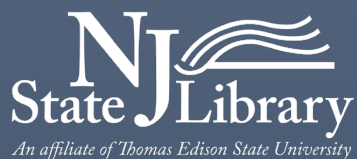


2020



New Jersey State Library

Digital Preservation Policies and Practices

Table of Contents

Overview 2

Purpose 3

Collection Policy.....4

Opportunities and Challenges.....5

Principles.....6

Collaboration.....7

Technical Aspects.....7

Sustainable File Formats.....9

 Preferred File Formats for Submitting Items to the New Jersey State Publications Digital Library.....14

Best Practices for File Naming.....15

 Preferred File Naming for the New Jersey State Publications Digital Library16

Copyright and New Jersey State Government Documents.....17

Further Resources..... 19

ATLANTIC COUNTY. 11

Atlantic county was formed from the townships of Galloway, Hamilton, Weymouth and Egg Harbor, Gloucester county, 1837, p. 96.
A portion of Hammonton was annexed to Burlington county 1902, p. 403.

Note.	MUNICIPALITIES.	Incorporated.	Page of Date.	FROM WHAT TAKEN.	1840.	1850.	1855.	1860.	1865.	1870.	1875.	1880.	1885.	1890.	1895.	1900.	1905.
	Absecon, city.....	1992	60	Formerly town of Absecon.....													616
	Absecon, town.....	1872	301	Galloway and Egg Harbortwps.....							626	567	567	501	522	530	
	Atlantic City.....	1854	278	Absecon Beach.....			204	687	746	1043	2009	5477	7942	18255	18329	27838	37593
	Brigantine City.....	1897	281	Brigantine Beach.....											188	99	95
	Buena Vista, township.....	1867	142	Hamilton township.....						948	837	885	1016	1299	1424	1646	2524

Compendium of censuses 1726-1905: together with the tabulated returns of 1915. This popular resource allows researchers to see the growth of counties and municipalities within New Jersey.

OVERVIEW

The New Jersey State Library's direct services are designed to meet the diverse and ever-changing information needs of state government, Thomas Edison State University staff and students, and New Jersey residents. Reflecting that need, this policy provides guiding principles regarding the operation, management, and scope of the New Jersey State Library digital preservation and access program, established by the New Jersey State Library in accordance with N.J.S.A 52:14-25.1 which authorizes the New Jersey State Library to collect and preserve state government publications for public access and preservation.



52:14-25.1 Annual or special reports, publications; copies filed in State Library.

52:14-25.1. All State officers, departments, commissions, committees, or agencies issuing annual reports or special reports required by law to be submitted to the Governor or to the Legislature of this State, and other State publications of a general informational character, where such reports are printed and electronically produced, shall file with the New Jersey State Library for purposes of permanent public access and distribution one electronic copy and six printed copies. In cases where such reports are made in electronic form only, one electronic copy and one printed copy shall be submitted to the State Library for preservation and permanent reference use. State officers, departments, commissions, committees and agencies shall designate an individual to act as a liaison to the State Library.

amended 1957, c.99, s.1; 1967, c.162, s.1; 2011, c.184, s.6.

Adopting strategic and technical measures for the long-term preservation of New Jersey's digital state government assets, primarily digital content produced by state government and its departments and authorities allows the State Library to ensure that the historical legacy of state government in digital form will be available to future residents, researchers, lawyers and the judiciary. The adoption of the following policies to capture, preserve, and make accessible the digital record of New Jersey's government will ensure the administrative integrity, information authenticity, government transparency, and legislative compliance of the State's digital assets. This policy outlines the State Library's strategic approach to minimizing the risks associated with technological changes and ensuring that the records remain reliable and accessible over time.

PURPOSE

The digital preservation program is intended to preserve, maintain, and make accessible state government publications both “born digital”, and those transformed to a digital format through digitization.

