Digitizing Records

Contents

Digitizing Records

Overview

Definition

Digitizing is the process of converting physical records into digital format. It is also known as digital imaging, scanning, or digitization. This can apply to the conversion of text-based documents, photographs, maps, microfilm, analogue sound, video, or film to digital formats.

*Note that digitization is a different process than digital transformation. Digital transformation includes changes in process and systems to streamline all records creation and capture into born-digital format. Digitization of analogue records can be a part of a larger digital transformation strategy, but is not, in itself, digital transformation.

Value Assessment of Digitization

Digitizing is most beneficial when:

- frequent access to high volumes of records is required
- there are multiple users accessing the same records
- the original format makes it difficult to access the records (e.g. large maps or microfilm)
- the original organizational structure of the files makes it difficult to search the records

There are many benefits that can be gained from digitization. There are also costs and risks. Business areas should look at the benefits and risks of conversion, storage, management, and maintenance of the records, and ensure information protection, security, and authorized disposal of the records once in digital form.

Summary of Benefits and Risks

While digital imaging is a popular option for access, as with any investment in resources, program areas should be aware of the potential benefits and risks. A cost-benefit analysis may include the following:

Potential Benefits

- Easier access to records

- Increased storage capacity / reduced office footprint

- Improved tracking and control of active files and content when supported by a system with adequate records management system functionality

- Improved access when combined with indexing

- Increased productivity

- Access can be managed

- Facilitates work process and information flow

- Easier to share information

- Reduction of hybrid systems (multiple formats to access full file)

- Extension of life of audio-visual and other records at risk of deterioration and loss

Potential Risks

- Increase to storage costs

- Hardware and software obsolescence (requires migration over time)

- Digitization itself is not a records management solution - issues could be compounded if good practices are not in place

- Requirements not thought through upfront, compromising the full benefits of the digital records

- Unmanaged accumulation of digital records, potentially resulting in loss of integrity, accessiblity, and reliability

- Paper originals still needed to be kept to meet requirements

- Hybrid systems increase, potentially reducing relability and integrity of the records

High-Level Considerations

Departments should carefully consider what records need to be in digital form for active business use. Upfront planning should be done to determine what conversion is needed, why, where records will be stored once digitized, how they will be accessed, controls required to protect the records, and costs over time.

Format

For many activities, the scanned version of records can serve as the recordkeeping copy and the originals can be destroyed. However, this is not always the case and proper authorization is required. For example:

- some records must be in hard copy to meet legal requirements (e.g. birth, marriage, and death registrations)
- some records have been identified to be transferred to the Archives for permanent preservation in their original format
- some records are considered high-risk/high-value and will need to have a robust recordkeeping system to ensure the digitized records have authenticity, integrity, and usability over time. Without such systems in place, the original may be required to minimize risk and meet requirements.

Analyzing the program area's business requirements is the best way to ensure that the digitized records and the original physical records will be managed appropriately.

Prior to scanning, it is recommended that the Government Records Office (GRO) is consulted to ensure that recordkeeping requirements are considered and the ongoing management of both the analogue and digital records are provided for in up-to-date records schedules. Other stakeholders, such as Legal Services, may need to be consulted during the analysis phase. See the Business Needs Analysis to start the consultation with the GRO.

Digital Systems and Storage

Prior to digitizing the records, program areas should identify the recordkeeping system that will be managing the records and where the records will be stored. At a minimum, digital records storage should:

- support organized business file structure(s) so that related records can be grouped together
- have appropriate security and access controls
- have acceptable retrieval times

Digitized images can require significant storage space depending on the quality, resolution, and compression ratios of the images. See Recordkeeping Systems for more information. The best approach is to go beyond storage and manage digitized images in an Electronic Document and Records Management System (EDRMS) - enabling records management rules to be applied, including the execution of authorized retention and disposal provisions.

Costs

Program areas should perform a cost-benefit analysis to determine the costs of digitization and the return on investment (ROI) over time. Program areas should factor in a number of facets, including long-term storage and migration if the records have a long retention period. The ROI should warrant the resources required to manage the records over time.

The following list is not exhaustive but provides introductory information for program areas to begin working through a cost analysis:

- initial system costs (hardware and software)
- staff time for setup, training, upkeep, and troubleshooting
- the digitization process (document preparation, digital quality control, image capture, metadata capture, and management)
- indexing requirements
- integration with other systems
- data storage and migration costs, including costs for records that need to be maintained medium- to long-term (over 10 years)

If records are to be destroyed at the end of their retention period and have short- to medium-term retention requirements (less than 15 years), the cost of conversion and investment of time may result in a negative ROI. If other benefits, such as accessibility for all staff, changes in workflow process, etc., are the impetus for the conversion, analysis should be done to see if the resources would be better applied to digital transformation of the process. Keep in mind that digitizing paper alone will not transform practices; recordkeeping systems managing the records also need to change. Unmanaged accumulation of digital records over time will result in unnecessary costs for government.

