```
*******************************************************************
*******************************************************************

   _\\\\\\\\\\\\_   _\\\\_   _ _ _
  *\\\\\\\\\\\\\\\\*  *\\\\*  *\\\\*  *\\\\*
  *\\\\//  \\\\\\*  *\\\\*  *\\\\*  *\\\\*
  *\\\\*    *\\\\*  *\\\\*  *\\\\*  *\\\\*
  *\\\\\_ _ ////*   *\\\\__\_  *\\\\*
  *\\\\\\\\\\\\\\*  *\\\\\\\\\\\  *\\\\*
  *\\\\/    \\\\*    *\\\\*  *\\\\*  *\\\\*
  *\\\\*    *\\\\*  *\\\\*  *\\\\*  *\\\\*_____
  *\\\\\_ _ ////*   *\\\\*  *\\\\*  *\\\\\\\\\\\*
  *\\\\\\\\\\\\\/   *\\\\*  *\\\\*  *\\\\\\\\\\\\*
  *\\\\\\\\\\\\\/   *\\\\*  *\\\\*  *\\\\\\\\\\\\*

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

*******************************************************************
*******************************************************************
```
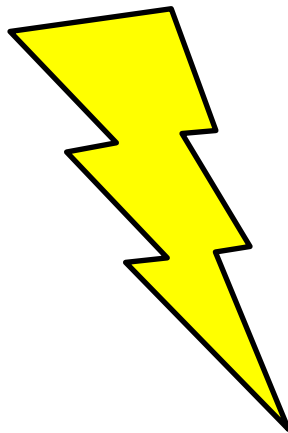
BENTLEY HISTORICAL LIBRARY

PROCESSING HYBRID AND DIGITAL COLLECTIONS

An AutoPro User Manual and General Reference
Version 1.1

Prepared by Michael Shallcross, Division of Digital Curation
Reviewed by Nancy Deromedi, Head, Division of Digital Curation

1. Introduction

1.1 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 scripts that automates key steps in preparing digital content for long-term preservation and access.

1.2 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 fulfill two goals: (1) to make digital processing more efficient by automating key workflow steps and (2) to reduce technical barriers and thereby permit archivists to focus their energies on the traditional archival functions of appraisal, arrangement, and description.

1.3 Overview of the Automated Processing Workflow

AutoPro is comprised of 28 Windows CMD.EXE shell scripts that move content through a 12 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. A brief overview of the workflow (with information on component applications) follows below.

Immediately after content is accessioned and deposited in the Bentley Library's interim repository (a secure Windows file server), AutoPro runs a virus scan (the University of Michigan employs Microsoft Forefront Endpoint Protection on all work stations) and creates a working backup so the SIP can be restored in case of an error or accidental data loss.[1]

AutoPro then searches 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.[2] 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. Please note that some disk image formats (.AFF, .AD1, .E01, etc.) will require the processor to employ FTK Imager to manually extract content.[3]

At this stage, AutoPro uses Bulk Rename Command to replace spaces and non-alphanumeric characters in folder and file names with hyphens and also record the original and new names in a log file.[4] AutoPro then runs DROID to search for files with missing or mismatched extensions, identifies correct extensions with the TrID File Identifier utility, and then verifies these results with the PRONOM format registry.[5] AutoPro preserves the TrID output (which includes a report on likely file types, based upon the target file's binary signature) in a log file, along with the original and new filenames.

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.[6] AutoPro searches for these at risk formats (based upon extension) and then employs

---

[1] 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.
[2] 7-Zip is an open source file archiving application. For more information see http://www.7-zip.org/.
[3] 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.
[4] Bulk Rename Command is a freely distributed file renaming tool. For more information, see http://www.bulkrenameutility.co.uk/Command.php.
[5] 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.
[6] 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

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 encoding), Aid4Mail (various email formats to .MBOX), and Microsoft Office File Converter (Office files to Open Office XML).[7] 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.[8] 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.[9] 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).

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 identify series and provide descriptive and administrative metadata about the materials. This step produces a Dublin Core XML file and Excel spreadsheet used to deposit material in Deep Blue,

---

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.
[7] 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 Convertor (http://www.microsoft.com/en-us/download/details.aspx?id=11454) is part of the freely available Office Migration Planning Manager.
[8] Identity Finder Data Loss Prevention (DLP) Endpoint (http://www.identityfinder.com/us/Business/IdentityFinder/EnterpriseClient) is proprietary software that can identify potentially sensitive information..
[9] 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.

the University of Michigan's DSpace repository.[10] Finally, AutoPro employs BagIt to transfer a copy of all material (and metadata) to a secure dark archives.[11] 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.

1.4 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.
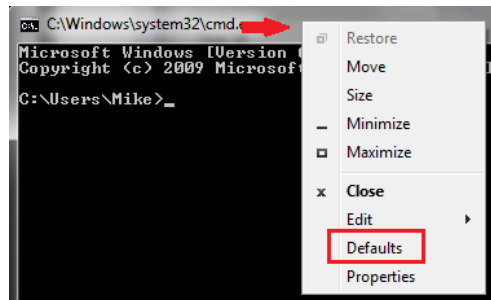
When the "Console Windows Properties" window opens, make sure that the boxes for the following items are checked under the "Options" tab:

---
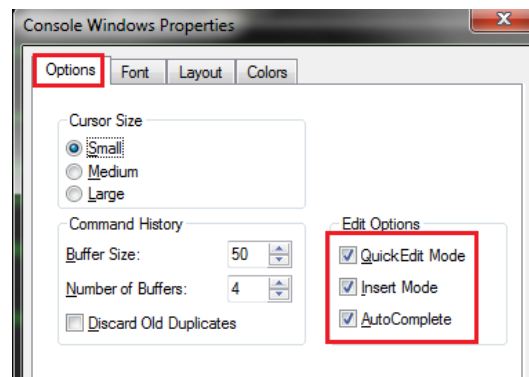
[10] Explore the Bentley Historical Library's archival community in Deep Blue at:
http://deepblue.lib.umich.edu/handle/2027.42/65133.
[11] BagIt is part of an open source set of transfer tools developed by the Library of Congress. For more information, see http://sourceforge.net/projects/loc-xferutils/.

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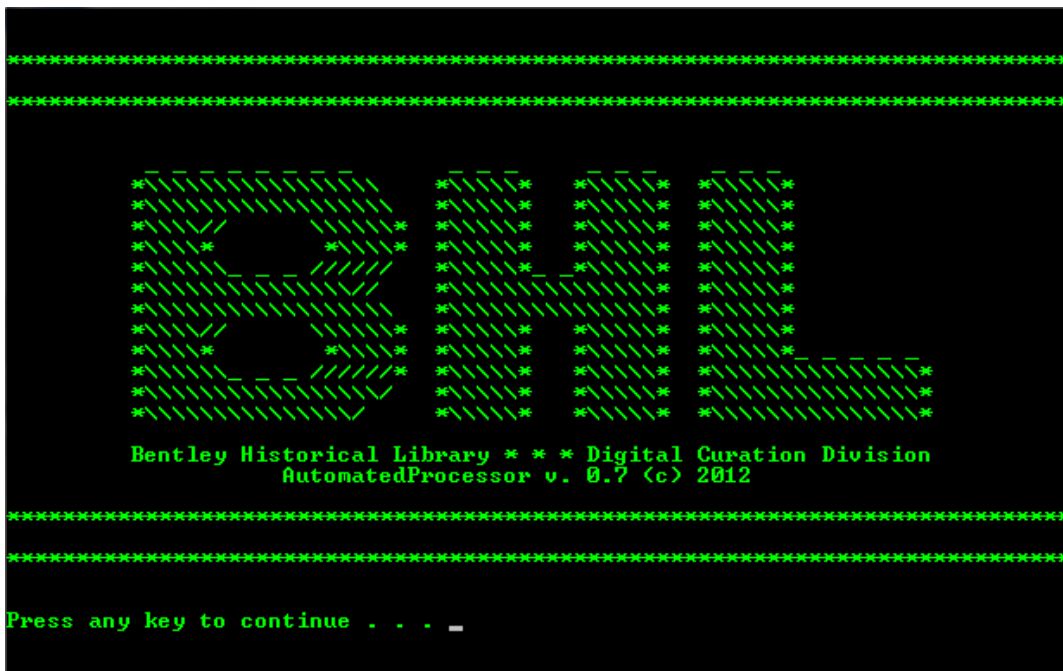
2.1 Launching AutoPro

Go to the 'Unprocessed' directory and locate the appropriate directory.

1. Right-click on the processing directory (i.e. \87209_0001\).
2. Select *AutoPro.bat* from the "Send to" context menu.
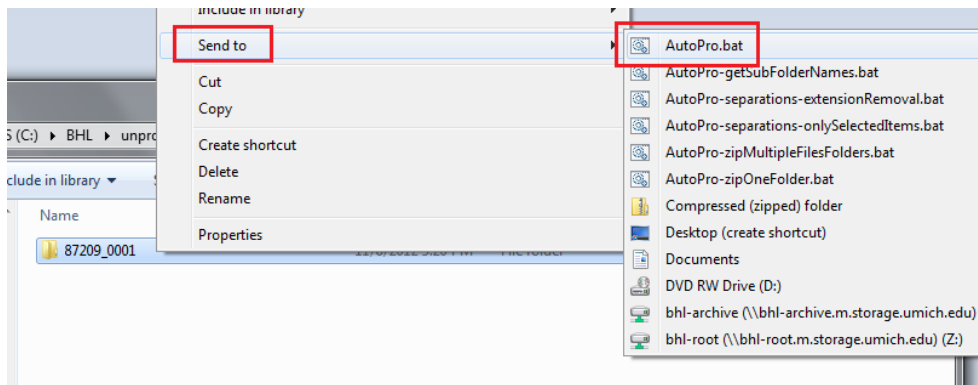
The AutoPro application will now open; hit any key to proceed.

2.2 Logging In

You will now be required to confirm the processing directory and enter your name so that an audit trail may be generated for the processing session.

2.2.1 Directory confirmation

In addition to making sure that you have selected the right processing directory, verify that the Collection ID for the deposit is correct. This ID will be the string of numbers prior to the underscore in the BHL digital deposit number (i.e. the processing directory's name). For example, deposit number 87209_0001 has a Collection ID of 87209.

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 aid to make sure that the Collection ID is the same. The Collection ID will be the first string of digits in the Call Number. **This simple step can save a lot of time and energy down the line!**

Enter *Y* if the processing directory (and Collection ID) is correct, *N* if a different directory needs to be selected, or *Q* to quit the application.

2.2.2 Identification of Processing Archivist

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

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

PROCESSING ARCHIVIST

At the prompt, enter your name (Last name, first name--i.e. Doe, Jane).

Shallcross, Michael
```

After entering your name, you will be asked if it is correct. Enter *Y* to proceed or *N* to re-enter your name.

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

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)
```

Once the processing directory 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 of the content.

3. Main Menu—Selecting 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 twelve steps in the AutoPro workflow. It is essential that the steps be completed in the order provided.