The primary objectives of the digital preservation program are to:

- Preserve digital New Jersey publications that are either born digital or digitized. Retain a copy of the original bit stream of born digital material for authenticity purposes, as well as potential access in the future.
- Where necessary, migrate records into formats that are acceptable for long-term preservation and access.
- Make strategic decisions, based on long-term resources, system sustainability, and identified efficiencies, to establish processes and technologies that support long-term preservation.
- Ensure record authenticity through preservation and technical solutions that promote and maintain composition (especially for assets with multiple objects or files) fixity, and validity.
- Establish operational processes and procedures to meet archival requirements pertaining to provenance, chain of custody, authenticity, and integrity.
- Provide tools that facilitate the discovery of online digital assets available from the State Library.
- Define policies and procedures for the preservation and availability of digital assets while maintaining intellectual property ownership and rights.
- Support and manage storage media for archival copies (preservation master, security copies, and access copies) in accordance with environmental, quality control, security, and other standards. Sustain a storage architecture that provides strategically selected redundancy in the event of technical failure, natural disaster, or human error.

COLLECTION POLICY

The digital preservation program is responsible for identifying, securing, and providing the means to preserve and ensure ongoing access to digital state government assets. Digital assets are those electronic objects that have been identified as having enduring cultural, historical, informational, and/or evidentiary value to the State of New Jersey.

Examples of digital assets include:

- directories
- reports
- statistical compendiums
- bibliographies
- newsletters
- bulletins
- state plans
- brochures
- periodicals
- committee minutes
- transcripts of public hearings
- maps
- and other documents produced for the public.

It does not include internal memos, administrative or training materials used only within an agency. Formats may be word processing documents, spreadsheets, datasets or digital images.

To become part of the digital preservation program, a digital asset will be evaluated in accordance with the New Jersey State Library collection development guidelines.

Digital assets selected for permanent retention will have ongoing usefulness or significance, based on the administrative, legal, fiscal, evidential, or historical information they contain that justifies their continued preservation.

The following types of records will be included under this program:

- Items that contain scanned or digitized material for which no analog counterpart exists.
- Born-digital assets produced by state government agencies.
- Library materials that are digitized and identified for long-term digital retention.

With the ever-growing volume of digital information, the State Library may need to allocate priorities for preservation action based on the relative significance of digital assets and the technical complexity of preserving and ensuring access to those assets.

OPPORTUNITIES AND CHALLENGES

The State Library has a legal mandate to preserve the publications of New Jersey's state government. This mandate has profound implications for the efficient management of democratic government and the cultural heritage of the people of New Jersey.

Though the costs of preserving digital assets may be high, the costs and implications of failing to preserve them are even higher. Loss of a digital asset not only means loss of the state's historical record, but also loss of the original investment of staff time and resources required to create the asset. This is especially true for born-digital materials, which can be lost either at the point where the asset becomes technically impossible to access or where access is so cost prohibitive (time, money, software) that recovery of the asset is not worthwhile.

There are two main challenges, both critical for the success of the preservation program. One challenge is the technical complexity inherent in the design, development, and operation of a digital preservation system. There is a perpetual race against the technology clock as hardware, software, and media formats evolve and become obsolete. Hardware, software, and network failures threaten to corrupt digital assets. Natural disasters can destroy data centers. Human failure is an ever-present challenge as digital assets are often unintentionally modified or deleted. The second challenge, less complex but no less important is one of communication and outreach. Establishing value and relevance among our potential users and stakeholders in state government.

There are also operational challenges such as adequately describing digital assets so that they are accessible, ensuring their technical characteristics are captured, and their archival provenance is maintained. Users must be able to easily access electronic publications.



From the State Library's digital postcard collection.

PRINCIPLES

The State Library's digital preservation objectives are best achieved through the development of practices that comply with an adequate, coherent, and widely understood framework for reliable, accountable, and manageable digital archives.

The digital preservation program is based on the following principles:

- Development of an interoperable digital repository framework using the best technology available— commercially or in house—incorporating open source options whenever feasible.
- Implementation of a range of community-based standards, best practices, and national and/or international standards that inform preservation procedures and technology as well as archival requirements such as provenance, chain of custody, intellectual property rights, and authenticity.
- To the best our capability, compliance with ISO 14721 the Open Archival Information System reference model (<http://www.oais.info/>) and ISO 16363 Trusted Digital Repository Standard for preservation. The later is more aspirational because of the lack of clarity surrounding the intellectual property rights of state government documents.
- Development of consistent documented guidelines and procedures for each stage of the lifecycle (i.e., creation, selection, acquisition, ingest, preservation action— including reformatting and producing derivatives—storage, identification/cataloging, access and use, and transform).
- Ongoing communication with other libraries and archives about opportunities for improved processes, monitoring the landscape for services offered by third party vendors.
- Review of policies and procedures on a regular basis, taking into account changes in the organizational, legal, and technical environment of the State Library, and government statewide.

