procedure after having previously started and quit, you will have the opportunity to edit this deposit-level metadata, if necessary.

### ***Check for Pre-Completed Metadata Spreadsheet***

If this is the first time you are describing the deposit, AutoPro will check to see if you have a pre-completed descriptive metadata spreadsheet.

```
*****
Bentley Historical Library * * * Digital Curation Division
AutomatedProcessor v. 0.7 (c) 2012
*****
Do you already have a metadata spreadsheet for the deposit that conforms to
the requirements for Deep Blue? (Consult with Nancy or Mike if necessary.)
(Y / N)
```

This spreadsheet **must** be completed in accordance with conventions established for depositing content into Deep Blue. Some metadata ‘massaging’ may be required; consult with Nancy or Mike if you have questions. If you have completed this spreadsheet, AutoPro will prompt you to indicate if there are any access restrictions and then move on to technical metadata extraction (see below).

### *Creating a New Item*

The first time you add descriptive metadata to a deposit, you will be immediately taken to the “Main Item Options” screen:

```
*****
Bentley Historical Library * * * Digital Curation Division
AutomatedProcessor v. 0.8 (c) 2013
*****
--> ITEM IDENTIFICATION AND DESCRIPTION
In most cases, an "Item" will consist of a single file. Multiple files may
be associated with an Item only if all the files represent a single coherent
intellectual unit.
NOTE: Be sure to discuss the organization of complex deposits with Digital
Curation.
Options:
-N: create a NEW Item
-Q: QUIT packaging and resume at a later date
(N / Q)
```

From this point, you may either create a NEW Item or QUIT and resume description at a later date.

After you have added metadata to an Item, you will return to this screen and will have additional options to either REVIEW existing items (which in turn will allow you to add additional content to an item or complete the description thereof) or FINALIZE the deposit’s packaging (see below).

### Item-Level Metadata

The first step in creating a new item involves the entry of descriptive metadata. You will first enter the title, which should reflect the intellectual arrangement of the material (i.e. series, sub-series, folder, etc.). AutoPro will prompt you to enter one title component (i.e., a level of the intellectual arrangement) at a time; enter a ‘D’ when have completed entering all components of the title:



```

ITEM METADATA:
This step will allow you to enter descriptive and administrative metadata
for an individual item.
NOTE: The item title should reflect the different sections of the content's
intellectual arrangement (series, subseries, file, etc.).
Enter a single level/section and hit "Enter." Enter D when DONE.
<Title component or D for DONE>: Research Materials
<Title component or D for DONE>: Surveys
<Title component or D for DONE>: d

```

You will then enter additional information about the item, including a description (i.e. scope and content note), the inclusive dates for when the content was created or originally used, the number of years access must be restricted ('0' if the content is open), and the 'content type' (information required by Deep Blue administrators).

```

ITEM METADATA:
DESCRIPTION (no double quote marks or parantheses): Surveys created by Joe Smith as part of his 2010 state senate run.
DATE RANGE (of original creation or use of content): 2009-2010
NOTE: Enter contributors' names using format "Lastname, Firstname" and separate multiple names with semi-colons.
ADDITIONAL CONTRIBUTORS:
NOTE: If the item contains multiple content types, separate them with semi-colons.
Deep Blue CONTENT TYPE(S): Text
NUMBER OF YEARS ITEM MUST BE RESTRICTED (0 if open): 0_

```

After entering this information, you will be given an opportunity to correct any errors.

