

DIGITAL PRESERVATION WORKFLOW



CORRESPONDENCE HERE

*Hells Everybody:- Will
come Thurs. morning
in meanwhile be good
all of you, Harry too. Don't
be feel a little lonely for
me, I do, I have got such
a cold and feel so gripping
but will be all right if I
only reach the Island again*

NAME AND ADDRESS HERE

*Mr Frank Scott
R.R. I Box 21.
Wash*



Table of Contents

Considerations Before Digitization	3
Project Planning	3
Selection Checklist	4
Documenting Your Plan.....	4
Identify and Involve Stakeholders	5
Budget	5
Technology.....	6
Scanner.....	6
Workstation.....	6
External Hard Drive	6
Software.....	7
Copyright.....	8
Digital Asset Management.....	11
File Naming and Organization	11
Protecting Your Physical Collections	11
Storage and Backup	12
Long-term Preservation Best Practices	12
Project Workflow.....	14
Workflow Overview.....	14
Standards for Scanning.....	16
Metadata	22
Exactly.....	24
Fixity.....	29
Tools	31
Appendix	32

Considerations Before Digitization

Information combined documentation from OWLS, Recollection Wisconsin, and CCDC Workshop

Project Planning

What should we digitize?

Before embarking on a digitization project, consider *why* you want to build a digital collection. Be thoughtful and considerate about why you selected the materials to help focus the project, and to help keep it manageable. Furthermore, think about how this digitization project will support the mission of your organization and community.

Why might you digitize?

Reasons for digitizing include new audiences, improving access to rare and/or unique materials, to protect fragile and/or heavily used materials, and to create educational resources for students and teachers.

Selecting Content for Digitization

When deciding what materials to select for your project, think about creating digital collections that help represent original materials related to state and/or local history.

Formats to include:

- Visual Materials
- Manuscripts
- Printed text
- Audio and video
- Physical objects

Consider the following criteria when selecting resources to digitize:

- Collection of high interest to your community.
- Items that are unique.
- Collection with a defined target audience to help influence a collection development rationale for digitization.
- Collections should have a definable scope, so that a project timeline for completion and staffing levels can be established.
- Collections that have clearly defined ownership. Rights and permissions for electronic distribution must be secured or securable.
- Original collections in an appropriate format and physical condition for scanning, objects that can be handled, nondegraded, with sufficient clarity of detail, appropriate size to fit the equipment (At this time, we are not set up to handle newspapers, microfilm, large books, and maps)•

- Collection with organization and has some item-level metadata (e.g. title, creator, date, etc.) available or inherent. Furthermore, image files uploaded to CONTENTdm must be accompanied by metadata. If necessary, staff should do research to provide usable access points such as pictured subjects and information on the circumstances surrounding the image (photographer, event, significance, timeframe, etc.).

In regards to newspapers, Recollection Wisconsin states that “due to a lack of centralized full-text searching, Recollection Wisconsin is not the ideal solution for access to digitized newspapers. If you have a historic newspaper collection you’d like to bring online, consider contributing to the Archive of Wisconsin Newspapers.”

Selection Checklist

For every collection that you select for digitization, consider the following questions. If you answer *no* to any one of the questions, this collection is not a good candidate for digitization.

- Are the materials in the public domain or can we get permission from the rights holder to make them available online?
- Are they unique and not already online?
- Are these materials of interest or value to genealogists, educators, researchers, community members or other potential audiences?
- Is there enough accurate information available to add useful context?
- Do you have appropriate equipment and skills to create a good-quality version? Or do you have resources available to outsource digitization if necessary?

Documenting Your Plan

Documenting your plan for digitization is important since many projects can take longer than expected. Throughout the process of digitizing a collection, staffing and volunteers might change. Documenting your plan protects your investment of time that staff/volunteers have put into the project by creating the digital images, creating metadata, and maintaining digital content. Remember to start small, set realistic goals, deadlines and define project roles.

When getting started, use the Digital Project Planning Worksheet from Recollection Wisconsin to outline your project before getting started. This will help document decisions and changes made, and will be helpful if staff or volunteers move on.

www.recollectionwisconsin.org/wpcontent/uploads/2012/11/DigitalProjectPlanningWorksheet2016.pdf

Identify and Involve Stakeholders

All stakeholders should be involved at the beginning of the project. This includes staff, volunteers, administration, and any potential partners.

Conduct a user needs assessment. Is the project worth doing? Who is your audience? What are their access requirements? Refer back to the Selection Checklist.

Consider who will fill key roles in the project, such as:

- Project leader
- Selecting collections
- Inventorying selected objects
- Creating metadata
- Scanning
- Inputting metadata
- Uploading files
- Technical support (hardware maintenance, software installation)
- Marketing/public relations
- Public services related to new digital collections
- Ongoing collection maintenance (quality checks, backups, archiving files, refresh and migration of archive files)

Budget

Consider how you will provide funding and staffing for this project in the short and long term. Potential costs include:

- Staff time
 - Scanning and metadata creation are very time-consuming
 - Time for quality control
- Consulting and outsourcing
- Software and hardware licenses/service contracts (if image processing software is purchased)
- Hardware
 - Windows PC with adequate memory and storage
 - Equipment upgrades (plan on new equipment every 3 years)
- Designated workspace for digital projects (cubicles, chairs, tables, glare free area, etc.)
- Project supplies
- Publicity/promotion costs

Technology

Scanner

- When choosing a scanner, consider the size and format of items to be scanned
- Consider the volume of scanning you plan to do. Some scanners are faster than others.
- Consider the knowledge-sharing benefit of other InfoSoup Memory Project participants using the same scanner and interface.
- Contact OWLS regarding opportunities to use OWLS's or other participants' scanners for smaller projects.
- If you have a very large collection to scan or unusual objects, consider outsourcing the imaging rather than purchasing expensive equipment.