COLLABORATION

The digital preservation program is committed to supporting collaboration between the Library, its departmental branches and other departments and agencies within state government to advance the development of the digital preservation of New Jersey state publications, seeking training and extending the breadth of available expertise, and expanding the digital content that is available to the public through cooperative efforts.

The New Jersey State Library will:

- Help identify and develop policies, procedures, and tools to support the management and preservation of digital information.
- Work with creators, publishers, and re-users of digital content to encourage practices that will enable, rather than hinder, preservation.
- Work with government agencies and officials to develop legislative and funding frameworks that will enable cost-effective preservation.

TECHNICAL ASPECTS

Authenticity

The State Library strives to ensure the authenticity of digital resources; the easily changeable nature of digital assets opens the possibility for unauthorized and undetectable changes. Confidence in the authenticity of digital records over time is particularly crucial owing to the ease with which alterations can be made. From the moment that digital resources are acquired, the State Library undertakes protective procedures to prevent, discover, and correct loss or corruption of digital assets due to either inadvertent or malicious intent. To facilitate this process the State Library uses a checksum checker called MD5. DSpace generates an MD5 checksum for file it stores and this checksum is used to verify the integrity of files over time.

In addition, the Library staff will endeavor to secure supporting evidence, ideally in the form of metadata, from those creating the resources. This will enable those who access the resources to evaluate the authenticity of all preserved digital resources.

Metadata

Metadata (descriptive, preservation and administrative) is fundamental to preserving and providing access to the state of New Jersey's digital assets. The preservation process includes maintenance of metadata submitted with the digital asset as well as the creation of additional metadata to manage the long term/active preservation and access of that asset. The Library staff are committed to identifying and collecting the needed metadata to preserve digital assets

in its collection and to provide access to those assets whilst remaining alert to the archival principle of LPMP (Less Process, More Product).

Formats

In acquiring digital materials, the State Library will define and communicate levels of preservation appropriate to each type of format. The State Library reserves the right to assign different preservation standards to different file formats.

Access/Use

The purpose of preserving digital assets is to ensure that the assets remain accessible in the future. The State Library must provide access using the most up-to-date technology available. For data that end users may download to their own computers, users will need to be informed of what software is necessary to render the digital record.

Where possible, the State Library will strive to make digital records available in openly documented formats for which rendering tools are readily available (e.g. PDF files, TIFF files, WAVE files).

In some cases, the digital record may be so tightly coupled with its originating software program that the record can only be rendered through the application used to create the digital record. In order to ensure continued access, the State Library may migrate assets to a new format. Some information may be lost during the migration but the fundamental information in the records will be accessible. However, the State Library as part of their preservation program, will maintain the original bit stream and make that available should it be needed. Sensitive and confidential information will require appropriate restrictions for access and use. Content intended for public dissemination will be available for public use.

Types of metadata within DSpace

DSpace contains four types of metadata about the archived content within the digital library – descriptive, administrative, structural and preservation.

Descriptive: Dublin Core is the default descriptive metadata schema used within DSpace though it is possible to use other schemas and also to customize the Dublin Core fields, an example is the addition of dc.date.removed for the Superseded New Jersey Administrative Code collection. Adopting an established metadata schema such as Dublin Core and restricting any customization to a minimum ensures that State Library metadata will be interoperable with other repositories and systems such as DPLA.

Administrative: examples of administrative data are information about rights and reproduction, legal requirements, version control, access restrictions, and statistical and audit trails.

Structural: provides information about the relationships between different parts of an object. It binds together components of complex items. The State Library uses structural metadata primarily within the public hearing collection to indicate the hearing and appendix sections.