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, you will be asked to verify your choice. Enter *A* to ADVANCE to the procedure, *M* to return to the MAIN MENU or *Q* to QUIT AutoPro.
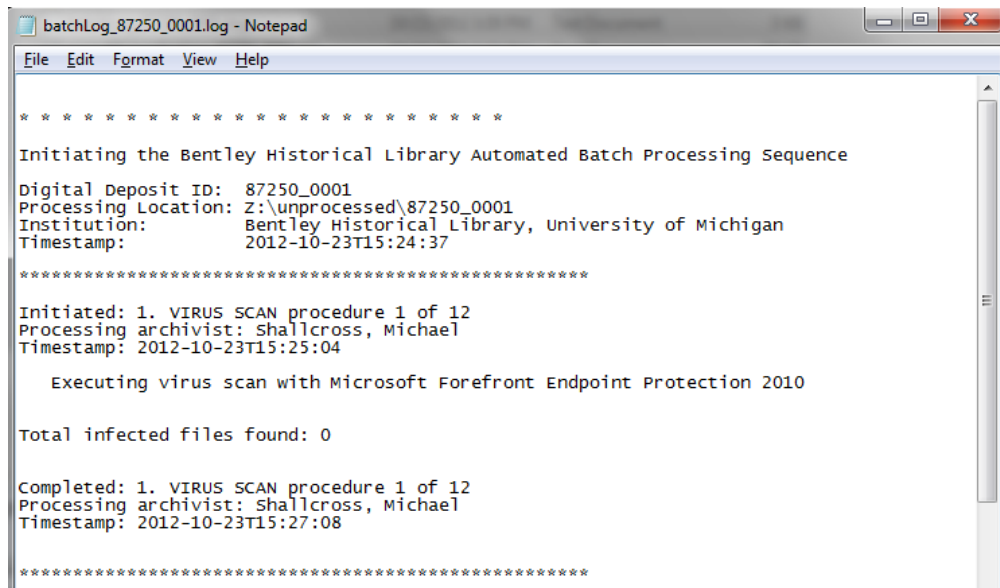
You will receive a similar prompt at the conclusion of each procedure in the AutoPro workflow. If you choose to quit, you may resume processing with the immediate next workflow step at a later date.

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	eventIdentifierV	eventDateTime	eventDetail	eventOutcome	linkingAgent	linkingAg	linkingAg	linkingAg	linkingAgent	linkingAgentRole	
2	virus scan	UUID	d76cfe99-9706-4f	2012-07-27T15:50:2	identification a	action complete	MARC21 Cod	MiU-H	Executor	tool	Microsoft Ant	software	
3	decompression of file	UUID	84c36d87-d675-45	2012-07-27T15:54:3	decompression	action complete	MARC21 Cod	MiU-H	Executor	tool	7-Zip 9.20	software	
4	personally identifiabl	UUID	f119bf41-1b56-47	2012-07-27T15:58:2	identify and re	action complete	MARC21 Cod	MiU-H	Executor	tool	Identity Finde	software	
5	message digest calcul	UUID	82262b07-9155-4c	2012-07-27T16:09:3	calculation of N	action complete	MARC21 Cod	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).

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.

4.1 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.

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

4.2 Backup Content

AutoPro creates a working backup copy of all content in the processing directory so that you have a restore point in the event that an irrevocable error or data loss occurs at some stage of the workflow. If the backup encounters any errors, AutoPro will alert you—please consult with Mike to resolve the issue.

If content in the processing directory exceeds 20 GB, you will receive an error message (see below).

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

Initiated: 2. BACKUP CONTENT

********************************************************************


This procedure will create a temporary backup copy of the processing
directory that will be deleted at the end of the AutoPro workflow.

   DIRECTORY IS TOO LARGE FOR BHL BACKUP AREA.

      --> Maintain backup copy on original media or in alternate location.

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

To conserve hard drive and interim repository space, we will need to make a backup of the content in another location (such as external hard drive, local desktop, or other location). Please consult with Mike or Nancy as to the appropriate place to backup content in such cases.

4.3 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.

4.3.1 Disk Image Extraction

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 section 4.3.2 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 Digial 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)
```
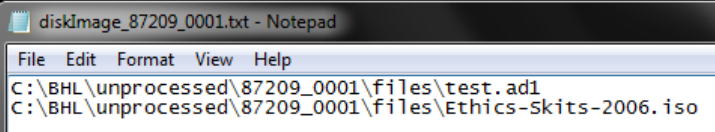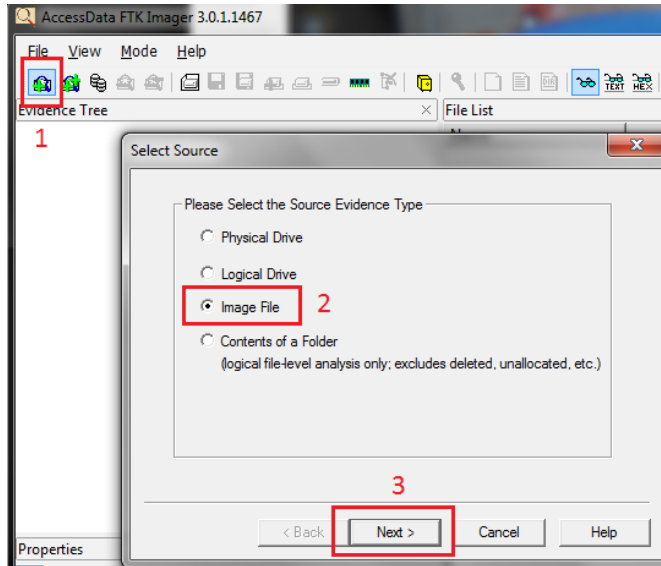
diskImage_87209_0001.txt - Notepad

File Edit Format View Help

C:\BHL\unprocessed\87209_0001\files\test.ad1
C:\BHL\unprocessed\87209_0001\files\Ethics-Skits-2006.iso

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."
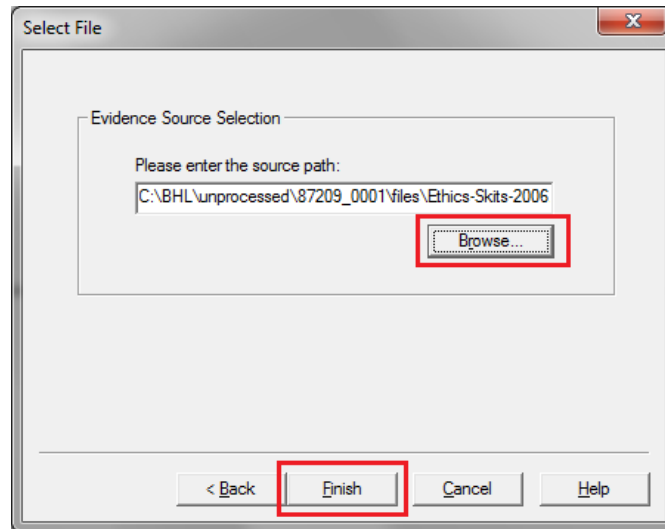

FTK Imager will now show the disk image in the upper-left "Evidence Tree" panel. Expand the disk image file (by clicking on the '+' next to its name), right click on directory immediately within, and select the "Export Files…" option.
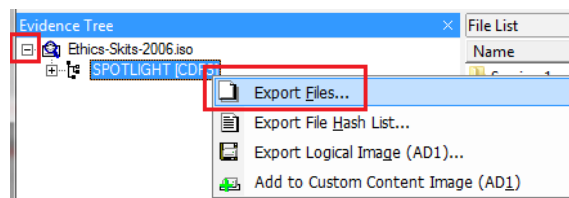
You will then be prompted to select a destination for the extracted files; to maintain original order, choose the same directory that holds the original disk image file.
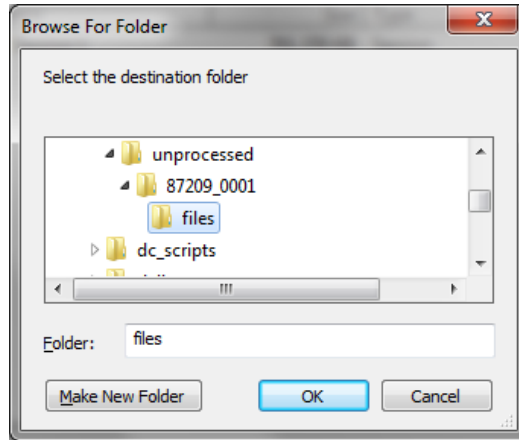
FTK Imager will immediately begin to extract files from the disk image:
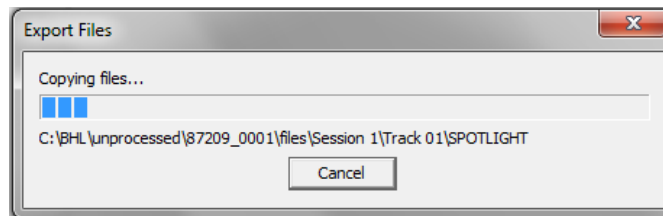
When the operation is successfully completed, you will receive a notice:
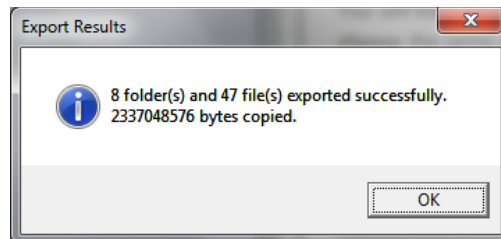
You may now remove the disk image (or "evidence item") from FTK Imager; select the image file and either click the "remove evidence item" from the toolbar or right-click and select the option from the menu. You are now ready to proceed to the next disk image!
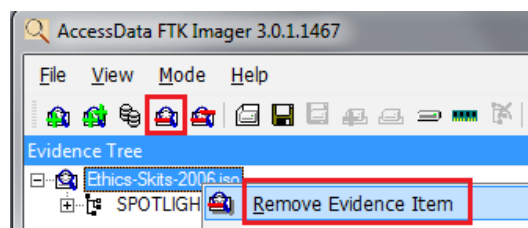
Once you are done extracting content from the disk image file(s), you may enter *C* into the AutoPro interface to COMPLETE this step. If you need to quit and resume file extraction at a later date, enter *Q* and recommence this procedure the next time you log into AutoPro.

```
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.

4.3.2 Archive File Extraction

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.

4.4 Name Normalization

AutoPro will normalize characters in folder and file names with the Bulk Rename Command to avoid issues that may arise in the processing workflow or access portals as the result of spaces or restricted characters. Therefore, all spaces and non-alphanumeric characters will be replaced with hyphens and accented characters will be replaced with non-accented characters. Name collisions will be avoided by the addition of a unique integer at the end of the filename.

Before running this procedure, **be sure to close any Explorer windows or other applications that might be using a file or folder** as these will prevent items from being renamed!