Workstation

- Check the minimum processing requirements for the scanner and the software you plan to use.
- Working with large image files requires good processing speed and lots of RAM.
- You will need significant hard drive space for saving images unless you plan to save them on an external hard drive. (Recommended for preservation purposes.)
- A larger monitor, 19-in or more, is ideal for working with multiple images

External Hard Drive

- If possible, purchase two USB or firewire drives with 1 TB of storage. Save your master TIFF files on both drives, and send a digital copy to OWLS.
- Serves as storage for large master files, conserving space on your working computers.
- Backup to a second medium if possible, such as cloud storage.
- Storage mediums that need to be addressed sooner, rather than later include:
 - External hard drives that are older than 3 years old
 - Typical life span of an external hard drive is 3-5 years
 - CD/DVD, Archival Gold Disks, or other optical discs
 - New computers do not have a disk drive
 - Life span is unknown
 - Thumb drives
 - Can pass virus' from one computer to another
 - Subject to corruption

Software

- Look for scanning software that includes basic editing features, including ability to crop, rotate, descreen, sharpen, and adjust a histogram.
- Photoshop or Photoshop Elements Image manipulation software might come packaged with your scanner.
- Photoshop Elements allows batch processing of images, so that you can quickly create a set of access JPGs from your archival TIFFs.
- Microsoft Excel for metadata collection, sorting, and uploading
- Optical Character Recognition (OCR) software to extract text from images. This may come with your scanner. OWLS also has OCR software.

Copyright

Combined documentation from OWLS, Recollection Wisconsin, and CCDC Workshop

OWLS and Recollection Wisconsin will be held harmless for all copyright challenges.

Possession of originals does not equal ownership of copyright.

Document your project's copyright clearance process.

Your digital collections must fit into one of the following categories:

- Items are in the public domain and may be copied and disseminated without restrictions.
- Items are protected by copyright but permission to use them has been (or will be) secured before uploading.
- Items are protected by copyright but fair use provisions apply.
- Items are protected by copyright but your institution owns the copyright.

Review any existing deeds of gift. If the copyright owner is available, work with him or her to secure the appropriate permissions. Orphan works—those for which the copyright owner cannot be identified or found—may be considered for inclusion. Consult with OWLS before beginning a collection which includes orphan works.

You will need to choose the wording of the copyright statement that will appear with the images in CONTENTdm, and discuss how your organization and the copyright owner will handle requests for images.

Recollection Wisconsin, as a DPLA Service Hub, is moving toward requiring the use of the standardized rights statements currently under development by Europeana and the Digital Public Library of America (DPLA). We recommend that you model your rights statements on those found here: <http://rightsstatements.org/en/>

We also recommend that you review these copyright sites:

Copyright Basics by the U. S. Copyright Office:
<http://www.copyright.gov/circs/circ01.pdf>

Peter Hirtle's chart helps you determine whether published or unpublished works are in the public domain:
<http://copyright.cornell.edu/resources/publicdomain.cfm> 8 Revised 07/06/2016

Mary Minow's article is a very thorough explanation of copyright as it relates to library digitization projects: <http://www.llrx.com/features/digitization.htm>

Interactive tool for evaluating Fair Use from the Office for Information Technology Policy of the American Library Association:
<http://librarycopyright.net/resources/fairuse/>

The Society of American Archivists (SAA) has published best practices for orphan works (works for which the copyright holder cannot be reached):
<http://www2.archivists.org/sites/all/files/OrphanWorks-June2009.pdf>

Determining the copyright status of the materials you make available online is essential to any digital project. Two important things to remember:

- Owning the physical item does not necessarily mean you own the copyright to that item
- Scanning an item does not change its copyright status or create a new copyright

Can you put it online?

Many of the historical materials in your collection may be in the public domain. This means that copyright protections do not apply and may be used freely by anyone, without permission or attribution.

What is considered public domain?

Books, articles, maps, pamphlets, and other published materials are considered in the public domain if they were published before 1923 or published between 1923 and 1977 without a copyright notice.

Letters, diaries, manuscripts, and other unpublished materials (including many photographs) pass into the public domain 70 years after the death of the creator or 120 years after the date created, if the creator is anonymous or death date is unknown.

In Copyright? Get Permission.

If the historical materials do not fall in the public domain, you must seek permission from the copyright holder before putting them up. You will need to contact the rights holder to request permission. Cornell University and Hathi Trust have templates of how to contact the copyright holder. Keep a record of when you contacted them and what responses you have received.

Should I put it online?

Privacy and Ethics

Even if the materials are in public domain, that does not mean that they should go online. Be mindful of privacy issues or other ethical concerns, for example:

- Materials created by or depicting Indigenous communities
- Materials created by or depicting protected populations (minors, inmates, medical patients)
- Materials containing personal or private information

Fair Use

You may wish to have a plan in place for responding to requests for reproductions, including any fees that you might charge.

Rights Statement

Every item put online must be accompanied by a statement describing the item's copyright status. Include two parts:

- Copyright status statement (in copyright, no copyright/public domain, or copyright undetermined)
- "For more information" statement (name of organization providing the item)

Examples:

- Copyright 2010, Board of Regents of the University of Wisconsin System. All rights reserved. For more information, contact University of Wisconsin-Milwaukee Libraries.
- Public domain. For more information, contact the W. J. Niederkorn Library, Port Washington, Wisconsin.
- Copyright status undetermined. For more information, contact the Langlade County Historical Society, Antigo, Wisconsin.

Digital Asset Management

Combined documentation from OWLS, Recollection Wisconsin, and CCDC Workshop

File Naming and Organization

- Keep folder labels and file names short
- Keep naming conventions simple and logical
- Use only lower case letters, numbers, dashes and underscores
- Don't use special characters such as ^"<>|? \ / : @' * & .()
- Don't use spaces. Alternatives to spaces in file names include:
 - Underscores, e.g. file_name.xxx
 - Dashes, e.g. file-name.xxx
 - No separation, e.g. filename.xxx
 - Camel case, where the first letter of each section of text is capitalized, e.g. FileName.xxx
- When using a sequential numbering system, using leading zeros to make sure files sort in sequential order. For example, use "001, 002, ...010, 011 ... 100, 101, etc." instead of "1, 2, ...10, 11 ... 100, 101, etc."