Preservation: provides information about the physical specification of an object's creation, it's format and condition, hardware and software requirements to render it, it's transformation into other formats (change history or "provenance") and its authenticity (fixity). The purpose of preservation metadata is to help future generations interpret and recreate the information objects.

SUSTAINABLE FILE FORMATS

DSpace can recognize and manage a large number of file formats. The most common formats currently managed within the New Jersey State Digital Publications Library are PDF/A, PDF, JPEG, MPEG, TIFF, WAVE, and MP4 files. Although out-of-the-box DSpace only auto-recognizes common file formats, files of any format can be managed by DSpace as DSpace provides a simple file format registry where you can register any unrecognized format, so that it can be identified in the future. However, because a main concern is the long-term preservation of submissions, not just recognition, limits are placed on the types of files ingest.

- **Supported:** New Jersey State Digital Publications Library fully supports the format.
- **Known:** New Jersey State Digital Publications Library can recognize the format but cannot guarantee full support.
- **Unsupported:** New Jersey State Digital Publications Library cannot recognize a format; such formats are listed as "application/octet-stream", or Unknown.

"Support", means "make usable in the future, using whatever combination of techniques (such as migration, emulation, etc.) is appropriate given the context of need." For supported formats, one might choose to bulk-transform files from a current format version to a future version, for instance when the standard is an open standard such as PDF/A. While no file format can guarantee perpetual accessibility, certain formats have distinct advantages over others in this regard. These formats are often referred to as "sustainable" formats. Sustainable formats often include the below features:

- 1) **Published Documentation and Open Disclosure:** Specifications for the format are published and accessible to the public. This means that anyone who wants to create tools to work with the format can do so with no restrictions of copyright. Formats that share these characteristics are commonly called "open-source" or "non-proprietary." Because anyone can create tools to access such formats, they have a low chance of becoming inaccessible in the future, even if the formats themselves become obsolete.

- 2) **Widespread Adoption and Use:** The more widely a format is used, the more likely it is to have multiple tools used to access and manipulate it. This reduces the chance of a format becoming inaccessible due to one software publisher going out of business. Widespread adoption also serves as an indicator of general format stability and as a safeguard against loss of accessibility. A wider user base means more stakeholders who have a vested interest in keeping a format going.
- 3) **Self-describing Formats:** These formats contain metadata (data about the data) within their structure that interprets the content, context and structure of the file. This means that descriptive information (e.g., the file name, date of creation, identification of data within the file,) can be kept within the file itself, and external documentation is not required. When discussing long-term preservation this is particularly important, since records often become disassociated from their original software environment and accompanying files. The more self-contained a format is, the better the chances of the data contained within being accessible in the future.
- 4) **Unencrypted Files:** Electronic records with long-term retention should not be encrypted in any way, as this can severely compromise the future accessibility of those records. Encryption methods change dramatically over time, and the specific software tools needed to access current encrypted records may not exist in the future. A good electronic records management system can handle security, restricting access to records as needed, while leaving the records themselves unchanged.

Text

Preferred:

PDF/A (Portable Document Format / Archives): A variant of PDF that is specifically aimed at long-term preservation, its specifications are published in the standard ISO 19005-1:2005. It sacrifices certain functions, such as the ability to have external hyperlinks or embed audio or video, for the sake of greater reliability. PDF/A embeds all necessary fonts within the file and so it is completely self-extracting without the need to access any external font information unlike straight PDF. PDF/A also embeds descriptive metadata within the file itself, making it self-describing. These two factors make PDF/A the preferred format for long-term preservation of textual electronic records, both born-digital and digitized. Files can be converted to PDF/A by a number of different software tools and plug-ins. Word has an option to save as PDF/A.

Other options:

XML (Extensible Markup Language): A standard format for structured documents and data on websites, XML is also a preferred format for the preservation of metadata associated with records. XML is maintained and developed by the World Wide Web Consortium (W3C) but is open source. XML enjoys nearly universal adoption and can be accessed and worked on by scores of freely available software tools. XML is self-describing, but requires association with an appropriate schema (also freely available) in order to properly render all formatting.