AutoPro will identify any files or folders that cannot be renamed and provide you with another chance for name normalization. If there are still issues after two attempts please consult with Mike.
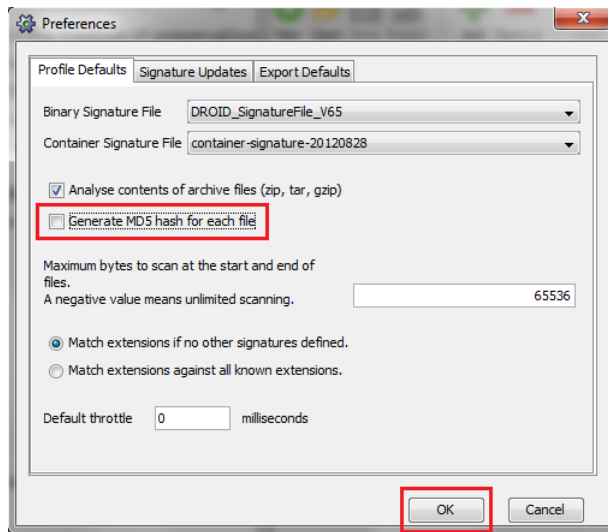
4.5 Identify Missing File Extensions

To aid in the creation of preservation copies and facilitate the eventual use of content by patrons, AutoPro will identify missing or incorrect file extensions. Upon initiation of this procedure, AutoPro will launch DROID and provide instructions to adjust the application's settings so that MD5 checksums are **not** calculated at this point. It is *very* important to do so! 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. 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*.
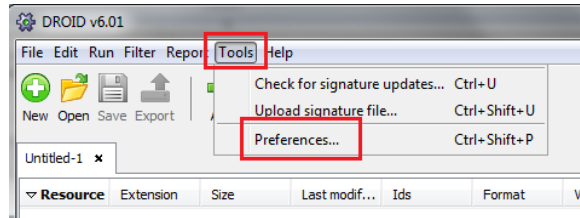

Make sure that the box next to "Generate MD5 hash for each file" is **NOT** checked. Click OK and then close the DROID window.


You will then be prompted to enter *S* once 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. Each of these files will be run through the TrID file identification utility, which uses binary signatures to identify appropriate extensions. The results of this operation are then cross-referenced with the PRONOM file format registry. If there are no conflicts, the new extension stands and the results are recorded in a log file. If TrID has misidentified a file, you will have the opportunity to manually identify its extension:

If you choose to manually identify the remaining files, a text file listing the full path, PRONOM unique identifier, and mime type for each file will open.
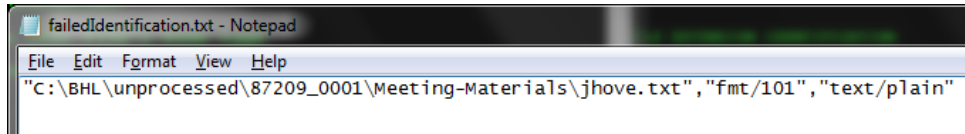
You may delete any files that do not need to be renamed and then hit *S* once the file has been SAVED.
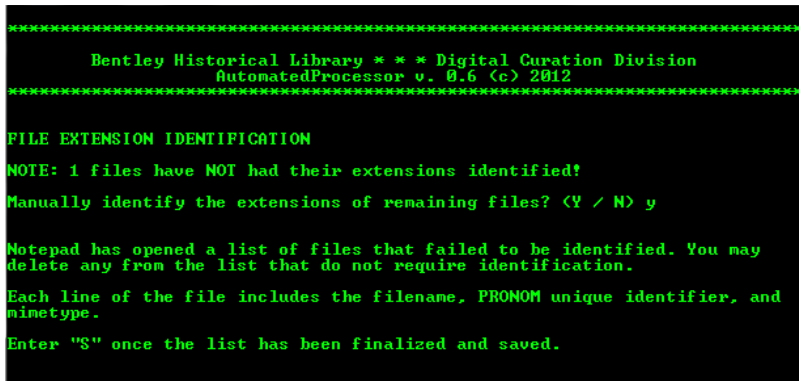
AutoPro will use TrID to make a preliminary identification and then consult the PRONOM file format registry to identify extensions associated with the file's PRONOM unique identifier (PUID).
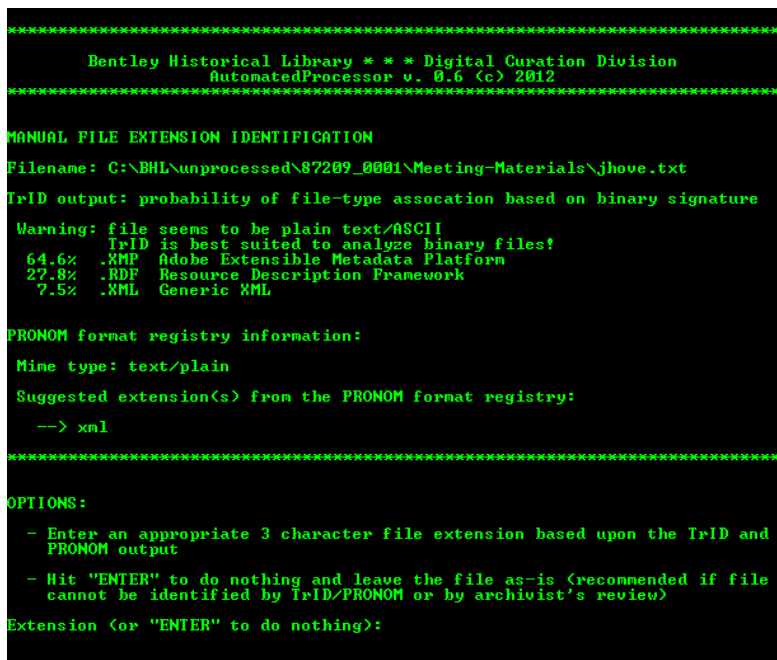
Based upon the results of the different resources (and, if necessary, your review of the file, with Quick View Plus) you may adopt one of the proposed extensions. Enter the three-digit extension without a period; AutoPro will apply it to the file and record the operation in a log file. If you are unsure of what the extension should be, simply hit "Enter"; AutoPro will leave the file unchanged and proceed to the next one on its list.

4.6 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!

4.7 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:

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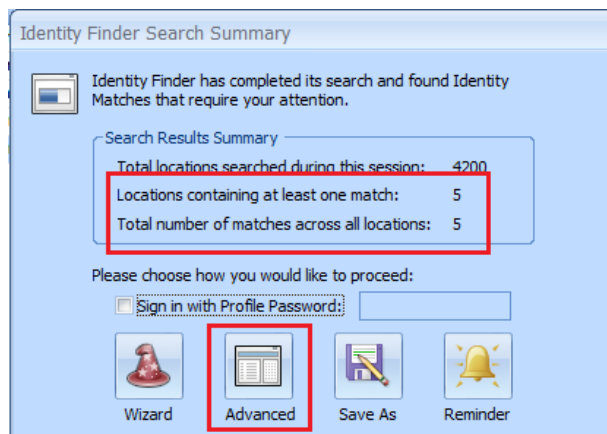
For no matches, you will simply enter *N* at the AutoPro prompt; if there are matches, enter *Y*, click the "Advanced" option and then follow the step-by-step instructions in AutoPro.

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*.
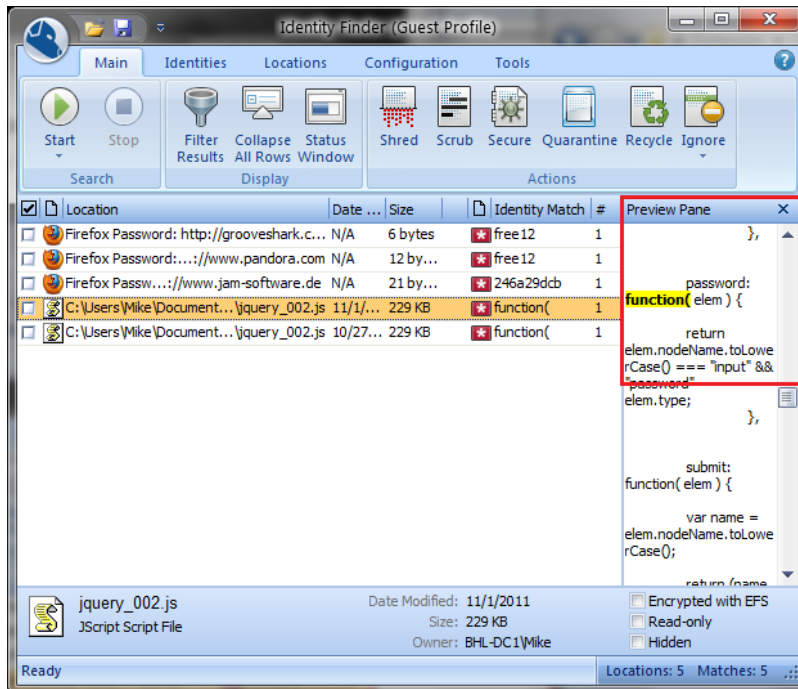
There are three possible actions you can take with content that has been found to contain an identity match:
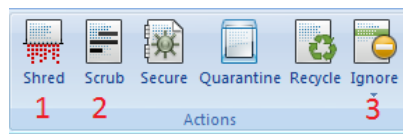
1. If the content is in non-unique or routine business documents (such as business expense reports or P-Card logs), you may use the "SHRED" option to securely delete content. Check the boxes next to the target filenames, click the "Shred" action, and then confirm your choice. Please note that this option 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.
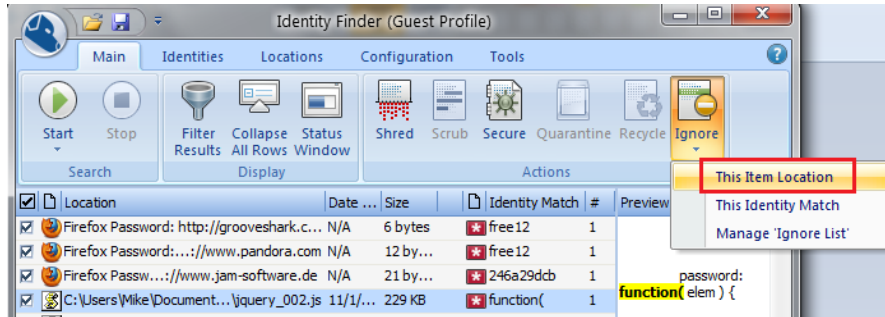
Finally, if the identity match is in a significant record that must be preserved but cannot be redacted, it may be necessary to impose an access restriction. Please consult with Nancy, Mike or a division head to determine the appropriate restriction.

4.8 Appraisal and Arrangement

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.

At the beginning of this procedure, AutoPro will display detailed instructions a separate CMD.EXE window in regards to the in-depth appraisal and review of content, separations, arrangement, and packaging of files.

After reading through the instructions, you may hit "Enter" and then choose from the following options to complete this procedure:

4.8.1 In-Depth Review of Content

AutoPro has two sub-menus of options to facilitate the appraisal and review of digital materials.

4.8.1.1 The "Characterize Deposit" Menu
Items within the "Characterize Deposit" employ TreeSize Professional and Windows utilities to provide visualizations of various aspects of the deposit.

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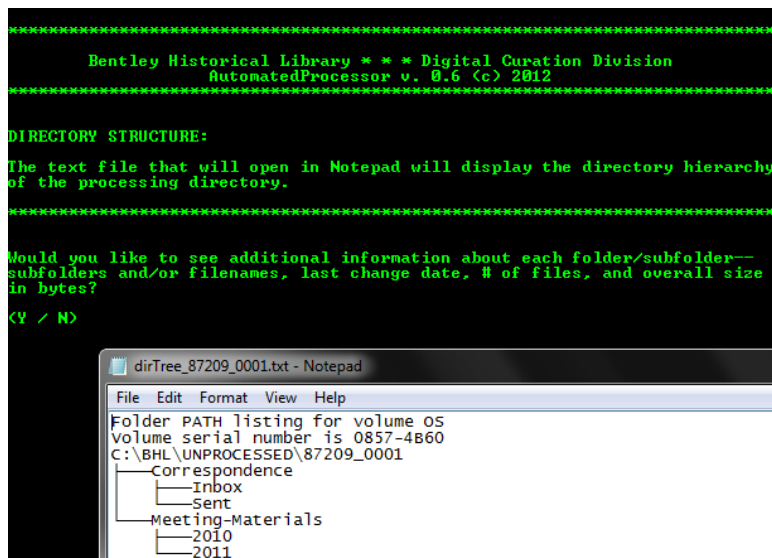
Reviewing the directory hierarchy may aid in the appraisal of the deposit and help you to understand the breadth and scope of content transferred to the library. Folder naming conventions and directory structure may also reveal organizational principles (if any) and areas of particular interest.