Create a document describing your file naming conventions and folder organization structure. A good strategy is to include in the directory a readme.txt file that explains your naming format along with any abbreviations or codes you have used.

Protecting Your Physical Collections

The digitization project is the perfect opportunity to assess the current condition of the original materials and address any preservation concerns. Remember, just because you are digitizing materials does not mean that you should stop caring for the physical item. Store your physical items in archival, acid free folders, and boxes. Gaylord Archival is a great source for archival materials. If you have any questions on the care of your archival collections, please contact OWLS.

Tips for safe handling of original materials during digitization projects:

- No food or drinks near collection items
- No pens near collection items - pencil only
- Wash hands regularly or wear gloves when handling original materials
- Always have plenty of room in your workspace to accommodate the material you are working with
- Never use collection items as a work surface
- Do not stack different items together such as books and photos
- Return items to their storage area at the end of the day

Storage and Backup

Digital Archival Storage

The main goal of digital archival storage is to manage digital objects over time and across different generations of technology. Proper digital archival storage and best practices ensures that files you create today can be opened and used in 5, 10 or 50 years from now.

The 3-2-1 Rule

- Maintain at least two, ideally three, copies of your digital files
- Use different types of storage mechanisms (don't rely on one form of technology)
- Separate your copies geographically (think natural disaster)

Storage Tips and Best Practices

- Separate your master copies from you access copies. Do not access archival files unless necessary.
- If using localized media (external hard drives), use common-sense handling. Minimize exposure to dust, jarring, temperature fluctuations, magnets, UV light.
- Migrate files regularly to new media. External hard drives have a shelf-life of 3-5 years.
- Document your storage decisions. Where is it? Who can access it? How?

Selecting Storage Media

Common options include local servers, cloud storage, or external hard drives. Keep in mind that external hard drives must be rotated out and replaced at least every three to five years.

Long-term Preservation Best Practices

- A master image should be created using file formats and compression techniques that conform to industry standards such as TIFF or lossless JPEG2000. OWLS recommends TIFF.
- Store the master image file in a designated location on an appropriate long-term storage medium. The file should be used to produce derivative image files as needed for a variety of current and future user needs.
- Avoid proprietary formats such as camera RAW or JPEG as a master version. Create multiple backup copies of all files on a stable medium. We suggest acquiring two external hard drives, and storing one off site (OWLS). CDs and

DVDs have been found to degrade over a relatively short time. Store media in an appropriate environment.

- Monitor and recopy data as needed.
- Outline a migration strategy for transferring data across generations of technology.
- Anticipate and plan for future technological developments.

Project Workflow

Combined documentation from OWLS, Recollection Wisconsin, and CCDC Workshop

Workflow Overview

Please do not hesitate to contact OWLS for support at any stage in the process.

1. Read guidelines, review online resources, and attend training as available. OWLS will provide training as staff time allows.
2. Select your first collection and document copyright clearance.
3. Define your new collection - think about the items you have chosen: what you know about the items (their metadata), and what info is applicable to all the items in the set (perhaps a copyright statement or note about where you obtained the images, i.e., the source).
4. Determine your file naming convention for the items you'll be scanning.
5. Choose which optional Metadata Fields you want to use in addition to those that are Required and Required if Available.
 - If more than one person will be assigning metadata, create a Metadata Dictionary for your collection.
 - Consult the appropriate controlled vocabularies to choose standard terms.
6. Contact the ContentDM Administrator at OWLS and provide to them:
 - The title of the collection.
 - A one-paragraph description of the collection for use on the website.
 - The metadata fields and local controlled vocabularies you want to use.
 - Sample records for 2-10 images.
7. Scan images according to Recollection Wisconsin guidelines, and assign them the appropriate file names as you save the files.
 - Have a clean, dedicated work area (no food or drink by photos).
 - When scanning several pages of a single document, use the same scanner settings for all pages, even if some have images and others do not.
 - While scanning, regularly stop and check the images for quality and consistency. Be especially vigilant when more than one person is involved with scanning.
 - File Types and Image Editing
 - Scan as TIFFs - archival version; these will be huge in file size
 - Optional JPGs - The software will convert your TIFFs to access JPGs at upload; but you can create your own set of access JPGs at this time if you anticipate needing them.

- Keep editing to a minimum, but complete the following steps as needed for quality control.
 - Straighten / Image Rotation
 - Crop
 - Adjust Histogram
 - Sharpen
 - Create transcript files in .txt format if applicable.
- 8. Complete the Excel metadata template appropriate to the item type you have scanned (available on OWLSnet).
- 9. Send your image files, transcript files, and metadata file to OWLS for uploading.
- 10. The OWLS CONTENTdm Administrator will check your files for quality control and compliance with guidelines and will contact you with any clarifying questions.
- 11. The OWLS CONTENTdm Administrator will upload, approve, and index the collection, and notify you when it is published.
- 12. Review the published collection and contact OWLS with any questions or corrections.
- 14. Link to your collection landing page from your library's web site and promote your collection to your community.

Standards for Scanning

Wisconsin Heritage Online Digital Image Guidelines

<http://recollectionwisconsin.org/wp-content/uploads/2017/08/WHO-Scanning-Quick-Guide.pdf>

Minimum Specifications and Resolution Chart

Original Material	Scanning resolution	Bit depth	Approximate scanned dimensions	Approx. size of preview image
<i>Photographs</i>				
16"x20" +	200 ppi	24-bit color	6400x8000 pixels	146 MB
8 ½"x11" - 16"x20"	300 ppi	24-bit color	3200x4000 pixels	36 MB
8"x10"	400 ppi	24-bit color	3200x4000 pixels	36 MB
5"x7"	625 ppi	24-bit color	3200x4000 pixels	36 MB
4"x5"	800 ppi	24-bit color	3200x4000 pixels	36 MB
4"x2 ½"	1200 ppi	24-bit color	3200x4000 pixels	36 MB
<i>Text</i>				
Print—no images	600 ppi	1-bit bitonal	Varies	Varies
Print—with images	300 ppi	8-bit grayscale or 24-bit color	Varies	Varies
Manuscripts	400 ppi	8-bit grayscale or 24-bit color	Varies	Varies

Scanning Notes

- Save all scans as TIFFS with NO compression
- Preview scanned images before adding calibration chart to see if pixel dimensions and file size are within recommended ranges. Then open scanning selection to include calibration chart.
- If you wish to reduce file size, you can convert from color to grayscale or from grayscale to bitonal after scanning.