HTML (Hypertext Markup Language): A standard format for structured documents and data on websites currently maintained and developed by the World Wide Web Consortium (W3C). HTML is open-source and universally adopted. Unlike XML, HTML does not contain descriptive metadata headings. This limits the machine-readability of HTML, particularly when attempting to perform advanced search functions within files.

Plain Text: The most basic form of text file, Plain Text can be rendered by any software that can read text, across any platform. Plain Text renders only basic characters, spaces and punctuation; however, it does not preserve formatting such as italics or bold letters. It is therefore typically used only for relatively small amounts of information such as software instructions or short notes. Plain Text is open-source and universally adopted. Common file extensions for Plain Text include .txt and .text

ODF (OpenDocument Format): An XML-based file format used for spreadsheets, charts, presentations and word processing documents. ODF was developed by Sun Microsystems, but is an open format, is freely available to anyone and has been published as an ISO standard (ISO/IEC 26300:2006). Owing to its relatively recent creation (2005) ODF is not as widely adopted as some other formats, but it is supported by almost all current office suites and word processing programs. File extensions for ODF files vary depending upon the specific type of file, but include .odt (word processing), .ods (spreadsheets) and .odp (presentations).

Still Images

Preferred

TIFF (Tagged Image File Format): TIFF was initially created in the 1980s in an effort to standardize file formats created by commercial scanners. The format has gone through a number of revisions since then, becoming an international standard for electronic images. The format is currently owned by Adobe Corporation, but the specifications are open and freely available. Unlike many image file formats, TIFF is uncompressed. This means that the

files are larger than a compressed format (such as JPEG) but there is no loss of data. This ensures that the file can be reproduced over time at its full fidelity. TIFF files can contain “tags” that store descriptive metadata about the file. TIFF files may have a file extension of .tif (Windows).

Other options:

JPEG 2000 (Joint Photographic Experts Group): JPEG-2000 was created by the Joint Photographic Experts Group in 2000 as a next-generation format for electronic images. The format is part of an international standard: ISO/IEC 15444:2004. JPEG-2000 files can be compressed in either lossy or lossless fashion, although only the lossless variety is acceptable for long-term preservation. The format is still relatively new, and thus does not have the same wide-spread use as TIFF. This makes it a slightly riskier choice for preservation, although usage of the format is growing. The lossless compression of JPEG 2000 provides some space savings over TIFF, but it may be better suited as a format for access rather than preservation. The standard file extension for JPEG 2000 is .jp2.

PNG (Portable Network Graphics): A file format initially created with the approval of the World Wide Web Consortium (W3C) as a replacement to GIF (Graphics Interchange Format). PNG is most often used to present images on the web and can be accessed with a wide variety of web browser and image display software. PNG uses a “lossless” compression algorithm which reduces the size of the file without losing any data. This means that images in PNG format do not suffer from “generation loss,” where the quality of an image suffers over time with repeated use. Specifications for PNG are open and freely available, and the format can contain extensive metadata within its structure.

Spreadsheets

Due to the complexity of spreadsheet structure it is challenging to perfectly represent data over time. Different software uses varied means to record formulae and link data, and so advanced functions are not always replicable in more open formats. The below formats represent the best approach for long-term accessibility, but both may be unable to represent certain formatting or functions of spreadsheets originally created in formats such as Microsoft’s XLS. The New Jersey State Publications Digital Library will save copies of spreadsheets with long-term retention in both the native (usually xls) format and in one of the below. This redundant method can preserve the maximum functionality of the spreadsheet while still protecting the core data from format obsolescence.

CSV (Comma Separated Values): A simple format which can be used to represent spreadsheet data. CSV files can be accessed with any spreadsheet software or text editor, but at the cost of potential loss of advanced functionality enjoyed by more proprietary

spreadsheet formats. There is therefore a tradeoff with using CSV: universal interoperability is excellent for long-term preservation, but the loss of advanced formulae may compromise the core data of the record. Basic spreadsheets containing tabular data without advanced functions may be better served by CSV than others.

ODF: (See previous entry for general data on ODF) The spreadsheet format of ODF, is a good choice for preservation of spreadsheets, as it supports more advanced functionality than CSV. However, spreadsheets originally created in other formats such as XLS may suffer some functionality loss upon conversion to ODF due to the non-standardized methods by which different software execute formulae.