4.8.1.1.2 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.
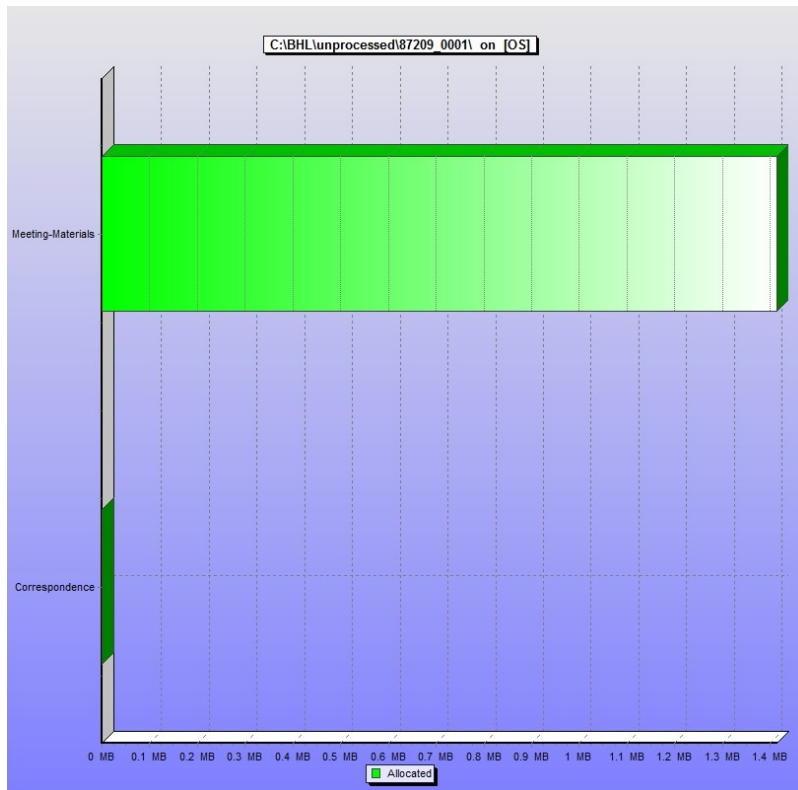
To examine the relative size of the contents of subfolders, simply complete the folder path in the AutoPro interface or enter *C* to complete the step.


4.8.1.1.3 File Extensions

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	Extension	Size	Allocated	Percent	Files	Description	
2	**Container Files**	**241.4 MB**	**241.4 MB**	**51.2%**	**3**	**Compressed Archives and disk images**	
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	**Mail Files**	**153.0 MB**	**153.0 MB**	**32.5%**	**18**	**Email messages and files of email clients**	
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	**Graphic Files**	**39.2 MB**	**39.2 MB**	**8.3%**	**8**	**Files containing pictures, images or mouse cursors**	
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	

4.8.1.1.4 Age of Files

The "Age of files" option generates a bar graph representation of the age range for content, identifying material produced within the past year, from one to 5 years ago, from five to ten years, and then older than 10. This information can be helpful in describing content in the finding aid and scope and content notes.
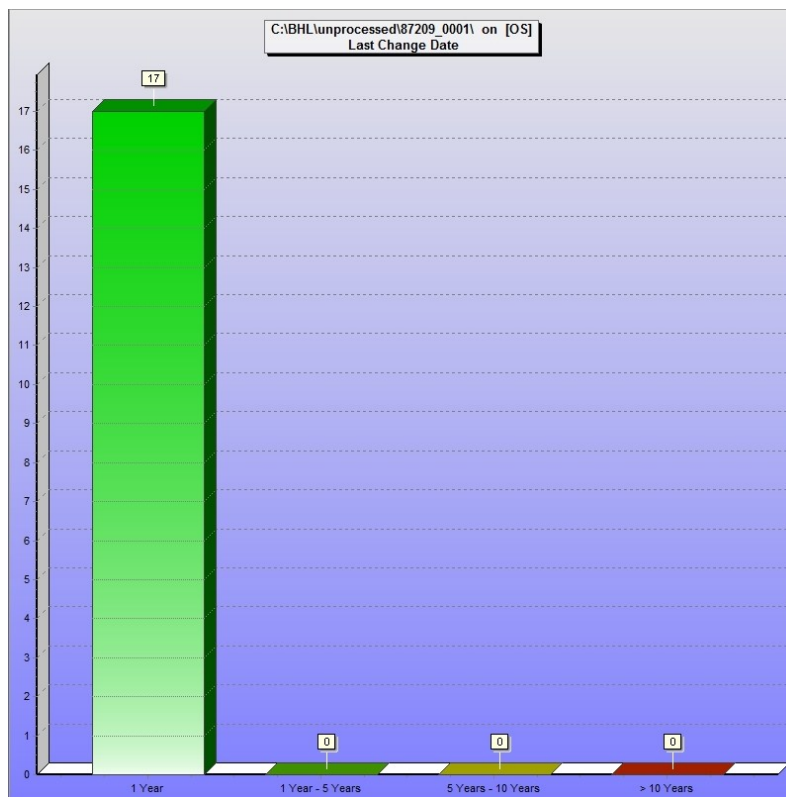
4.8.1.1.5 Duplicate content

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\fido\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 AM
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 I	18.3 KB	10/3/2012 1:06 PM	10/28/2011 1:37 PM
11	Agenda_for_News_an	(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 I	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.

4.8.1.2 The "Review Content" Menu

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.

4.8.1.2.1 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.
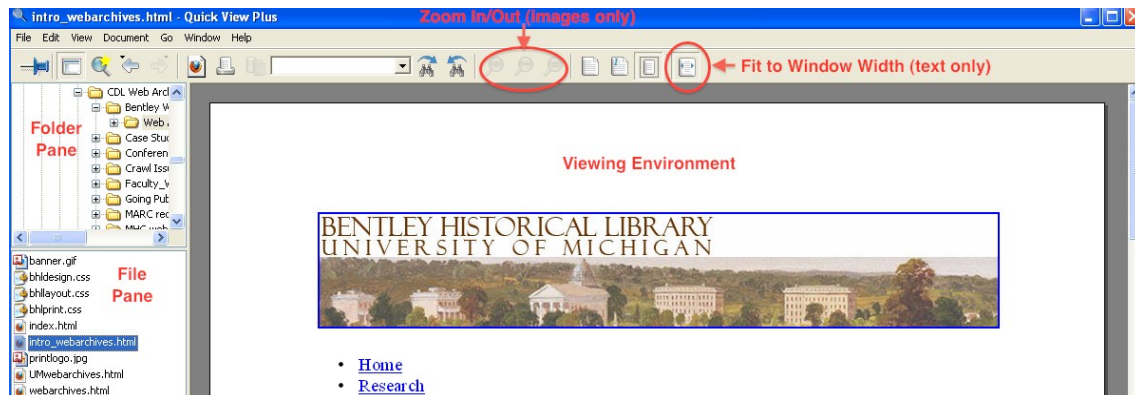
While viewing a file, you may search for text in the ribbon's text box, zoom in/out (for images only) and adjust the document size so that it fits entirely in the view are or is fit to the area's width.

4.8.1.2.2 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.
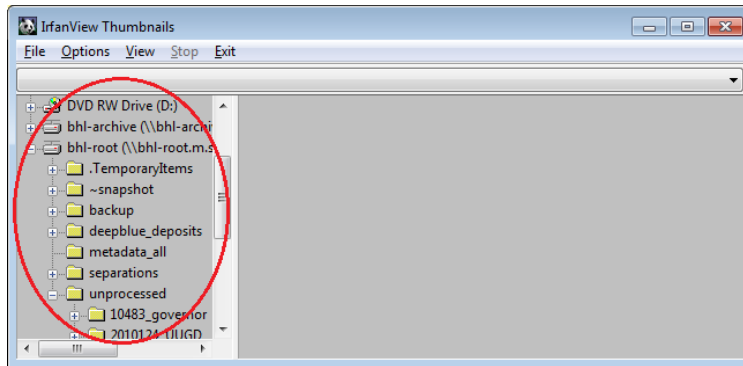

Once you have selected a folder with images, IrfanView will automatically load thumbnails of all image files into the viewing pane. You may then click on a thumbnail to take a closer look at the full-scale or use the navigation pane to browse to other folders.
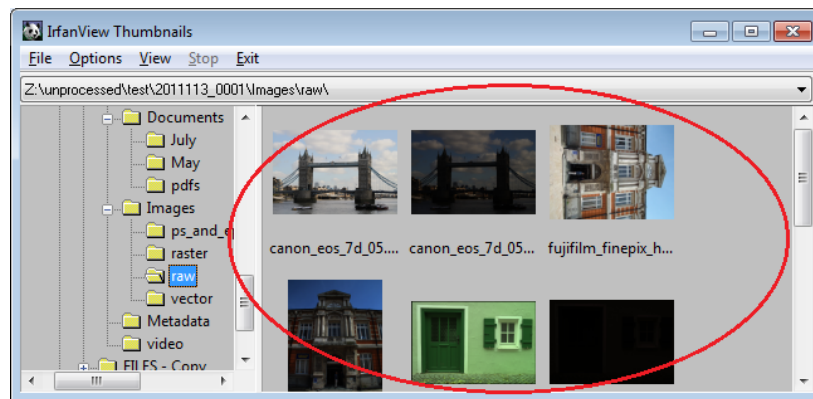
4.8.1.2.3 Inkscape:
Inkscape may be used to view vector image files (such as .SVG, .AI, .WMF, etc.) that cannot be opened in IrfanView of Quick View Plus. AutoPro will open Inkscape and you will then need to click on the open file icon (or use *File > Open* in the navigation menu) to access vector image files.
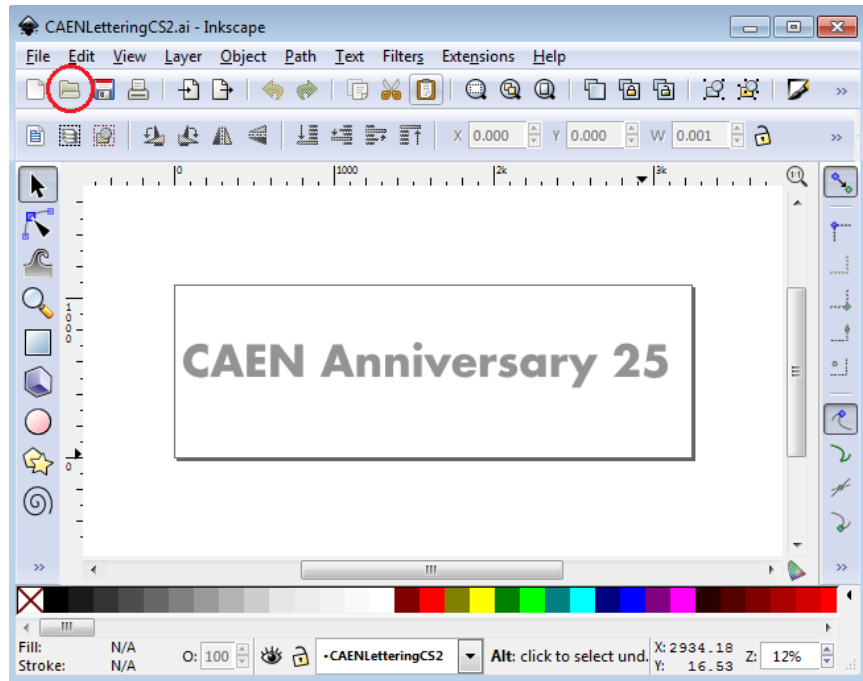
4.8.1.2.4 VLC Media Player:

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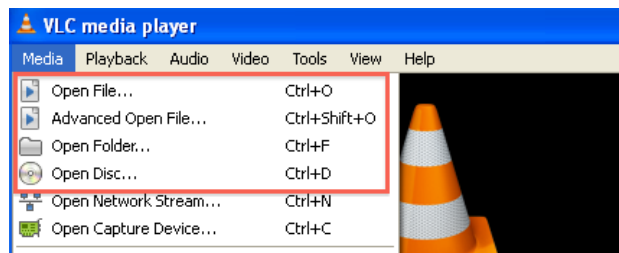
Play controls are located at the bottom of the Media Player window; in addition to *Play*, *Pause*, and *Stop* buttons, the archivist may fast forward or reverse progress by adjusting the slider on the progress bar.

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

4.8.2 Separations

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.

4.8.2.1 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:
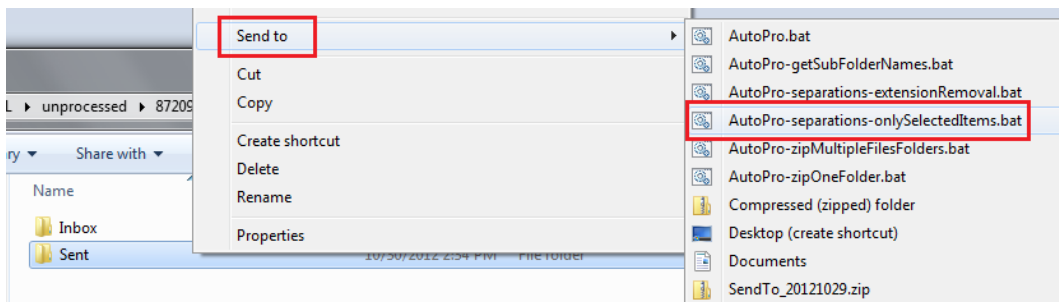
This option may also be used on multiple files and/or folders, selected by clicking on the left mouse button while holding down the "Shift" key (to separate an entire range of files) or the "Ctrl" key (to choose a select number of files and/or folders for separation):
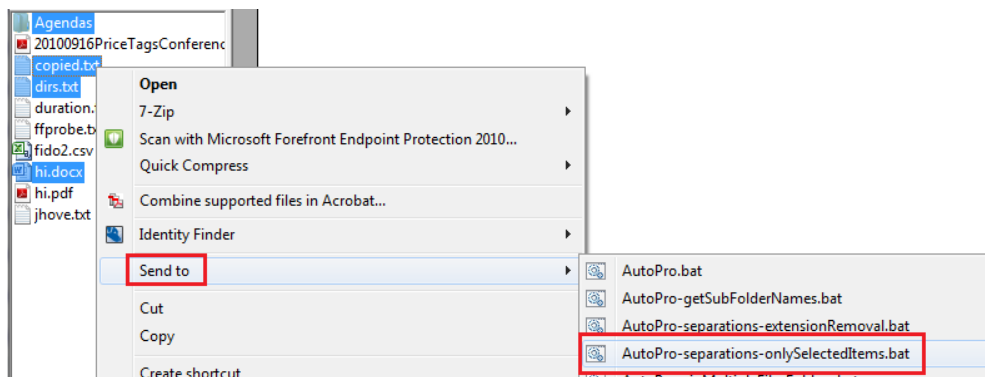
Please note that when selecting multiple files and/or folders for separation, all content must reside in the same parent directory.

4.8.2.2 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.
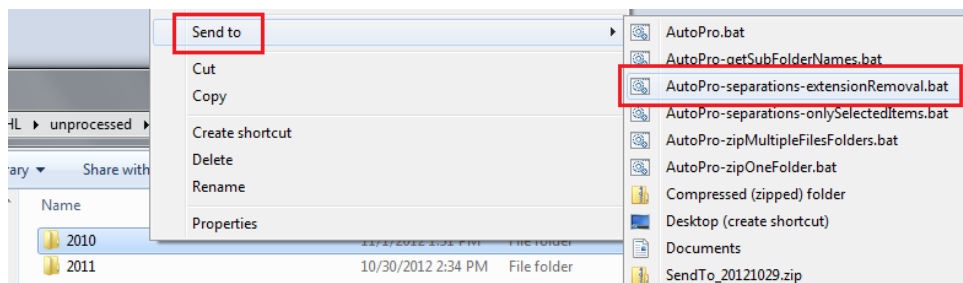
A new CMD.EXE window will open and prompt you to enter the extension for the files you wish to remove. Be sure to enter the appropriate extension (as used in Windows) and to precede it with a period (i.e. ".ini").

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.

4.8.3 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 are analogous to series. You may therefore need to create a top-level folder in order to organize files (or subfolders) into series. It may also be necessary to impose some organizational principle if the files were copied pell-mell from their source location with no structure. All arrangement must be manually performed via a Windows Explorer window. If you create a folder avoid using spaces and/or non-alphanumeric characters in the names (i.e. "John_Doe_Audio_Files").

Be sure to complete the intellectual arrangement before you move any files! Once you have identified series and more precise levels of arrangement, you will have a defined strategy of how to 'physically' arrange material. Furthermore, by producing a draft of how the digital deposit will be represented in a finding aid, you will be able to complete Procedure 9: Add Descriptive Metadata in a more efficient manner.

AutoPro produces a record of the initial arrangement the first time that the application is run on a digital deposit. It will create a final record of arrangement at the close of this procedure—be sure that you have completed all separations and packaging before completing the procedure.

4.8.4 Packaging

To simplify the description, management, and dissemination of content, the Bentley Library will enclose entire directories or selected files and folders in uncompressed .ZIP files. In some cases (i.e. where the deposit is small or the individual files are extremely large or significant), the content will be described and deposited at the item level; in others cases, however, it will be necessary to package.

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 how the content will be deposited and presented in Deep Blue.

To ensure that we do things right the *first* time, there will be **no exceptions** to the above!

Ideally, you will only package content in the top two levels of processing directory's folder hierarchy, although in some cases it may be necessary to package content in the third level:
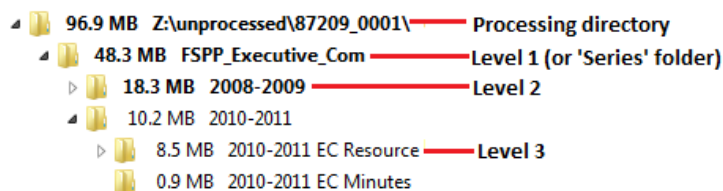

These decisions will be guided by the intellectual arrangement, convenience of management, and ease of access for our end-users. In thinking about access, we should consider not only how content will be linked from the finding aid but also how it will be displayed and downloaded from Deep Blue. Size is therefore an important consideration: while it might be easy for the library to describe and manage a

single 40 GB file, it could prove difficult or even impossible for an end-user to download such a package from Deep Blue. As a general rule, then, **consult with Nancy or Mike if you will produce .ZIP packages larger than two (2) GB.**

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.

4.8.4.1 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.
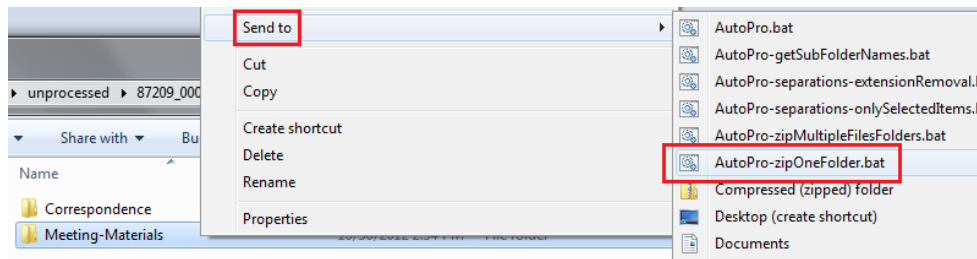
This operation may be performed on a top-level folder within the processing directory (i.e. \87209_0001\) or a subfolder thereof. After selecting the batch file, a new CMD.EXE window will open to track the progress of the operation. AutoPro uses 7-Zip to package and verify the .ZIP files; be patient, as large directories will take a longer time. The batch file will package all content within the target folder into a .ZIP file of the same name, located in the same parent directory (see below).

After the .ZIP file has been created and verified, AutoPro will notify you if an error occurred during the packaging process. In addition, if the folder contained subfolders, you will have the option to view a list of their names (only those at the highest level of the directory structure) that may be copied and pasted into the finding aid to give Bentley Library researchers a better idea of the contents of packaged digital objects.

At the next prompt, enter *C* to CLEAN the original directory (i.e. delete the unpackaged content) if the packaging procedure has proceeded without errors. Alternatively, you may enter *D* to DELETE the .ZIP file if an error has occurred or you need to reconsider the packaging strategy. AutoPro will immediately close after it has completed your request.

4.8.4.2 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 all of the target files/folders must be located in the same parent directory and that the resulting .ZIP file will be created in the same directory. 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.
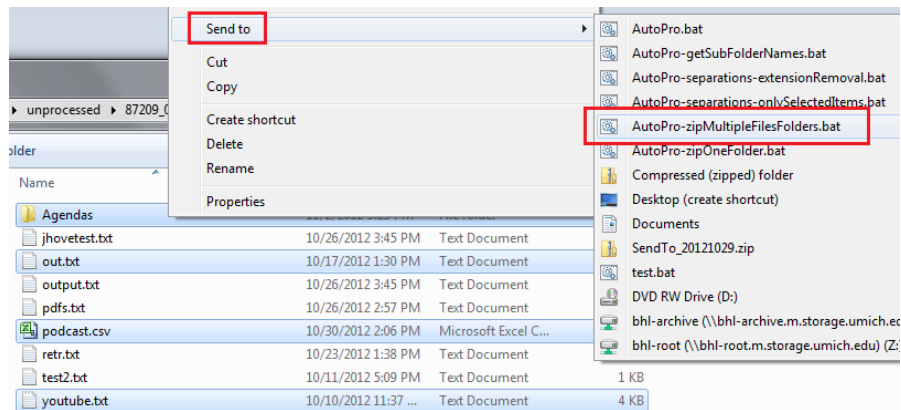
You will then be asked to enter a filename; use a descriptive, DACS-compliant name for the material, avoiding spaces and non-alphanumeric characters. Please note that AutoPro will supply the .ZIP extension.

