GUIDELINES FOR

CREATION AND PRESERVATION OF DIGITAL FILES

October 2018

WYOMING STATE ARCHIVES

INTRODUCTION

This document provides guidelines for the creation and preservation of digital files. They pertain to both born-digital records and digitized copies of records in an analog format (e.g., paper or photos) and provide information on digitization, file formats, file naming, storage, and preservation.

The decisions you make regarding how to create, name, and store digital files will affect your agency's ability to preserve them for long-term access and use. The result of balancing image quality and storage capacity when digitizing records should be increased access to information at an affordable price. Selecting nonproprietary file types when possible will reduce the risk of software obsolescence over time. Employing a consistent file naming system and metadata schema will help you find records quickly. Using modern storage media with a robust backup plan in place and developing a preservation strategy will help protect and maintain your files. Taken together, these steps will greatly increase your agency's capacity to manage, access, and preserve its digital files.

State Archives Services for State Agencies

These guidelines have been written with organizations that digitize and store their own records in mind. However, the State Archives has services available to assist state agencies in digitization and digital records preservation. For more information about these services, please contact the State Archives.

DIGITIZATION

Government agencies digitize records to increase access, streamline workflows, and reduce the need for physical storage space. Digital files made available over the web allow government agencies to provide information to partners or the public quickly and efficiently. In addition, when optical character recognition (OCR) software is used, digital images can be text-searchable, which makes information easier to find.

While digitization can save agencies time in accessing records and money in storage, it is an investment. Scanners, scanning software, and storage media are required up-front and should be updated on a routine basis. Keeping your software and equipment current is important to the long-term preservation of your records and will help ensure a trustworthy management and storage environment for as long as the records retention schedules require. Agencies sometimes ask us if they can digitize (or scan) their paper records and then discard the paper copies. The Uniform Electronic Transactions Act (Wyoming State Statutes 40-21-101 through -119) allows "[e]ach governmental agency [to] determine whether, and the extent to which, a governmental agency will create and retain electronic records and convert written records to electronic records."

As long as the agency meets the requirements set forth in the Uniform Electronic Transactions Act and ETS Rule 5, they can digitize records and discard the paper copies. However, there is an additional requirement for records scheduled for permanent retention. Records scheduled for permanent retention may be digitized, but scanned copies of the records must be deposited in the State Archives' Digital Archives before the paper can be discarded. This fulfills the requirement in Statute 9-2-406(a) that the Director of the Department of State Parks & Cultural Resources maintain and secure all public records and the requirement in Statute 9-2-411 that calls for agencies and political subdivisions to obtain approval from the State Records Committee prior to destroying permanent public records.

Terms

Digitization: A process by which a document or photo is scanned and converted from analog format to a computer-readable digital format. After scanning, the document or photo is represented by a series of pixels arranged in a two-dimensional matrix called a bitmap or raster image. This image can then be kept on a network for storage and use.

Pixel Bit Depth: Pixel bit depth refers to the number of bits used to define each pixel. The higher the bit depth, the more tones (color or grayscale) can be represented in a digital image. Digital images can be bi-tonal, grayscale, or color. In general, higher bit depths are recommended for master images to accurately represent the original document.

Standard pixel bit depths

Bit-depth	Displays	Recommended for
1-bit or "bi-tonal"	black and white	Typewritten documents
8-bit grayscale	256 shades of gray	Black and white photographs, half-tone illustrations, handwriting
24-bit color	Approximately 16 million colors	Color graphics and text, color photographs, art, drawings, maps

Resolution: The quality of a digital image is dependent upon the initial scanning resolution. Resolution refers to the number of dots, or pixels, used to represent an image, expressed commonly as "dpi," dots per inch. You may also see the terms "ppi" (pixels per inch) and "lpi" (lines per inch) used. As the dpi value increases, image quality increases, but so does the file size.

Recommendations

The desired image quality and the storage capacity of your computer system play large roles in determining what pixel bit depth and resolution to use. The greater the bit depth and resolution, the more storage space the scanned image will require. Larger images take longer to deliver over the Internet, something to consider if that is a service you provide. If online access is important to your agency, you may want to scan high-resolution masters for long-term preservation and lower resolution copies for web delivery.

In most cases, the State Archives recommends scanning standard black and white documents bi-tonal at 300 dpi. The size and quality of the original document may affect how we scan, but that is our usual resolution. Please see the table below for recommendations on scanning photographs and other record types.

Common Scanning Resolutions for Master Files

Material	Recommended resolution (8-bit grayscale and 24-bit color)
Textual records	300-600 dpi
Photographs, negatives, slides	4000-8000 pixels in long dimension
Material	**Resolution used by State Archives (8-bit grayscale and 24-bit color)**
Textual records	300 dpi
Photographs, negatives (4"X5" or larger)	600 dpi
Smaller Photographs, slides, negatives (ex. 35 mm, 120, 220, etc)	600 dpi and target size set to 8" in long dimension

Because the standards for digital audio and video are complex and quickly changing, please consult the Federal Agencies Digitization Guidelines Initiative website or contact the Wyoming State Archives for more information.

- Federal Agencies Digitization Guidelines Initiative:

http://www.digitizationguidelines.gov/

- Council of State Archivists Minimum Digitization Capture Recommendations

https://www.statearchivists.org/resource-center/resource-library/minimum-digitization-capture-recommendations/?ccm_paging_p=12

FILE FORMATS

The file format used to create and store your content is a primary factor in their future viability and usage. Technology continually changes and all contemporary hardware and software should be expected to become obsolete over time.