Video

Uncompressed video can take up huge amounts of space in a storage environment, and thus formats utilizing “lossless” or “near-lossless” compression have become more acceptable in some cases. Compression of these types utilizes algorithms to reduce the size of a file without irrevocably losing any data. This can be compared to “lossy” compression, which sacrifices some data to achieve smaller size. Lossy compression is unacceptable for long-term preservation because it permanently alters the structure of digital content and can lead to gradual reduction in quality over time.

Preferred:

MPEG-4 (Motion Picture Experts Group): MPEG-4 is an open-standard format developed by the Motion Picture Experts Group as a format for encoding video content for dissemination on the web. There are two main encoding versions, and numerous subcategories, of the format. Documentation for all varieties of MPEG-4 is extensively published as part of an international standard: ISO/IEC 14496-14:2003. The compression of a given MPEG-4 video file will depend upon the specific software and coding used in its creation and can range from lossy to lossless. For long-term preservation only lossless or near-lossless compression should be used. MPEG4 supports the embedding of descriptive metadata to help support future access. A number of software tools, both free and paid for, are available to convert existing video files to MPEG-4 format.

Other option:

Motion JPEG 2000 (Joint Photographic Experts Group): Motion JPEG-2000 is a derivative of JPEG 2000 which codes and displays video. The format is part of an open international standard: ISO/IEC 15444-3:2004. Motion JPEG-2000 files can be compressed in either lossy

or lossless fashion, although only the lossless variety is acceptable for long-term preservation. The format is still relatively new, so adoption is not yet as widespread as older video formats. A number of software tools are available that can convert other video formats into Motion JPEG2000, and it can support a variety of descriptive and structural metadata.

Audio

Preferred:

BWF (Broadcast WAVE Format): A variant of the WAVE format, BWF (sometimes called BWAVE) was developed by the European Broadcasting Union with long-term preservation in mind. BWF takes the existing WAVE file structure and adds additional metadata support. The specifications for BWF are open and freely available, and the format is a de facto standard for digital audio for those in the radio, motion picture and television industries. It is also used extensively by audio archives throughout the world. The format is self-describing, as it contains its own structural and descriptive metadata. BWF files are uncompressed and can be played by any software that is WAVE compatible. In order to display, add or modify metadata in a BWF file, however, one must use software that specifically supports the format. Free software is available that can attach BWF metadata to existing WAVE files. The file extension for BWF is .wav, the same as standard WAVE files.

WAVE (Waveform Audio File Format): WAVE is a format created by Microsoft and IBM in the early 1990s. Though proprietary, the format is fully documented and has been used as the basis for the preservation oriented variant BWF (see entry below). WAVE files are uncompressed, so they lose no audio data as with some other audio formats. The format also enjoys near-universal adoption, as it is compatible with virtually every audio player available, across computer platforms. Software utilities to convert other formats to WAVE are plentiful and inexpensive (or free). WAVE has limited metadata capabilities, so is a second choice for long term preservation behind BWF (see below). WAVE can still be an acceptable format for nonpermanent audio, provided that appropriate external metadata is associated with the WAVE files.

Preferred File Formats for Submitting Items to the New Jersey Publications Digital Library