Scanning Chart Rationale

Scans that meet these recommendations will produce a print of reasonable quality up to 20"x24" and should be sufficient for most publishing requirements. However, they are not intended to serve as preservation files to replace the original materials.

Image Capture Guidelines

Choose a wide gamut

Use a wide-gamut RGB color space for capturing and editing RGB master files. Scanning software allows a choice of color for the resulting files. A wide-gamut space for print (such as Adobe RGB) and the smaller sRGB color space for web are the standards. sRGB is the default color space of Windows XP, web browsers, numerous image browsers, image editors and other software on the Windows platform. If your scanner doesn't include an option for choosing a color space, it's almost certainly sRGB. sRGB is a small gamut color space and you will shortchange your images where a wider gamut is supported. **Use Adobe RGB, not sRGB.**

Scan preparation

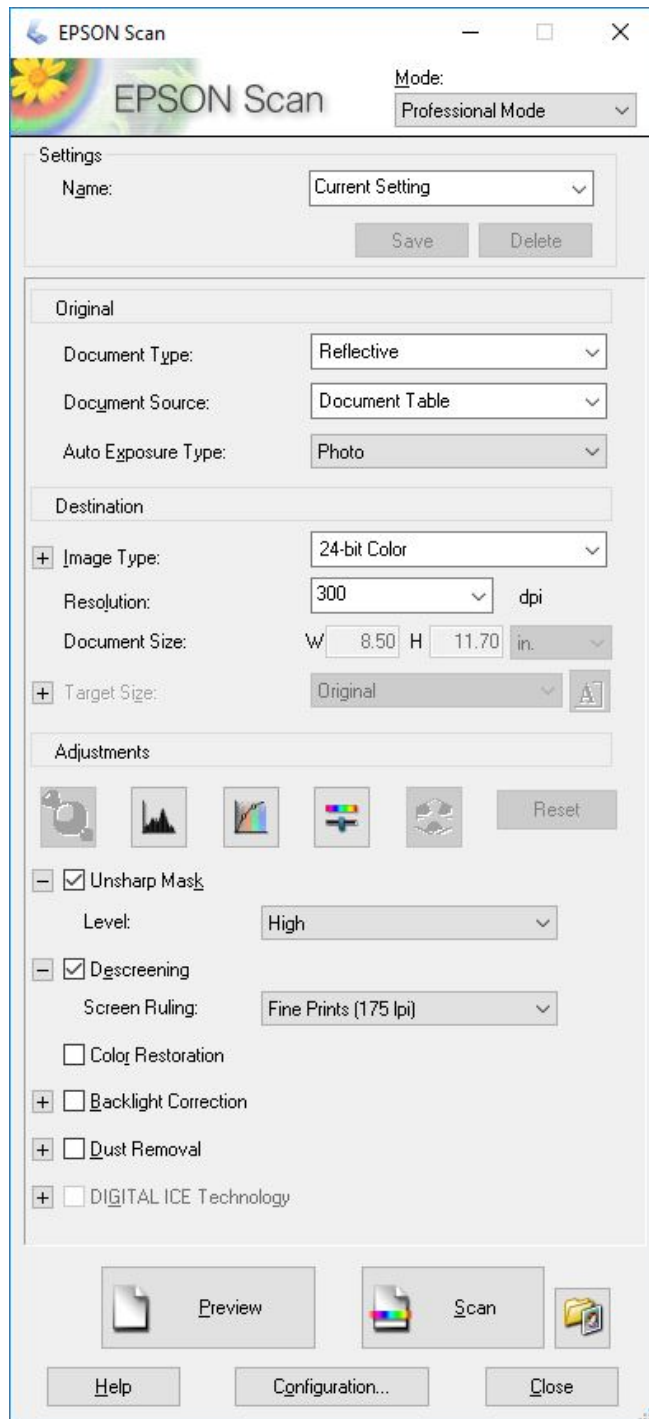
The scanning hardware and software must be configured to ensure optimal image capture and image processing. Digital files should be evaluated by reviewing the image's histogram, which illustrates how pixels in an image are distributed by graphing the number of pixels at each color intensity level rather than by relying on how the image looks on a monitor. In a histogram, each pixel in an image file is assigned a numeric value between 0 and 255 with 0 representing black and 255 representing white. Do NOT use the default automatic settings when scanning, which will set the lighted area to 255 and the darkest area to a value of 0. This eliminates information in the highlights and shadow areas.

- Clean the scanner bed (the glass)
- Place the image and calibration chart on the scanner
- Prescan
- Fit the marquee (box) around the image and the calibration chart.

Scan the image

Open "EPSON Scan" (or whatever scanner your site has)

1. Make sure that "Professional Mode" is selected
2. Document type: Reflective
3. Change image type and resolution to reflect your needs for digitization (see chart at bottom)