### Associate Bitstreams with the Item

After the item metadata has been saved, AutoPro will prompt you to enter information about associated files (or 'bitstreams') in a spreadsheet. To identify related files (i.e., the original version as well as its preservation and/or access copy), we use a concept of 'bitstream groups.' All versions of a file belong to the same group; additional files associated with the item will belong to other groups. In addition, a bitstream group may only have one file in it. You will also note the inclusive dates for when the content was originally used/created and provide a brief description. For files in the same bitstream group, the description should clearly identify the relationship between the files:

	A	B	C	D	E	F
1	Full Path to File	Bitstream Group	Date Range	File Description		
2	C:\BHL\unprocess	1	2009	Original version of surveys and surv		
3	C:\BHL\unprocess	1	2009	Preservation copy of surveys and re		
4	C:\BHL\unprocess	2	2010	Follow-up surveys from 2010		

After you indicate that the spreadsheet has been saved, AutoPro will record file sizes, checksums, and mime types and then generate an XML manifest and EAD for the item. You will then be taken to the Main Item Option menu where you will then have an opportunity to create a new item, review other items, quit, or finalize the packaging and description:

```
--> ITEM IDENTIFICATION AND DESCRIPTION

In most cases, an "Item" will consist of a single file. Multiple files may
be associated with an Item only if all the files represent a single coherent
intellectual unit.

NOTE: Be sure to discuss the organization of complex deposits with Digital
Curation.

Options:
-N: create a NEW Item
-R: REVIEW Items created for this deposit
-F: FINALIZE packaging for entire deposit
-Q: QUIT packaging and resume at a later date

<N / R / F / Q> _
```

#### Create an Item for Archived Websites

If you indicated earlier that the deposit includes an archived website, AutoPro will ask if you are describing the web content. If so, you will be prompted to enter the year the website was captured before continuing with additional item-level metadata:

```
ITEM METADATA:

This step will allow you to enter descriptive and administrative metadata
for an individual item.

Is this item an archived website? <Y / N> y
Date website was captured <YYYY>: 2009

NOTE: The item title should reflect the different sections of the content's
intellectual arrangement <series, subseries, file, etc.>.

Enter a single level/section and hit "Enter." Enter D when DONE.

<Title component or D for DONE>:
```

Your description of the item should note the precise date of the capture (if known) as well as important features and functions of the archived website. The description from a School of Natural Resources and Environment website may be used as a template:

Archived version of the School of Natural Resources and Environment website as of April 20, 2004. Documents the academic programs, accomplishments, resources, events, and people at the School of Natural Resources and Environment. Content includes important news and announcements, information for alumni, and content geared towards students, such as course descriptions and funding opportunities.

AutoPro will take this information and create an HTML access page (BHL-start-here.html) that is stored directly within the .ZIP package. This file will contain descriptive and administrative metadata related to the site as well as hyperlinks to individual resources (identified in Step 6).

### **Reviewing an Item**

You have the opportunity to review items when you resume step 7 (after previously quitting) or when you select option *R* (“Review Items created for this deposit”) from the Main Item Options screen. During the review process, AutoPro will display an item ID (generated for internal use by AutoPro), title, description, and status (either “completed” or “NOT completed”) for all the items you have created. Please note that **ALL items must be formally completed** for AutoPro to correctly generate metadata records.

```
*****
Bentley Historical Library * * * Digital Curation Division
AutomatedProcessor v. 0.8 (c) 2013
*****

ITEM REVIEW

ITEM ID no.: 1
TITLE: Meeting Materials - State Republican Party Executive Committee
DESCRIPTION: Meeting materials collected by Samuel Smith during his service on the Michigan state Republican Party's Executive Board
STATUS: Packaging complete

PLEASE NOTE:
1. You may not add files to an Item once it has been finalized.
2. Be sure to "COMPLETE" the packaging of each Item.

Options:

-A: ADD content to an existing Item
-C: COMPLETE packaging of a Item
-U: VIEW contents of Item

-N: create a NEW Item
-F: FINALIZE packaging for entire deposit
-Q: QUIT and resume packaging content at a later date

<A / C / U / F / N / Q>
```

From this screen you may ADD additional content to an existing item (if **not** complete), VIEW the contents of an item, or COMPLETE the packaging and description of an item. When you select any of these options, AutoPro will ask for the item ID number:

```
At the prompt, enter the appropriate Item ID number; do not include any extra spaces.
1_
```