TYPE OF PUBLICATION	BEST CHOICE	OTHER OPTIONS	DSPACE COMPATIBLE
Documents such as reports, hearings & newsletters.	PDF/A-1 (ISO 19005-1:2005 compliant PDF/A) OpenDocument Text (.odt)	PDF/A-1b (.pdf) (ISO 19005-1 minimally compliant PDF/A) Microsoft Word Document (.doc) Microsoft Open XML Document (.docx) Rich Text Format (.rtf)	Yes - recognizes best and other formats
Spreadsheets	OpenDocument Spreadsheet (.ods) Comma-separated file (.csv) Tab-delimited file (.txt) PDF/A-1a (.pdf) (ISO 19005-1 compliant PDF/A)	Microsoft® Excel® Spreadsheet (.xls) Microsoft® Excel® Open XML Spreadsheet (.xlsx) Other delimited text files (space-delimited, colon delimited, etc.)	Yes - recognizes best and other formats
Audio	Broadcast WAVE Format LPCM (.wav) WAVE Format LPCM (.wav)	AIFF (uncompressed) (.aif, .aiff) Standard MIDI (.mid, .midi) Windows® Media Audio WMA (.wma) MPEG3 (.mp3) MP4 AAC (.m4a)	Yes - recognizes best and other formats
Video	MPeg4 (: ISO/IEC 14496-14:2003)	Motion JPEG 2000 (ISO/IEC 15444-3:2004)	Yes - recognizes best and other formats
Presentations	OpenDocument Presentation (.odp) PDF/A-1a (.pdf) (ISO 19005-1 compliant PDF/A) for presentations without animation	Microsoft® PowerPoint® Presentation (.ppt) Microsoft® Open XML PowerPoint Presentation (.pptx)	Yes - recognizes best and other formats
Images	TIFF (.tif, .tiff) uncompressed JPG 2000 (.jp2) PDF/A	JPEG (.jpg, .jpeg) PNG (.png) PDF/A-1a (.pdf) (ISO 19005-1 compliant PDF/A) GIF (.gif)	Yes - recognizes best and other formats

BEST PRACTICES FOR FILE NAMING

In the same way that physical collections are organized and numbered so that items can be found on the shelf so are digital files organized. In fact, the way that physical items are named in the state publications section is the foundation for digital file names.

Good file names are essential to current and future accessibility. Efficient management of a digital library begins with accurate file-naming. Our file naming convention should incorporate best practices and this section explores the file-naming rules that should be applied to all digital publications. Digital publications include documents, spreadsheets, databases, images, video, and audio.

- 1. Each file should have a unique combination of numbers and letters usually based on the call number**

Just like call numbers each digital file should have a unique file name. Do not rely on the file folder to provide context.

- 2. Avoid using special characters (:&*\/#\$. ,)**

The characters listed above are frequently used for specific tasks in an electronic environment. For example, a forward slash is used to identify folder levels in Microsoft products, while Mac operating systems use the colon. Periods are used in front of file-name extensions to denote file formats such as .jpg and .doc; using them in a file name could result in lost files or errors.

- 3. Use underscores rather than periods or spaces**

As mentioned above, periods already have a specific function in a file name, which is to tell the computer program where the file-name extension begins. Spaces are frequently translated in a Web environment to be read as "%20".

- 4. Err on the side of brevity**

Generally, about 25 characters is the maximum length to capture enough descriptive information for naming a publication but fewer characters are better.

- 5. Files will be moved from their original location**

Files are frequently copied to other folders, downloaded, and emailed. It is important to ensure that the file name, independent of the folder where the original file lives, is sufficiently descriptive. The State Library's digital files are organized in a series of folders in njdigidox based on call number. However, this way of storing records is only efficient as long as the files stay in their original folders in their original context. As soon as "1919.pdf" is copied to another folder, or emailed to a patron, the context provided by the folders in which the document is nested is lost. Context can be important because it provides authenticity and trustworthiness. For instance, if the following files were pulled out of their appropriate folders, they would appear to be the same file: njdigidox\901\40\1919.pdf and njdigidox\901\31\1919.pdf

- 6. Use descriptive file names**

The file name should include enough descriptive information independent of where it is stored. For the digital library this includes a call number and date.

Preferred File Naming for the New Jersey Publications Digital Library

Publication	Call #	Correct	Incorrect
<i>Annual Report of the NJ Dept. Education 1921</i>	974.901 E31 1921	e311921	1921
<i>Annual Report of the NJ Dept. Agriculture 1921</i>	974.901 A40 1921	a401921	1921
<i>Public welfare in New Jersey, 1948-1952</i>	974.90 I52 1953	i521953	1953
<i>A Sketch of the NJ Dept. Motor Vehicles</i>	974.90 R424 1953	r4241953	1953
<i>A Chronology of the NJ Motor Vehicles Dept.</i>	974.90 R424 1953a	r4241953a	1953a
<i>Study of NJ School Educational Needs</i>	974.90 E24 1953a/z	e241953a_z	1953a/z
<i>NJ Economy at a Glance: December 2019</i>	974.905 E19 Dec 2019	e19_201912	Dec_2019
<i>News Release Oct 17, 2019</i>	974.905 L19 2019 Oct 17	L19_pr_20191017	20191017
<i>Medical History of Atlantic Co., NJ</i>	LH1414	LH1414	Atlantic
<i>Old Sussex Almanack, Sept. 2019</i>	J/Per S98 2017 Summer	sussex Almanack- summer2017.pdf	Summer2017
<i>NJ Legislative Manual, 1878</i>	J328 M294 1878	J328m2941878	1878