After confirming (or changing) the filename, AutoPro will create a .ZIP file of that name in the same directory and then verify the completeness of its contents. If there are any errors in the process, you will be notified by an error message. As with the other packaging option, you will have the option to view a list of subfolders (if any are present immediately within the .ZIP file) so that their names may be copied and pasted into the finding aid. Once the process is completed and everything is satisfactory, you may enter *C* to CLEAN the original, unpackaged copies from the processing directory or *D* to DELETE the .ZIP files and package things differently.

4.8.5 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.

It is very important that you wait to complete the step until after all separations, arrangement, and packaging (if necessary) have concluded, as it search for and remove any empty directories in the deposit and create a final record of the deposit's arrangement.
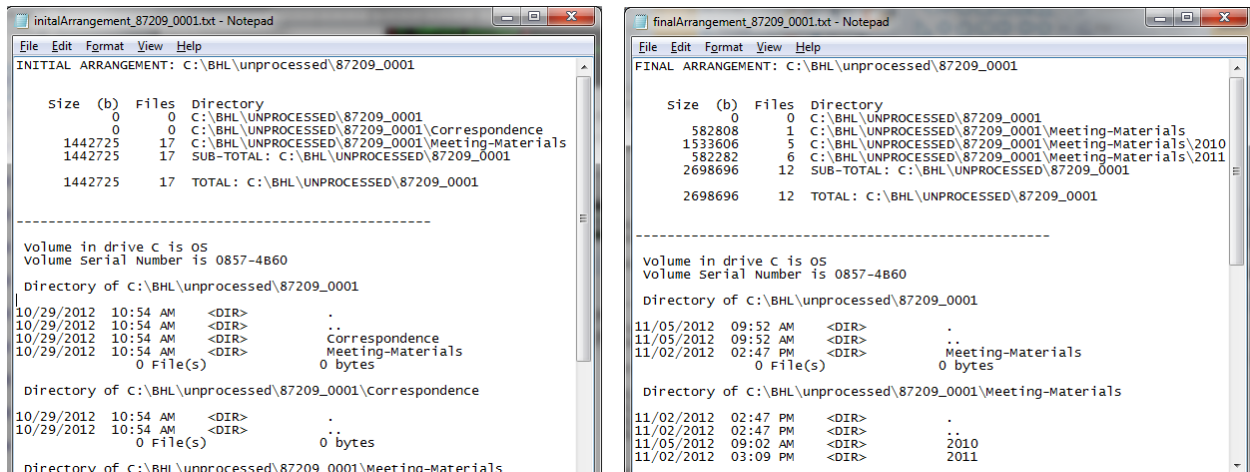
Taken together, the initial and final arrangement log files document the changes wrought by the Bentley Library on the structure and content of the deposit. Both files include a listing of all the folders in the processing directory (with the number of files and relative size of each), statistics for the deposit as a whole, and then an item-level manifest of the contents of every folder.

After formally completing this step, you will also have an opportunity to view a .CSV spreadsheet of information (number of files and final size in bytes) on the .ZIP files created during the packaging phase.

These statistics may be used in the series scope and content notes in the finding aid (see the BHL *Guidelines for Describing Digital Materials in Finding Aids* for more information). If necessary, you may save a copy of the spreadsheet in a local space for later reference.

4.9 Extract Technical Metadata

Upon initiation of this procedure, AutoPro will open DROID so that you may adjust the application's settings. It is essential that you verify these settings, as DROID will provide item-level checksums for all files (including those packaged inside .ZIP files) that may be used at a later date to determine the integrity of content. After approving any updates, click on the *Tools* menu item and select *Preferences*.
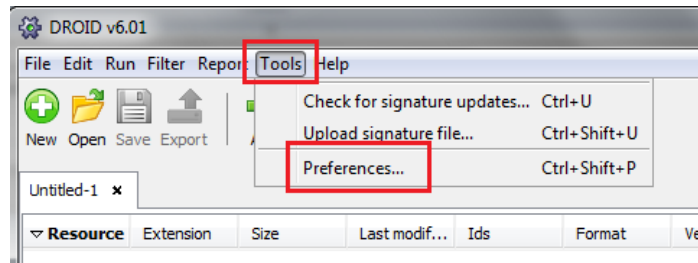
Once the Preferences window opens make sure that the boxes next to "Analyse contents of archive files" and "Generate MD5 hash for each file" are checked.
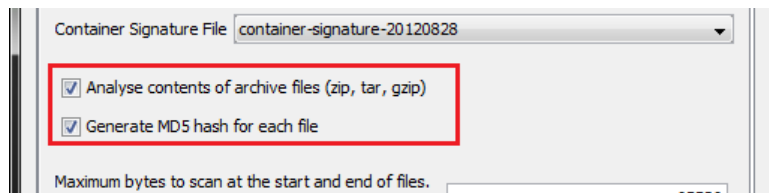
Once you have made the necessary changes, click "OK" and close the DROID window; AutoPro will prompt you to enter an *S* to verify that the settings are SAVED. The command line version of DROID will now launch and AutoPro will report on the progress as the application creates a profile of the content and then generates a .CSV spreadsheet. **Please be patient!** Deposits with a large number of files or very large individual files (audio, video, email, etc.) may take a long time to be processed—in some cases, hours. Please inform Mike if it appears that DROID is frozen or taking an exceptionally long time to complete its operation.

After DROID has finished, AutoPro will open a spreadsheet of the report for your inspection and then prompt you for a response (see below).

If there is an error (the most common will be a blank spreadsheet if the process was interrupted), you may run the process again—just be sure to close the Excel window for the report so that a new one may be generated! Otherwise, enter *C* to COMPLETE the procedure.

4.10 Add Descriptive Metadata

This step in the processing workflow allows you to add administrative and descriptive metadata to the deposit in preparation for its deposit in a long-term repository. It has been designed to standardize the creation (and documentation) of metadata about the Bentley Library's digital holdings.

As mentioned above, before initiating this step, you **MUST:**

- Complete your intellectual arrangement.
- Prepare a draft of your finding aid and submit it for review.
- Discuss how the content will be deposited and presented in Deep Blue.

To provide an overview of the procedure, you will first provide general, deposit-level administrative and descriptive metadata before identifying "Series" (and providing basic metadata for the title, scope and content note, contributors and access restrictions). You will then associate files/packages with the series and provide additional information for this content (descriptions and date ranges) while AutoPro records technical information (file size and checksums). 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.

4.10.1 Consult with Digital Curation

Upon initiation of this step, AutoPro will remind you to finish the intellectual arrangement and description that should have begun with the Appraisal and Arrangement of content (see section 4.8).

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. In addition, having a draft of the finding aid will allow you to save time by copying and pasting information into the AutoPro interface.

4.10.1.1 Overview of Deep Blue

Before discussing how material will be described and deposited in Deep Blue, it will be helpful to have a basic understanding of some DSpace features. A *collection* corresponds to a Bentley Library record group or manuscript collection:

An *item* in Deep Blue is roughly analogous to a Series in an archival arrangement:

Each "item" may in turn be associated with one or more *files* (individual files or packaged .ZIP files):

4.10.1.2 Strategies for Depositing Content in Deep Blue

While DSpace was not designed for archival materials, the Division of Digital Curation has established several options (and associated conventions) for describing and depositing materials in Deep Blue. It is important that you consult with Digital Curation to determine the best strategy for depositing a particular collection in the repository. The three main options are:

1. Create a single item for a series and associate one or more files with that item: This strategy works well when there are a relatively small number of files or packages (i.e. fewer than 20) and the material in the series is fairly homogenous. Consider the below example from the Men's Glee Club records:

Title:	Chronological File, 1886-1969, 2000-2011
Authors:	Men's Glee Club
Issue Date:	2012
Abstract:	The Chronological File represents the basic records of the collection. Arranged by academic year, it consists primarily of printed concert programs, newspaper clippings and other publicity material, occasional financial records--generally summary statements of club finances for an academic year, details regarding tours--particularly the annual spring tour--and minutes. Also includes digital versions of "Laudes," the Glee CLub newsletter, from 2000-2011. This series contains four individual documents of particular note. A fifteen page club history written by John Payne is found in 1958-1959. Narrative logs of the club's international trips in 1955, 1959 and 1967 were maintained and are filed in the appropriate years
Other Identifiers:	87250_0002 87250_0003
Appears in Collections:	Men's Glee Club (University of Michigan) Records

Files in This Item:

File	Description	Size	Format	
1886-1893.zip	Chronological file; collected records from 1886-1893	2824Kb	ZIP	View/Open
1911-1942.zip	Chronological file; annual records from academic year 1911/12 to 1941/42	303846Kb	ZIP	View/Open
1944-1957.zip	Chronological file; annual records from academic year 1944/45 to 1956/57	562536Kb	ZIP	View/Open
1957-1969.zip	Chronological file; annual records from academic year 1957/58 to 1968/69	974595Kb	ZIP	View/Open
laudes.zip	Digital versions of "Laudes," the University of Michigan's Mens Glee Clubs newsletter, from 2000 to 2011	156873Kb	ZIP	View/Open

By creating a single Item to represent the Series, the archivist consolidated the materials so that a researcher would not have to open multiple Items to access different files. All of the packaged .ZIPs are furthermore similar so that there is no need to provide the additional description that could be had if each file had its own Item.

2. Create a single item for each individual file within a series: This strategy is most often employed with collections such as the Michigan Today records or the News and Information Services podcasts. With the podcasts, for example, there is a single series and each archival recording is represented with a direct link ("Download file") to the item in Deep Blue (see below):

News and Information Services (University of Michigan) Podcasts 2006-2010

full text *File Size: 123 K bytes* | Add this to my bookbag

Search within this finding aid:

Title Page
Summary Information
Access and Use
History
Collection Scope and Content Note
Subject Terms
Contents List
　　Audio Podcasts, 2006-2010

Container / Location

Title

Audio Podcasts, 2006-2010 [series]:

The 70 percent solution, 2008-04-28
MMPEI is working with other organizations on campus to develop a tool to help scientists and engineers project whether new cleaner technologies will take hold. Speakers: Nicole Casal Moore, Gary Was. (70percent_mixdown.mp3; 5:39; audio/x-mpeg) [download file]

Adult kids and their parents, 2009-05-29
New research shows that it's common for adult children and their parents to annoy each other. The research identifies common sources of irritation, and how parents and adult kids cope. Speakers: Diane Swanbrow, Kira Birditt. (AdultKids.mp3; 3:37; audio/x-mpeg) [download file]

M-flu results positive, 2008-10-23
The results suggest that masks and handwashing can reduce the spread of symptoms by as much as 50 percent. Speakers: Laura Bailey, Allison Aiello. (aiello_oct_08.mp3; 5:32; audio/x-mpeg) [download file]

Given the relatively "flat" arrangement and the importance of having a direct link to each item in the repository, it makes sense to establish a unique Item for each file:

Browsing "News and Information Services (University of Michigan) Podcasts" by Title

Jump to: 0-9 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
or enter first few letters: [　　　　] Go!