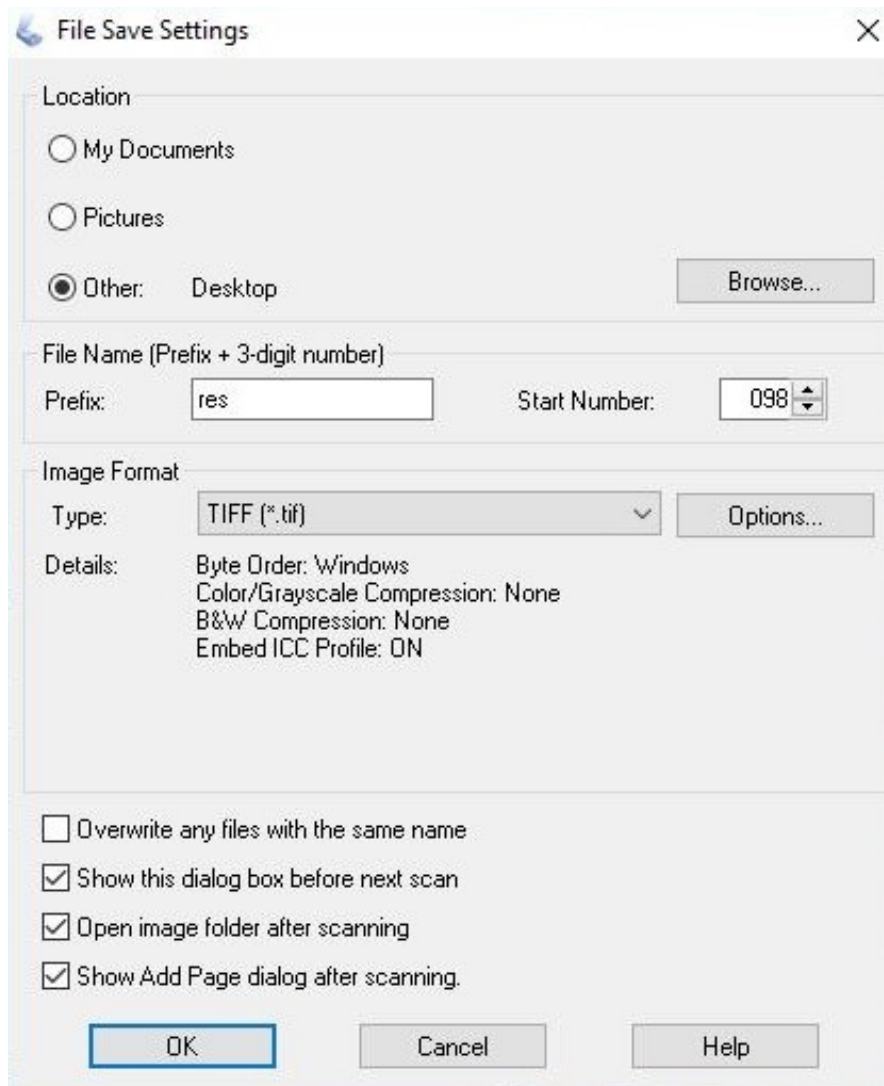


4.

Next,

- Select the folder image located next to “Scan”
- From here, you can locate where you want the files to be scanned to
 - Recommendation: Create a special folder for the collection you are digitizing, this will make sure no items get misplaced and will take up less time in the long run since there will be no need to move files after
- Select what prefix you want and start number.

- Try to keep page numbers of the physical item the same as what you saved the digital image as. For example, you digitized page 1 of a collection, the file name would be owls_exa_001
- For covers and other pages that do not contain a page number, add letters of the alphabet. For example, you digitized the cover of a collection, the file name would be owls_exa_00A
- For any questions about naming conventions, please see “File Naming” on page 11



File Save Settings

Location

☐ My Documents

☐ Pictures

☒ Other: Desktop Browse...

File Name (Prefix + 3-digit number)

Prefix: Start Number:

Image Format

Type: Options...

Details: Byte Order: Windows
Color/Grayscale Compression: None
B&W Compression: None
Embed ICC Profile: ON

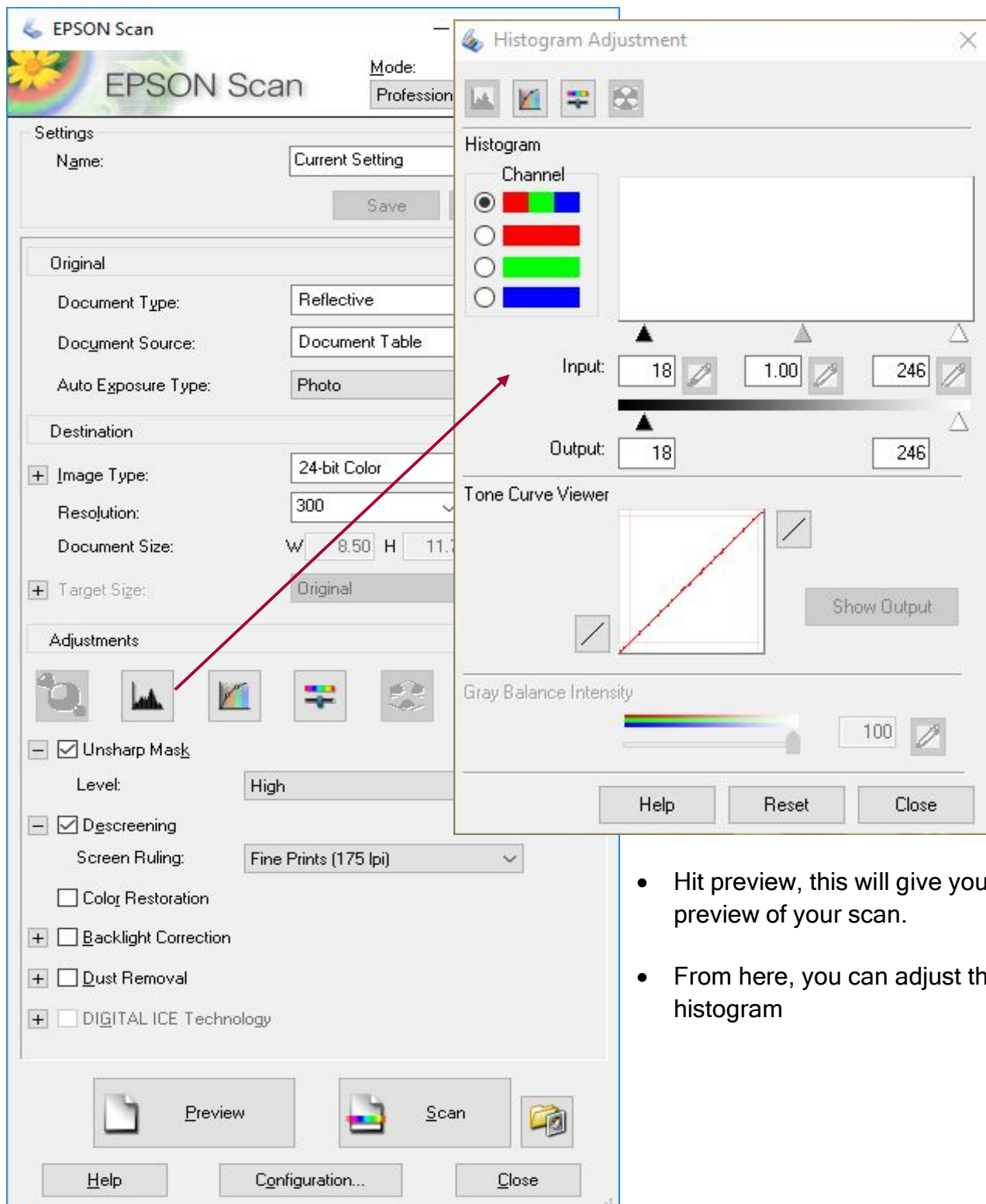
☐ Overwrite any files with the same name