Consider how your data will be read if the software used to produce it becomes obsolete and how best to manage and share your data for future use. File formats created with these principles in mind are more likely to be accessible in the future.

- Non-proprietary
- Open, documented standards
- Unencrypted
- Uncompressed, if space is available

Examples of preferred formats

(see Digitization section for conversion of analog content)

File Type	Preferred Format
Image	jpeg, jpeg-2000, tiff
Text	txt, html, xml, PDF or PDF/A Open Office XML
Audio	afif, wav
Video	mp4, avi
Databases	xml or convert to csv

Examples of proprietary formats and alternatives

Proprietary Format	Alternative Format
Excel (.xls, .xlsx)	Comma Separated Values (.csv)
Word (.doc, .docx)	Plain text (.txt) or, if formatting is needed, PDF or PDF/A
PowerPoint (.ppt, .pptx)	PDF or PDF/A
Photoshop (.psd)	Tiff
QuickTime (.mov)	mpeg-4 (.mp4)

These are examples of commonly used proprietary formats. For long-term accessibility, consider generating a copy in one of the preferred formats listed in the previous section. For advice on generating these copies, contact your IT staff or the State Archives.

The following links provide more information on format descriptions and their characteristics:

- Library of Congress' Sustainability of Digital Formats:

http://digitalpreservation.gov/formats/fdd/descriptions.shtml

- Council of State Archivists File Format Comparison Projects:

https://www.statearchivists.org/resource-center/resource-library/guidelines-file-format-comparison-projects/?ccm_paging_p=9

FILE NAMING

If you create and follow a specific strategy for how you name original files, you will be able to more easily identify, locate and share those files. Ideally, members of your organization should be able to look at a record's file name and use that information to recognize the contents and characteristics of the record and make decisions about it.

When developing your file naming policy, you may wish to include some of the following elements:

- Create unique file names. Duplicate file names will cause confusion.
- File names should be simple and easy to understand.
- Avoid using special characters such as: ? / $ % & # . \ : < >
- Use underscores (_) and dashes (-) to represent spaces.
- Use leading zeros with the numbers 0-9 to facilitate proper sorting and file management.
- Dates entered in this format will remain in chronological order: YYYY_MM_ DD or YYYYMMDD. Variations include YYYY, YYYY-MM, YYYY-YYYY.
- Keep the file name as short as possible and always include the three character file extension (e.g., .jpg or .doc).
- Include the version number in the file name by using 'v' or 'V' and the version number at the end or beginning of the document. (e.g., 2014_Notes_v01.doc). Avoid using the words "version" or "draft"

METADATA

Metadata is used to describe a record, its relationships with other records, and how the record has been and should be treated over time. Metadata often includes items like file type, file name, creator name, and date of creation. Metadata enables proper data creation, storage, and retention. In addition, standardized metadata helps validate the trustworthiness of your recordkeeping system and the legal admissibility of your digitized records in court.

There are two commonly used approaches to storing metadata. Metadata can be stored separately from the digital files in a database or it can be embedded in a digital file. Most software applications automatically create metadata and associate it with files, generally making the standardization of metadata simpler. One example of automatic and standardized metadata is the header and routing information that accompany an e-mail message. Another is the set of properties created with every Microsoft Word document; certain elements such as the title, author, file size, etc., are automatically created, but other elements can be customized and created manually. By standardizing the process it will be easier to manage, access, and preserve the files long-term. Normally, some combination of automatically and manually created information is best for precise and practical metadata.

The Dublin Core metadata schema, made up of 15 "core" elements has emerged as one of the basic means of creating metadata about resources that can be shared widely.

For more information see the following resources:

- Dublin Core Metadata Initiative: http://dublincore.org
- Federal Geographic Data Committee (FGDC). Metadata: http://www.fgdc.gov/metadata

STORAGE OF DIGITAL FILES

It is critical to store master digital files in a way which insures that back-up copies are secure, tamper proof and available if needed. The Wyoming Department of Enterprise Technology Services include the following recommendations in their Data Backup, Storage and Restoration Guidelines:

- Primary backup schedule: The primary backup should occur at least daily
- Offsite backup: At least one set of the backup should be stored at a sufficient distance away to escape any damage from a disaster at the main site. (The Archives recommends at least 45 miles away)
- Backup testing: Restoration of backup data should be performed and validate at least every six month.

The Wyoming Department of Enterprise Technology Services has also implemented a broader Data Backup, Storage and Restoration Policy. Both the policy and the guidelines apply to all state agencies as defined in Statute 9-2-2904(a)(i).

PRESERVATION STRATEGIES

Once you have decided on a file format and a storage plan, the challenge will be to keep those files accessible and viable. There are two, often compatible approaches for long-term electronic record preservation:

- Conversion. When you convert a record, you change its file format. Often, conversion takes place to make the record software available in an open or standard format. For example, you can convert a record created in Microsoft Word by saving it as a Rich Text Format (RTF) file or to PDF/A.
- Migration. When you migrate a record, you move it from one computer platform, storage medium, or physical format to another. For example, you may need to migrate records from old magnetic tapes to new ones or to a different medium entirely to ensure continued accessibility.

FOR MORE INFORMATION

For more information about these guidelines, the Digital Archives, or digital records in general, please contact the Wyoming State Archives.

Wyoming State Archives
Barrett Building
2301 Central Avenue Cheyenne, WY 82002
(307) 777-5586