Sort by: [title ▾] In order: [ascending ▾] Results/Page [20 ▾] Update

Showing results 1 to 20 of 463

next >

Issue Date	Title	Author(s)
8-Nov-2006	"After the Gallatin River, Outside Bozeman, Montana" - Poem by Britta Ameel 06 MFA	*News and Information Services, University of Michigan*
8-Nov-2006	"Blue Tips" - Poem by Sean Norton	*News and Information Services, University of Michigan*
3-Nov-2006	"Boys with Crucifices" - Poem by Ray McDaniel	*News and Information Services, University of Michigan*
8-Nov-2006	"December" - Poem By Erin Morris	*News and Information Services, University of Michigan*

Creating an individual Item for each file also allowed the archivist to apply more detailed descriptive and technical metadata about the content:

dc.contributor.author	News and Information Services, University of Michigan	en_US
dc.contributor.other	Moore, Nicole Casal	en_US
dc.contributor.other	Was, Gary	en_US
dc.date.accessioned	2011-05-24T15:19:30Z	-
dc.date.available	2011-05-24T15:19:30Z	-
dc.date.issued	2008-04-28	en_US
dc.identifier.other	87265_0001	-
dc.identifier.uri	http://hdl.handle.net/2027.42/84169	-
dc.description.abstract	The Michigan Memorial Phoenix Energy Institute (now the University of Michigan Energy Institute, or UMEI) is working with other organizations on campus to develop a tool to help scientists and engineers project whether new clean energy technologies will take hold.	en_US
dc.format.extent	5:39	en_US
dc.format.mimetype	audio/x-mpeg	en_US
dc.language.iso	en_US	en_US
dc.rights	Copyright is held by the Regents of the University of Michigan.	en_US
dc.title	The 70 Percent Solution for Reduction of Greenhouse Gases	en_US
dc.type	Recording, oral	en_US
dc.description.bitstreamurl	http://deepblue.lib.umich.edu/bitstream/2027.42/84169/1/70percent_mixdown.mp3	-
Appears in Collections:	News and Information Services (University of Michigan) Podcasts	

3. <u>Create multiple Items for a single Series and associate one or more files with each</u>: One of the greatest challenges of working with Deep Blue is that its relatively flat hierarchy makes it difficult to represent the often complex relationships between archival materials. To better represent the hierarchies within archival Series, we will use the Item titles to better represent the provenance and order of materials.

As an example, the Men's Glee Club toured China in 2012 and the trip resulted in a single archival series comprised of some 40 files and .ZIP packages. The series was arranged into four subseries that represented major cities visited on the tour. Each of these subseries was then divided by major events or sites visited while in that particular city. To compound matters, the content consisted mainly of very large audio and video files.

Rather than create a single Item for the whole series or even individual ones for each city, the archivist created multiple Items that categorize files down to the third level of the intellectual arrangement:

2012	China Concert Tour 2012 - Beijing - Arrival	Men's Glee Club
2012	China Concert Tour 2012 - Beijing - Forbidden City	Men's Glee Club
2012	China Concert Tour 2012 - Beijing - Great Wall of China	Men's Glee Club
2012	China Concert Tour 2012 - Beijing - Jade Factory	Men's Glee Club
2012	China Concert Tour 2012 - Beijing - Olympic Park	Men's Glee Club
2012	China Concert Tour 2012 - Beijing - Tea Ceremony	Men's Glee Club
2012	China Concert Tour 2012 - Beijing - Tiananmen Square	Men's Glee Club
2012	China Concert Tour 2012 - Beijing - Tsinghua University Meng Minwei Concert Hall	Men's Glee Club
2012	China Concert Tour 2012 - Chongqing - Ba Shu Secondary School	Men's Glee Club
2012	China Concert Tour 2012 - Chongqing - Dazu Stone Carving Park	Men's Glee Club

This approach permitted more detailed metadata to be added to the items and also enables the user to directly access specific (and very different) aspects of the series without having to download excessively large files.

Title:	China Concert Tour 2012 - Beijing - Great Wall of China
Authors:	Men's Glee Club
Issue Date:	2012
Abstract:	Photographs and video of the Men's Glee Club's visit to the Great Wall of China in Beijing on May 5, 2012. The club visited Beijing from May 3 to May 5, 2012.
Other Identifiers:	87250_0001
Appears in Collections:	Men's Glee Club (University of Michigan) Records

Files in This Item:

| File	Description	Size	Format	
Great-Wall-Photos.zip	Photos of the Men's Glee Club at the Great Wall of China	103651Kb	ZIP	View/Open
Great-Wall-Video.zip	Video of the Men's Glee Club singing	227515Kb	ZIP	View/Open

This strategy thus provides additional access points into the collection, preserves the intellectual arrangement of the finding aid, and ensures that patrons can download files of manageable sizes.

Given these different strategies, it will be important for you to consult with Digital Curation to determine the best approach for your deposit.

4.10.2 Add Deposit-Level Metadata

AutoPro will require you to enter descriptive and administrative metadata about the deposit as a whole. A plain text file will automatically open and instructions guide you through entering information related to the creator, collection title, donor, donor ID, donation year, copyright statement, and an abstract.
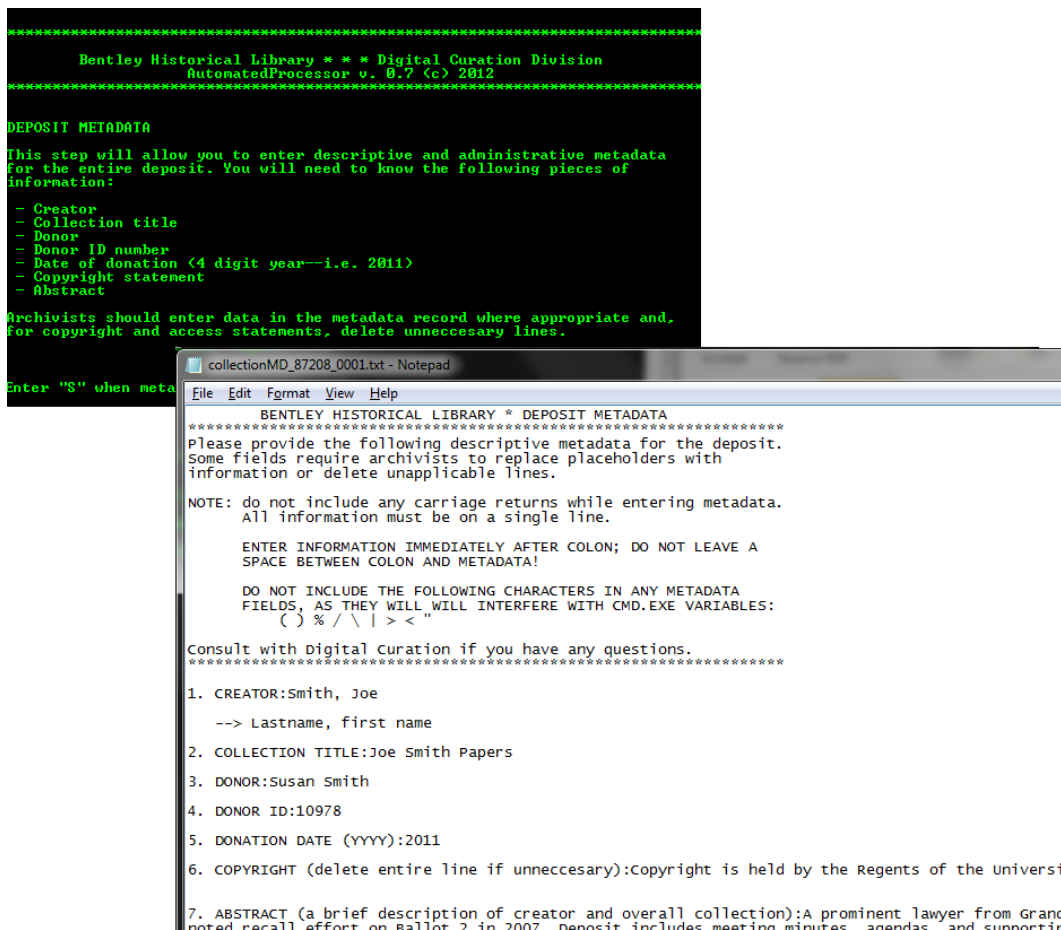
Please enter information immediately after the colons on the respective lines and do not include any carriage returns. When you have finished entering metadata, save the file and then verify that the save operation with AutoPro. The program will then parse the text file, display the results, and give you an opportunity to edit the information as entered.

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

4.10.3 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. 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 do not already have a pre-completed metadata spreadsheet, respond to the AutoPro prompt with an *N* and move on to the next section of the User Manual.

If you do have such a spreadsheet (prepared by the donor or created through the reuse of existing metadata), it should be organized with the following columns:

- IDENTIFIER.OTHER: unique BHL digital deposit ID
- DC.TITLE: title of the Deep Blue Item
- DC.DESCRIPTION.ABSTRACT: description of the Deep Blue Item
- DC.CONTRIBUTOR.AUTHOR: primary creator of content
- DC.CONTRIBUTOR.OTHER: additional content creators (for people, enter as "Smith, Jane" and separate multiple entries with semi-colons
- DC.DATE.ISSUED: the year content is deposited in Deep Blue

- DC.DATE.CREATED: the date (expressed in years; may be a range) that content in BHL custody was last modified. Applies to all materials associated with the Item.
- DC.COVERAGE.TEMPORAL: the date (expressed in years; may be a range) that reflects when the information content of the digital material was actually created or used. Applies to all materials associated with the Item.
- DC.TITLE.FILENAME: filename(s) of content associated with the Item; separate multiple entries with semi-colons.
- DC.FORMAT.MIMETYPE: Mime type of files associated with the Item; separate multiple entries with semi-colons. If there are multiple entries, these should be ordered so that they match the order of their respective filenames.
- DC.DESCRIPTION.FILENAME: descriptions of the file(s) associated with the Item; separate multiple entries with semi-colons. (NOTE: semi-colons should therefore be avoided in these descriptions.) If there are multiple entries, these should be ordered so that they match the order of their respective filenames.
- DC.DESCRIPTION.MD5.CHECKSUM: this field is not required and checksums will be maintained in the DROID report. If included, separate multiple entries with semi-colons and maintain appropriate order of files.
- DC.TYPE: this element indicates the type of content in the deposit and is used with .ZIP files; see User Manual section 4.10.11 for more information on content types.
- DC.RIGHTS.ACCESS: this element applies to the entire Item and indicates any access restrictions.
- DC.RIGHTS.COPYRIGHT: this element applies to the entire Item and indicates the content's copyright holder.
- DC.LANGUAGE.ISO: this indicates the language used for the Item's metadata entry; in all cases it should be "EN-US"

The spreadsheet should be saved in the deposit Metadata folder (located at \logs\(depositId)\Metadata) and must conform to the convention deepBlue_(deposit id).xlsx. See the example below:
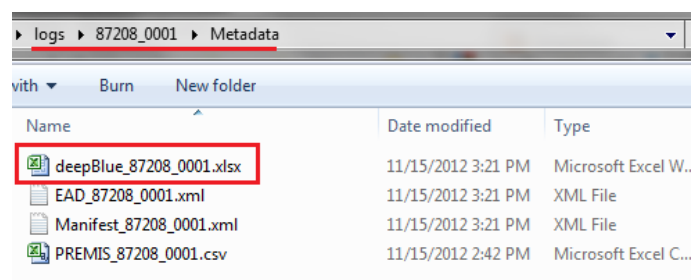
Be sure that the spreadsheet conforms to the above conventions; AutoPro will then prompt you for some general information about the deposit as a whole. You will need to provide date range information about when the files were originally created/used and last modified (since these two dates are often different) as well as what access restrictions are necessary for the deposit. (NOTE: TreeSize Professional can be helpful in determining the range for the deposit's overall last modified dates.)

```
At the prompt, enter the inclusive date range for when the materials in the
deposit were originally created/used. Three example entries are provided:

Examples: a) 2000      b) 2010-2012      c) ca. 1999-2001

2010-2011


At the prompt, enter the inclusive date range for when the files currently
in the Bentley Library's custody were last modified. If the last modified
date is the same as the original creation/use date, hit "Enter".

2012


Summarize the deposit's access restrictions:

   -O: Content is OPEN for research
   -R: Access to content is RESTRICTED; see finding aid for details

<O / R> O
```

If the content of the deposit has already been described in a spreadsheet, this step will conclude the procedure.