☒ Show this dialog box before next scan

☒ Open image folder after scanning

☒ Show Add Page dialog after scanning.

OK Cancel Help



- Hit preview, this will give you a preview of your scan.
- From here, you can adjust the histogram

- Ideally, the histogram should be like a hill. Here is an example of what you want your histogram to look like



- When adjusting the histogram, make sure to only adjust on the ends of the histogram
- Once you have adjusted this, hit close
- Now you are ready to scan!
- Remember to always do quality review checks as you are scanning. This will help make sure that you don't have to re-do an entire project, or come back at a later point and rescan. This is also idea since you already have the settings set, and this will help make sure the images are uniform.

Capture Standards

UW-Madison Digital Collections

<https://www.library.wisc.edu/digital-library-services/uwdcc/resources-documentation/reformatting-standards/>

Flatbed Scanning

- 300 dpi for text-based material (EFacs)
- For photographic material aim for 2000-5000 pixels on the longest edge of the images. This generally means:
 - 300 dpi for material larger than 5" x 7"
 - 600 dpi for material smaller than 5" x 7"
- 24-bit color
- All adjustments on the Epson Scan utility are unchecked.
- Histogram settings are created for each machine based on performance assessment using the GoldenThread Analysis software
- TIFF images

High Speed Scanning

- 300 dpi (color/grayscale) or 600 dpi (bitonal)
- 24-bit color, 8-bit grayscale, or 1-bit bitonal depending on material
- TIFF images

Metadata

Metadata is commonly defined as "data about data." Metadata is frequently used to locate or manage information resources by abstracting or classifying those resources or by capturing information not inherent in the resource. Typically, metadata is organized into distinct categories and relies on conventions to establish the values for each category. For example, administrative metadata may include the date and source of acquisition, disposal date, and disposal method. Descriptive metadata may include information about the content and form of the materials. Preservation metadata may record activities to protect or extend the life of the resource, such as reformatting. Structural metadata may indicate the interrelationships between discrete information resources, such as page numbers.

(<https://www2.archivists.org/glossary/terms/m/metadata>)

Descriptive metadata for digital cultural heritage collections aim to answer some basic questions for uses:

- What am I looking at?
- Who created it?
- When was it created?
- Where was it created?
- Who owns it?
- How can I use it?

Metadata for Recollection Wisconsin and DPLA

Each item contributed to Recollection Wisconsin must include for descriptive metadata fields:

- Title
- Subject
- Type
- Rights

Recollection Wisconsin strongly recommends including four additional metadata fields when possible:

- Date
- Description
- Creator
- Place

For more information on detailed guidelines and examples for each required and recommended metadata field, visit Recollection Wisconsin Metadata Essentials.

(<https://recollectionwisconsin.org/wp-content/uploads/2016/12/RecollectionWisconsinMetadataEssentials.pdf>)










Exactly

What is Exactly?

Exactly is a free and open source application that is easy-to-use for remotely and safely transferring any digital data from a sender to a recipient.

Exactly allows recipients to create customized metadata templates for senders to fill out before submission. Exactly can send email notifications with transfer data and manifests when files have been delivered to the archive.

So, what is in an Exactly BagIt file?

Name	Date modified	Type	Size
 data	7/25/2018 12:21 PM	File folder	
 bag-info.csv	7/25/2018 12:21 PM	OpenOffice.org X...	1 KB
 bag-info.txt	7/25/2018 12:21 PM	Text Document	1 KB
 bag-info.xml	7/25/2018 12:21 PM	XML Document	1 KB
 bagit.txt	7/25/2018 12:21 PM	Text Document	1 KB
 FileSystemData.txt	7/25/2018 12:21 PM	Text Document	12 KB
 manifest-md5.txt	7/25/2018 12:21 PM	Text Document	16 KB
 tagmanifest-md5.txt	7/25/2018 12:21 PM	Text Document	1 KB
 TransferComplete.txt	7/25/2018 12:21 PM	Text Document	1 KB

Exactly creates a bag of checksums, manifest, bag info, and the images you digitally preserved under “data.” By using Exactly, this allows Fixity to run the checksums created against the digital collections to make sure the collections are holding their integrity. By holding their integrity, this means that they are not becoming corrupt, or changing. This allows us to make sure that items will be seen the same in 50 years as they are today.