- If you elect to ADD content, you will follow instruction outlined in the section “Associate Bitstreams with the Item”
- If you elect to view content, AutoPro will display the filename and description for all content associated with the item before returning you to the “Item Review” screen.

```

*****
Bentley Historical Library * * * Digital Curation Division
AutomatedProcessor v. 0.8 (c) 2013
*****
VIEW FILES ASSOCIATED WITH ITEM
ITEM: Meeting Materials - State Republican Party Executive Committee
      FILENAME: january-2011.zip
      DESCRIPTION: Michigan Republican Party executive committee meeting materials from January 2012
      FILENAME: June-2011.zip
      DESCRIPTION: MEETING MATERIALS FROM JUNE 2011
Press any key to continue . . . -

```

- If you elect to complete packaging and description of the item, AutoPro will record file sizes, checksums, and mime types and then generate an XML manifest and EAD for the item. You will then be taken to the Main Item Option menu where you will then have an opportunity to create a new item, review other items, quit, or finalize the packaging and description

From the “Review Item” menu, you may also create a NEW item, FINALIZE the deposit, or QUIT and resume at a later date.

### ***Finalize Description and Technical Metadata Extraction***

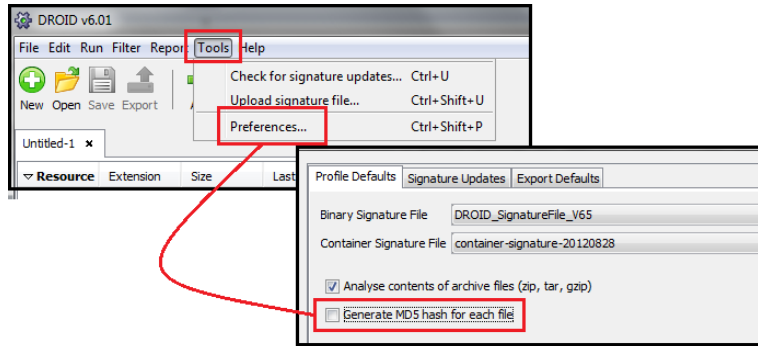
When you enter option F (FINALIZE) from the Main Item Option menu, AutoPro will complete the generation of metadata files and then prepare to extract technical metadata from material with DROID. This application will also generate MD5 checksums for content (including materials within .ZIP files); you will thus be prompted to confirm that DROID is correctly configured:

```

*****
Bentley Historical Library * * * Digital Curation Division
AutomatedProcessor v. 0.9 (c) 2013
*****
TECHNICAL METADATA EXTRACTION
This step requires you to configure DROID so that it will generate MD5
checksums.
Is DROID configured to generate MD5 checksums? <Y / N / QUIT>

```

If you don’t know, answer No so that you can be sure that this feature is enabled. After DROID opens, it will check for any updates; if any are available, approve their installation. Next, click on the *Tools* item in the top navigation menu and select *Preferences*. Make sure that the box next to “Generate MD5 hash for each file” is checked. Click OK and then close the DROID window.



After you confirm the correct settings, DROID will run and produce a report; this step may take a long time for large collections. DROID also has a tendency to ‘hang’ on very small deposits—contact Mike if this occurs. (You may need to enter CTRL-C to interrupt the procedure and then answer ‘N’ when you are asked if you want to terminate the batch process).

### Step 8: Transfer to Long-Term Storage

This and the following workflow steps will only be completed by Mike, Nancy, or other staff with access to the Deep Blue deposit folder and the BHL Dark Archive. Be sure that you are connected to the appropriate repositories and logged in to the UMROOT domain with your Windows AD password before beginning!