Special collections or projects (such as the superseded New Jersey Administrative Code) may warrant their own file naming conventions. See a digital library administrator when in doubt.

COPYRIGHT AND NEW JERSEY STATE GOVERNMENT DOCUMENTS

Answer yes or no to the following questions in order to navigate through the decision matrix

Is the material a publication or document published by the state of New Jersey or one of its agencies?

- If the answer is **YES**, it is highly likely no permission is needed

The State of New Jersey on <https://www.nj.gov/nj/legal.html> states:

The State of New Jersey has made the content of these pages available to the public and anyone may view, copy or distribute State information found here without obligation to the State, unless otherwise stated on particular material or information to which a restriction on free use may apply. However, the State makes no warranty that materials contained herein are free of Copyright or Trademark claims or other restrictions or limitations on free use or display. Making a copy of such material may be subject to the copyright of trademark laws. See [State of New Jersey Conditions of Use Notice](#).

Look for a copyright notice on the publication. If you see one, do not digitize or capture.

- If the answer is **NO...**

Is the item a work of New Jersey (or any government entity) laws, codes, judicial decisions etc. without commentary?

- If the answer is **YES**, no permission is needed

The law is always in the public domain, whether it is government ordinances, statutes, judicial decisions or regulations. What is not in the public domain are those legal materials that have commentary or context offered along with the legal text.

- If the answer is **NO...**

Is the material a work of the U.S. government?

- If the answer is **YES**, no permission is needed
- If the answer is **NO...**

Was it published prior to Dec. 31, 1924?

- If the answer is **YES**, no permission is needed
- If the answer is **NO...**

Was the item published prior to 1978 and does not have a copyright notice, either the symbol or the text?

- If the answer is **YES**, no permission is needed

Under the copyright laws that were in effect prior to 1978, a work that was published without copyright notice (such as Copyright or a ©), fell into the public domain

- If the answer is **NO**...

Was it published between 1924 and 1964 and copyright was not renewed?

- If the answer is **YES**, no permission is needed

In the US, books published before 1964 had to get their copyrights renewed at the Library of Congress Copyright Office in their 28th year, or they would fall into the public domain. Check here to see if this was done: <https://onlinebooks.library.upenn.edu/cce/>

- If the answer is **NO**...

Is the item something you are digitizing for preservation purposes?

- If the answer is **YES**, no permission is needed

The Digital Millennium Copyright Act has a provision that explicitly allows libraries and archives to make up to three copies of a work for preservation purposes. The item being preserved can be in any format (text, image, video, sound) but it cannot be made available outside the library.

- If the answer is **NO**...

Either seek permission or forgo digitization or capture.

There is more nuance to many of these questions than is presented here. These sites are helpful for further research:

[ALA Copyright slider](#)

[Preservation copying resources from ALA](#)

[Harvard State Copyright Resource Center](#)

FURTHER RESOURCES

[Digital Projects 101: A Resource Guide](#)

[DSpace Guides and FAQ](#)

[Harvard State Copyright Resource Center](#)

[Library of Congress Digital Preservation](#)

[NARA Digital Preservation](#)

RESOURCES USED

Illinois Digital Environment for Access to Learning and Scholarship: [IDEALS Digital Preservation Policy](#)

North Carolina Dept. of Cultural Resources: [File Format Guidelines For Management And Long-Term Retention Of Electronic Records](#)

National Archives, UK: [Selecting Storage Media for Long-Term Preservation](#)

Purdue University: [Research Repository \(PURR\) Digital Preservation Policy](#)