This will serve as a guide on how to run the Exactly software. If you have any questions during the process, please contact Exactly:

<https://www.weareavp.com/contact/>

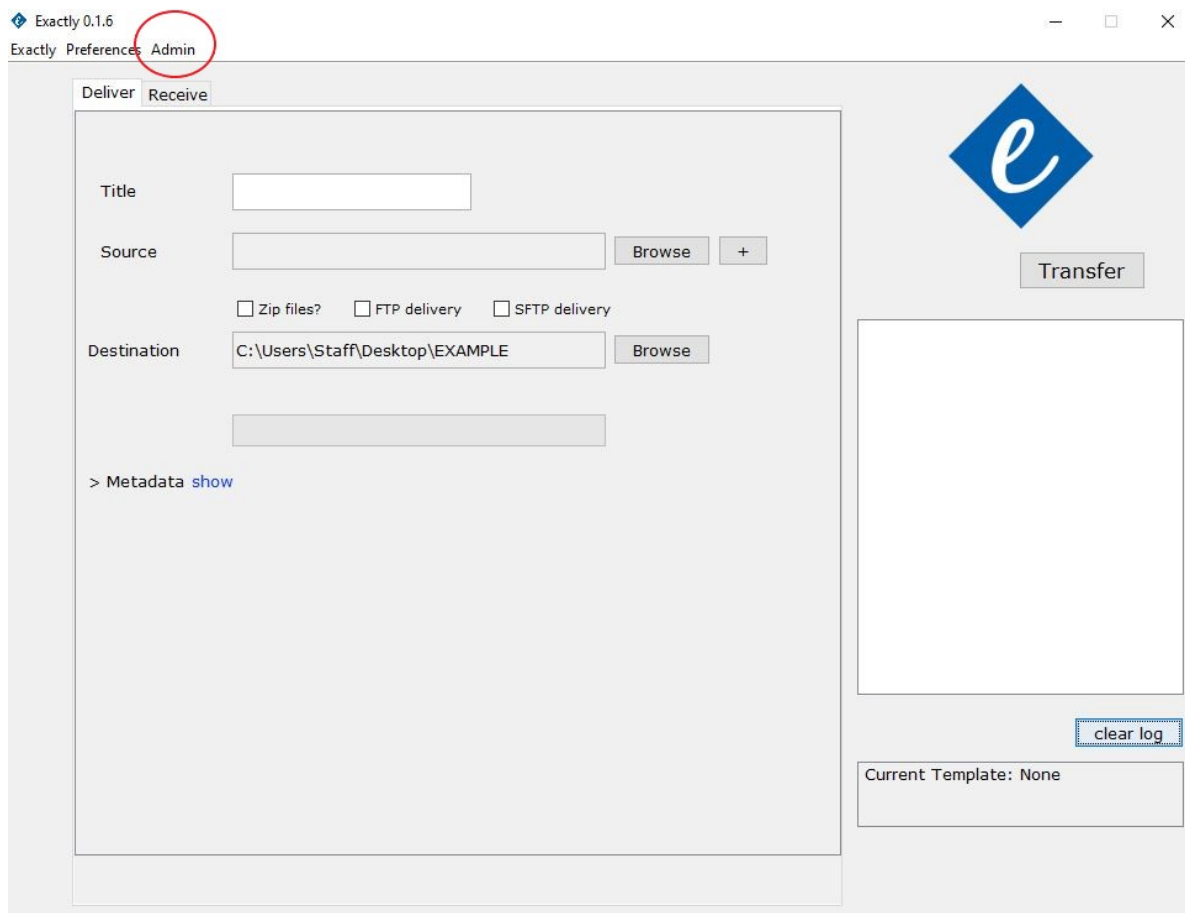
Email Settings - Admin Menu (important)

Administrative settings should be created by the receiving entity and included in preset configurations. If the receiving entity has not provided settings in the preset configurations, see below.

Email Settings

- Enter in email “send” settings in order for email notifications to be sent. You may need to consult your email provider for appropriate settings for sending email from the application.
- The “Email Notifications” checkbox must be enabled in order for notifications to be automatically sent by Exactly.

You may have to adjust your security settings on your email account (for example with Gmail) in order to allow third party applications to send email on your behalf.



The screenshot shows the 'Exactly 0.1.6' application window with the 'Admin' menu item circled in red. The 'Admin' menu is open, showing 'Deliver' and 'Receive' tabs. The 'Deliver' tab is active, displaying a form for configuring email delivery settings. The form includes fields for 'Title', 'Source', and 'Destination'. The 'Source' field has a 'Browse' button and a '+' button. The 'Destination' field has a 'Browse' button. There are checkboxes for 'Zip files?', 'FTP delivery', and 'SFTP delivery'. A 'Transfer' button is located on the right side of the form. Below the form, there is a 'clear log' button and a 'Current Template: None' label. The 'Exactly' logo is visible in the top right corner of the window.

Exactly 0.1.6
Exactly Preferences Admin

Email Settings

Mail Server:

User name:


Password:

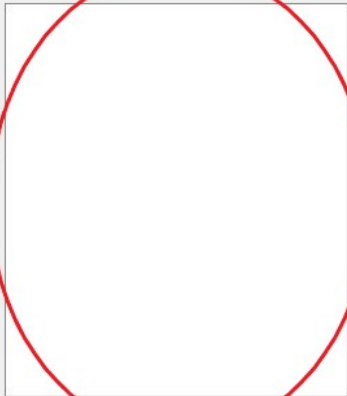
Port: ☒ Enable email notifications

Encryption Method

☐ SSL Protocols ☒ TLS Protocols ☐ None

Provide smtp configurations for sending email.





Current Template: None

The white box will let you know if something was successful or not. Reminder, the server settings that come preset work well with Gmail, however will need to be adjusted for other emails.

Package Delivery

Here is where you will provide basic (required) package delivery information or package-level metadata in order to successfully transfer your files. Consult with the archive or other receiving entity on appropriate conventions and destinations.

First, make sure that you have selected the “Deliver” tab.

1. Title: You must provide a title for your package. Example: Shawano
2. Source: Locate the file that you would like to transfer by clicking “Browse.”
3. Destination: Where you want the Bag to be located. For example, moving a digital collection from a computer, to an external hard drive.

The screenshot shows the 'Exactly 0.1.6' application window. The 'Deliver' tab is active, and the 'Receive' tab is also visible. The interface includes the following elements:

- Title:** A text field containing 'Shawano'.
- Source:** A text field containing 'Jusers\Staff\Desktop\SampleExactly\Shawano', a 'Browse' button, and a '+' button.
- Destination:** A text field containing 'C:\Users\Staff\Desktop\EXAMPLE', a 'Browse' button, and an empty text field below it.
- Options:** Three checkboxes labeled 'Zip files?', 'FTP delivery', and 'SFTP delivery', all of which are currently unchecked.
- Metadata:** A link labeled '> Metadata show'.
- Transfer Quad (Right Side):**
 - A large blue diamond logo with a white 'e'.
 - A 'Transfer' button.
 - A large empty rectangular box.
 - A 'clear log' button.
 - A status bar showing 'Current Template: None'.