4.10.4 Create a New Series and Other Options

The first time you add descriptive metadata to a deposit, you will be immediately taken to the "Main Series Options" screen after you have entered general information about the deposit and indicated that a pre-completed metadata spreadsheet is not present. (Please note that if you have previously initiated the addition of descriptive metadata, you will instead have the opportunity to review existing series— see section 4.10.14 for more information about the "Series Review"). You will also encounter this screen each time you formally complete the identification and association of content with a series.

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

SERIES:

--> IDENTIFICATION AND DESCRIPTION

Archivists will define Series in which content will organized for deposit
in Deep Blue and record descriptive metadata.

Do not conclude metadata entries with periods and avoid using non-
alphanumeric characters.

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


Options:

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

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

From this point, you may create a NEW series, REVIEW existing series (which in turn will allow you to add additional content to a series or complete the description thereof), FINALIZE the deposit's packaging (see section 4.10.15), or QUIT and resume at a later date. The User's Manual will proceed with the creation of a new series, but explanations of the other options will be found elsewhere in this guide.

4.10.5 Add Series-Level Metadata

Upon electing to create a new series, a "Series Metadata" text file (similar to the "Deposit Metadata" file) will open.

In addition to entering the series title and a descriptive note, you will need to address:

- Path to series: in arranging content, we have assumed that the top-level folders in the processing directory correspond to "Series." In packaging materials for deposit, you may have a number of files within a top-level folder that will be deposited individually or the entire top-level folder may have been packaged in an uncompressed .ZIP file (see examples below).
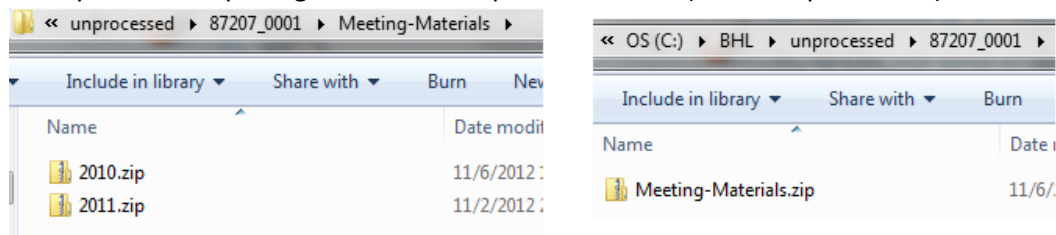
If there is a top-level folder that corresponds to a "Series," enter the full path to it; if the entire folder has been zipped, simply delete this line from the metadata file.

- Series contributors: Note individuals (or organizations) that have materially contributed to the Series, include names (last name, first name), separating entries with a semicolon. Delete the entire line from the file if unnecessary.
- Access Restrictions: select the appropriate access restriction and, if necessary, replace "YYYY" with a four-digit year. Delete all the non-applicable entries for access restrictions.

See a completed example below:

Enter *S* when you have saved the metadata file; AutoPro will then give you an opportunity to make corrections as needed.

4.10.6 Add Content to the Series and Other Options

After the Series metadata has been saved, AutoPro will proceed to the "Main Content Options" screen.

You will encounter the above screen each time you have completed the association of content with the Series. From it, you will have the option to ADD new content to the series, COMPLETE the packaging and description of the Series (in which case you will be returned to the Main Series Options screen described in section 4.10.5), or QUIT and resume at a later date. The User's Manual will assume you are going to add new content to the Series.

4.10.7 Enter Path to File Associated with Series

In the next screen, AutoPro will allow you to provide the full path to a file to be associated with the Series (if needed, you may also complete the Series description or quit and resume the procedure at a later date). To reduce spelling or other errors, it is recommended that you copy and paste the path or use the CMD.EXE AutoComplete functionality.

AutoPro will then prompt you to verify that this is the correct file; if an incorrect path has been entered, you will be directed to re-enter the correct one.

4.10.8 Describing the File

After the path has been entered, you will enter a description of the file (or, in the case of a .ZIP package, the contents thereof). Please note that this description should be identical to the description of the material in the finding aid's container list.

AutoPro will give you the opportunity to edit or re-enter the entry; if necessary, you may hit the "up" arrow to bring back the description via the CMD.EXE command history so that you avoid re-typing the entire description).

4.10.9 Content Creation and Last Modified Dates

To accurately represent the digital provenance of content, you will attempt to identify when the materials were originally created/used and also when the content in the Bentley Library's custody was last modified. Please note that for scanned or digitized materials, we will consider the original creation/use dates to be those of the analog/paper versions (and not the date on which the digital surrogate was generated).

AutoPro will present you with the last modified date for the material and ask if it accurately represents the period when it was originally created and actively used. (For .ZIP files, it has recorded the last modified dates for all content during the packaging operation; it may therefore present you with a date range that will include newly created preservation copies of content.) While this may be the case, it is also highly possible that the last modified date is much more recent than the original creation date.

 If there is a discrepancy between the last modified and creation dates, AutoPro will then prompt you to enter the original dates.

To avoid errors in date-processing, please:

- Only include four digit years (no days or months)

- Enter as YYYY or YYYY-YYYY (i.e. 2010 or 2009-2012)
- Use *ca.* (not "circa" or "Ca."); also, do not include an *s* (i.e. 2000s)

AutoPro will ask you to verify that the date is correct and properly formatted; if not, you will be permitted to re-enter the information.

4.10.10 Mime Type Identification

To meet deposit requirements for Deep Blue, AutoPro will record the mime type of the content, comparing the file's extension to a list of known mime types stored locally. If the file extension (and mime type) are not recognized, you will be asked to provide a mime type. The following website provides a thorough listing: http://www.webmaster-toolkit.com/mime-types.shtml

Consult with Digital Curation if you need assistance.

4.10.11 Content Type Identification

To comply with a request from the Deep Blue administrator, we will identify the contents of a .ZIP file. AutoPro will provide you with a controlled vocabulary of content terms; select those that apply and enter them on one line, separating multiple entries with hyphens.

4.10.12 Return to the Main Content Options Screen

Having entered all descriptive metadata for the file, you will now be returned to the Main Content Options screen described in section 4.10.6 of the User's Guide. From this screen, you may add additional files to the Series, complete the packaging and description of content for the series, or quit AutoPro to resume at a later time.

4.10.13 Complete the Packaging and Description of a Series

When you elect to complete the packaging and description of a series, AutoPro will write out the metadata for the series and associated files to the Deep Blue spreadsheet and XML manifest. You will then be returned to the Main Series Options screen described in section 4.10.5 of the User's Guide.

From here, you may create a new series, review series that have already been created, finalize the entire deposit (as described in section 4.10.15), or quit and resume at a later date.

4.10.14 Review a Series

You will access the "Series Review" screen in one of two ways:

1. You have resumed the "Add Descriptive Metadata" procedure after previously quitting it. In this case, you will arrive at this screen immediately after you initiate the procedure and are given the option of revising the deposit-level metadata (as described in section 4.10.1).
2. You select option *R* ("Review Series created for this deposit") from the Main Series Options screen described above in section 4.10.5.

In either case, this screen will display each of the Series you have created for the deposit, listing a Series ID (generated for internal use by AutoPro), title, description, and status (either "completed" or "NOT completed"). Please note that **ALL Series must be formally completed** for AutoPro to correctly generate metadata records.

There are two main groups of options on this screen. First, you may perform actions on a particular series: ADD additional content to an existing series (if **not** complete), VIEW the contents of a Series, or COMPLETE the packaging and description of a Series. When you select any of these options, AutoPro will ask for the Series ID number:

- If you elect to add content, you will be taken to the "Enter Path" option described in section 4.10.6 of the User's Manual.
- If you elect to view content, AutoPro will display the filename and description for all content associated with the Series before returning you to the "Series Review" screen.

- If you elect to complete packaging and description of the series, AutoPro will perform the steps described in section 4.10.13 of the User's Manual.

The second group of options available on the "Review Series" screen are ones we have already seen: you may create a NEW Series (starting with the creation of Series-level metadata as described in section 4.10.5 of the User's Manual), FINALIZE the deposit (as described in section 4.10.15), or QUIT and resume at a later date.

4.10.15 Finalize the Deposit

When you have created all requisite series, provided metadata, and associated and described all files with the appropriate series, you may finalize the deposit. In this step, AutoPro will calculate the overall size of the deposit as well as date ranges to represent the creation and last modification of content and complete the Deep Blue metadata spreadsheet, Dublin Core XML manifest, and the EAD administrative information. You will now be ready to proceed to the next procedure.

4.11 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!

4.11.1 Transfer to Deep Blue Deposit Folder

Upon initiation of this step, AutoPro will ask if the materials will be deposited in Deep Blue so that the general public may have access. If the content is restricted in any way, respond with *N* and proceed to the next step.

If the content is unrestricted and permitted to be deposited into Deep Blue, you will be asked to indicate the drive letter on your work station that is mapped to the BHL Interim Repository.

You will have an opportunity to correct your entry before AutoPro copies the files to the deposit folder. If any errors occur, you will be alerted so that you can retry the operation. Consult with Mike if any issues arise.

4.11.2 Transfer to Dark Archives

To transfer content to the BHL Dark Archives, you will need to indicate the drive letter to which it is mapped on your work station and AutoPro will request verification. The material (and metadata) will be packed into "Bags" with the Library of Congress BagIt tool and then copied to the appropriate deposit directory in the BHL Dark Archives. If any errors occur, you will be alerted so that you can retry the operation. Consult with Mike if any issues arise.

4.12 Clean Up

In the final step of the digital processing workflow, AutoPro will delete the processing directory, the backup restore point, and other temporary files. If content has only been transferred to the BHL Dark Archives, do not perform this operation until content has been captured by the "snapshot" functionality of Mainstream storage.

AutoPro will first check to see if the Separations should be deleted as well; check the transfer agreement if you are unsure of the disposition of separated material. 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.

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

TRANSFER LOG FILES

Upload C:\BHL\logs\87206_0001\0001.zip to BHL IFS space:

   --> ../bhl/Private/html/processing_log_files/0001/

   NOTE: Consult with Mike if problems arise.


Has 0001.zip been successfully uploaded? (Y / Q to QUIT) _
```

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.

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

CLEAN UP

     - C:\BHL\unprocessed\87206_0001
     - C:\BHL\tmp\87206_0001
     - C:\BHL\backup\87206_0001
     - C:\BHL\logs\87206_0001
     - C:\BHL\separations\87206_0001

Options:

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

(D / Q) _
```

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 . . . _
```