#### *Transfer to Deep Blue Deposit Folder*

Upon initiation of this step, AutoPro will check if the materials will be deposited in Deep Blue; if so, AutoPro will also ask if the deposit requires additional email processing (via procedures developed by MLibrary LIT). Finally, you will enter the drive letter where the \deepblue\_deposits\ folder lives:

```

Initiated: 8. TRANSFER to LONG-TERM STORAGE
*****
This procedure will transfer content to the BHL Dark Archive and, if
necessary, to the \deepblue_deposits\ folder in preparation for deposit to
DSpace.

Will content be deposited in Deep Blue? <Y / N> y
Is this a dedicated email deposit? <Y / N> n
At the prompt, enter the drive letter where the "deepblue_deposits" folder
is located. Do NOT enter a colon.

Options:
  -QUIT: QUIT and resume transfer at later time
  -Drive letter (i.e. X, Y, or Z, etc.)

<QUIT / Drive letter> Z

```

AutoPro will then check to see what drive letter the dark archive is associated with:

```

*****
      Bentley Historical Library * * * Digital Curation Division
      AutomatedProcessor v. 0.9 (c) 2013
*****

TRANSFER TO DARK ARCHIVE

NOTE: This procedure may only be completed by staff with full access to the
      bhl-root and bhl-archive storage locations.

At the prompt, enter the drive letter to which "bhl-archive" (the BHL Dark
Archive) has been mapped (i.e. Z or Y). Do NOT include a colon.

Options:
  -QUIT: QUIT and resume transfer at another date
  -Drive letter (i.e. X, Y, or Z, etc.)

<QUIT / Drive letter> Y

```

AutoPro will then package the deposit according to the BagIt specification and then move the material to the dark archive, where the bag will be validated. Please note that large deposits may take a long time.

### Step 9: Clean Up

In the final step of the digital processing workflow, AutoPro will delete the processing directory, temporary files, and logs with an additional option of deleting separations.

```

Initiated: 9. CLEAN UP
*****

This procedure will clean the processing directory after content has been
successfully transferred to Deep Blue and/or the BHL Dark Archive.

BE SURE TO CLOSE ANY WINDOWS EXPLORER WINDOWS OR APPLICATIONS THAT MAY BE
USING FILES IN THE TARGETED FOLDERS.

The procedure will package log files and delete the following directories:
  - C:\BHL\unprocessed\98567_0001
  - C:\BHL\temp\98567_0001
  - C:\BHL\logs\98567_0001

Delete Separations directory, too? <Y / N> Y_

```

After you make your selection, AutoPro will package the log files so that a copy may be deposited in the Bentley's IFS space for quick reference.

```

TRANSFER LOG FILES
Upload C:\BHL\logs\98567_0001\0001.zip to BHL IFS space:
--> ../bhl/Private/html/processing_log_files/98567/
NOTE: Consult with Mike if problems arise.

Enter "S" when 0001.zip has been successfully uploaded. S

```

Once you have indicated that the .ZIP file of logs has been uploaded, AutoPro will request a final verification that you are ready to delete the working copies of materials.

```
CLEAN UP
- C:\BHL\unprocessed\98567_0001
- C:\BHL\tmp\98567_0001
- C:\BHL\logs\98567_0001
- C:\BHL\separations\98567_0001

Options:
-D: DELETE folders
-Q: QUIT procedure and perform clean up at a later date

<D / Q> D_
```

If you prefer to wait, you may quit the program and return at a later time. Following the deletion of content, the digital processing workflow is concluded; you may press any key and AutoPro will close.

```
*****
Bentley Historical Library * * * Digital Curation Division
AutomatedProcessor v. 0.7 (c) 2012
*****

You have completed processing deposit 87206_0001 with AutoPro, the
Bentley Historical Library's Automated Processor

Good bye!

*****

Press any key to continue . . . _
```

## 5. Version History

<b>Version</b>	<b>Date</b>	<b>Prepared By</b>
3.0	2013-10-29	Michael Shallcross
2.0	2013-04-09	Michael Shallcross
1.0	2012-10-29	Michael Shallcross