4. Before hitting transfer, add any metadata that is applicable.
5. Transferring files: Using Exactly to transfer files is easy. Once you have the application configured, and your metadata is complete, all you have to do is to press the “transfer” button in the Transfer Quad (upper right).
 - The status bar and log (right) will tell you the status of your transfer.
 - When your transfer is completed, the log will indicate “transfer completed successfully.”
6. Email notification of successful transfer.

Exactly Digital Transfer



kristinarwarner@gmail.com

Wed 7/25/2018 12:21 PM

To: Kristina Warner ↗



Shawano-20180725-ma...
23 KB



Download Save to OneDrive - Outagamie Waupaca Library System

Transfer completed: Wed Jul 25 12:21:51 CDT 2018

Transfer Name: Shawano

Target: C:\Users\Staff\Desktop\EXAMPLE\Shawano\Shawano

Application Used: Exactly

User: kristinarwarner@gmail.com

Total file count: 1

Total Bytes: 901594005

Payload Oxum: 901594005.144

Bagging Date: 2018-07-25

Bag Size: 880462.9 KB

Source-Organization:

Organization-Address:

Contact-Name:

Contact-Phone:

Contact-Email:

External-Description:

External-Identifier:

Bag-Group-Identifier:

Bag-Count:

Internal-Sender-Identifier:

Internal-Sender-Description:

Within the email notification, you will also get the manifest for the Bag. This also can be found within the Bag you created.

Fixity

From UW-Milwaukee

Configuration

- Download Fixity from AVPreserve (www.avpreserve.com/avpsresources/tools/)
- From main screen, Preferences>Email Settings
 - SMTP Server: smtp.outlook.com
 - Email Address: the email FROM which you want fixity alert emails to originate. This should be a generic department account where possible to maintain business continuity.
 - Enter password to email account
 - Port: 587
 - Encryption Method: TLS protocols
- Click “Check Credentials” to send a test email, which will verify you’ve entered the settings correctly

Creating and running a project

- From the main screen, File>New Project. Project name cannot contain spaces or special characters
- Select one or more directories to scan for fixity. You can select multiple directories in case you want to scan some folders within a larger directory but not others
- Enter one or more email addresses to which scan results should be sent.
- Select when and how often you want to scan the directory
- Check “if missed, run on restart” to ensure the scan still happens if you’re not in the office when you schedule it
- Optimally, check “Email only upon warning or failure” to limit the amount of email you get from routine checks
- File>Save settings to save the project
- File> Run Now to run the fixity check for the first time

Receiving and interpreting results

There are two sets of files created as the result of Fixity’s check: the Reports and the History. Both files are saved as TSV files.

Reports

Once a check has finished running, the email that is sent will link to the report file for that check. Otherwise you can find them in your fixity directory under “/reports/”

- The first few lines are summary information, which tell you how many of each “category” of file there are in check:



- “Confirmed” files are in the same location as the previous check, and content has not been changed.
 - “Moved or Renamed” files are either in a different location or have been renamed since the last check, but the content is substantially the same based on the checksum
 - “New Files” have been added to the collection since the last checksum
 - “Changed files” have changed content (based on the checksum) since the last check
 - “Removed files” were not found by Fixity in the expected or any other location
- For the first report, run on a collection, all files should read as “new”
 - For subsequent reports, make note of any files in statuses above
 - In general, statuses of “Changed” and “Removed” are usually most serious from a file integrity standpoint

History

The History files contain the raw checksums computed by Fixity, from which the program derives the report by comparing one History File with another. In general, you will have no need to access this file, but it can be useful for comparing checksums independently with another tool, if necessary from particularly valuable files. The History files can be found in your Fixity directory under “/history/”

- The header of the History file indicates the directory against which Fixity was run, the time of the specific report, and the checksum algorithm used (Fixity defaults to SHA256, but MDS is another common algorithm.)
- The list of files includes the checksum and full path to the file to which it belongs. You can independently verify the fixity of a file by copying checksums from the same file from different history files to a 3rd party checksum comparison tool, such as Checksum Compare

For more information, please see “How-to-Guide: Fixity”

Tools

Advance File Renamer

<https://www.advancedrenamer.com/>

Exactly

<https://www.weareavp.com/products/exactly/>

Fixity

<https://www.weareavp.com/products/fixity/>

Checksum Compare

<http://checksumcompare.sanktuaire.com/home>

Appendix

Audit

Process of systematic review and validation of storage environment, for example, the analysis of monitoring logs.

Authenticity

The quality of trustworthiness of a record – in this context a digital object. Authenticity provides the assurance that a record is what it purports to be and has demonstrably not been tampered with or otherwise corrupted, Authenticity arises from the reliability, integrity, and usability of the record.

Backup

Static copy of data, usually for data recovery purposes.

Checksum

Typically expressed as a text string or hash value, checksums are outputs generated by an algorithm and compactly express the data in a file or other data block. Checksums can be used to detect errors or changes to files stored on a computer or transferred from one computer to another since any change to the bit order of a file will result in a change to the file's checksum. Checksums are also known as file hashes or message digests.

Digital object

An entity which is the target for preservation. Depending on context, it may refer to an information object or data object.

Digital preservation

The process of maintaining a digital object for as long as required, in a form which is authentic, and accessible to users.

Fixity

“Unchangedness” of data, usually evidenced by identical and persistent checksums generated from the same file over time.

Hard disk drives

A form of magnetic media that have magnetic platters that are read by arms that spin.

Migration

The process of converting data from an obsolete structure to a new structure to counter software obsolescence.

Monitoring

Logging or recording various aspects of the storage configuration, including hardware, activity, and data integrity.

Redundancy

Creation and retention of multiple near-identical copies of the same data.

Replication

Automated copying of data from one primary storage location to another or several other storage locations. Replication is distinct from redundancy in that it dynamically updates the secondary storage locations.