Mitigating Risk Over Time

If the digital records will be required for a medium- to long-term retention period (more than 15 years), then the life expectancy of the supporting systems and formats will need to be considered as it is likely

that the format of the records will require conversion and the systems replaced. Choosing industry standard formats and systems that have robust capabilities now will help to prevent issues in the future.

As well, active management and migration of the records over time will help to keep the records reliable and accessible.

Work processes should support current business needs, but the integrity and accessibility of the information in the records over time should also be considered. Analyze the value of the digital asset against the retention period and the disposition. For example, maintaining digital images for 10 years versus 100 years requires different levels of commitment of resources and recordkeeping capability to ensure the records will be there when needed, for as long as required. See our guidance on high-risk/high-value records and information for some considerations related to risk and value.

Mitigating risk can be done in a number of ways. See Recordkeeping Framework for Departments and Agencies: Policies and Requirements and Managing Records to Mitigate Risk as starting points. If not considered upfront, program areas may still need to print to file or apply additional manual processes and procedures to adequately meet requirements.

Recordkeeping Requirements

Under *The Archives and Recordkeeping Act,* government bodies are responsible for creating full and accurate records of their activities and for ensuring that they are managed according to standards and processes established by the Archives of Manitoba.

A department/agency planning a digitization project must ensure that:

- records schedules are updated to:
  - document the change in business process
  - specify the retention requirements of the electronic records
  - enable authorized disposal of the original source records
- the system(s) supporting the records ensure the records have authenticity, integrity, and reliability for as long as required

In cases where the records have been identified for permanent preservation, the Archives of Manitoba will assess which format is best suited for archival designation.

Detailed Considerations

File Naming / Indexing

File names should be:

- unique
- meaningful, i.e. descriptive
- consistently structured and straightforward

Indexing should:

- provide the most minimal structure required to effectively organize the records
- be meaningful, i.e. descriptive; avoid using jargon and generic terms
- be consistently structured
- be organized in a way that enables records management rules to be applied, i.e. retention

Other tips:

- use lowercase letters and dashes; avoid special characters, slashes, and spaces
- keep length under 30 characters
- if numbered, take into account the maximum number of items to be digitized and reflect that in the number of digits, and use leading zeros to facilitate sorting (e.g. if digitizing 10,000 objects do not have the numbering be 1, 2, 3 etc. but rather 00001, 00002, etc.)

For more information about file naming see the File Structures guidance.

Technical Options

There are many technical options associated with digitization related to file format, file size, colour resolution, and compression (see glossary for definition of terms). Technical specifications are always evolving but the primary consideration is to ensure the legibility and usability of the digital image. The technical specifications adopted for a particular project should ensure that the images are appropriate for the purpose for which they are being digitized; that the essential characteristics of the source records are captured; and that the digital records can be retained for as long as required.

Project-specific standards should be clearly defined and documented before a digitization project is started to ensure that quality can be evaluated. The standards may be different depending on the goal of the digitization project; for example, the technical specifications might be less stringent when digitizing for access purposes only. When making access copies, you may choose to meet lower standards to create smaller files and save storage space and costs, knowing that the non-digital source documents will be retained. Or conversely, rigorous specifications may be warranted to ensure the best quality for the record type, based on the intentions of use, retention period, and disposition.

File Formats

Departments/agencies should create the best quality image possible with available resources, the purpose of the images, and recordkeeping requirements. For master digital images, the recommended file formats are TIFF or PDF/A.

Digitized images in a lower standard file format, such as PDF (for text) or JPEG (for photographs), can be sufficient in some situations. Some examples include:

- when access copies are produced but the original source records will be retained
- when the records being digitized are scheduled for destruction after a short retention period
- when derivative images are made from master images for the purpose of easier use and file sharing

PDFs or PDF/As may be the most appropriate format if the file will be added to with additional PDF pages.

Resolution

Resolution is the amount of picture data in a specific area of an image. Resolution is measured in pixels per inch (ppi) or dots per inch (dpi). Images scanned at a higher resolution will produce a better image. Some source records require higher resolution; for example, photographs require a much greater resolution than text documents. Larger records, such as maps and plans, must be scanned at an even higher ppi/dpi to ensure details are clear. The higher the resolution, the larger the file size.

Note: File size may be one consideration but it should never be the primary consideration when selecting technical specifications. The primary consideration should be the creation of legible, accessible, and usable images that meet your business needs and preserves the source records' essential characteristics for as long as required. If this cannot be guaranteed with the resources available, then the digitization project may not be viable or provide a return on investment.

Colour Resolution

The colour resolution, or bit depth, allows a greater range of colours or shades of grey to be represented by a pixel.

Digital images can be bitonal, greyscale, or colour. The decision to create bitonal, greyscale, or colour images should consider the colour of the source record, and whether the digital image will replace the source record as the record of business. Capturing a record at lower than recommended bit depth will possibly result in a digital image that is visibly different from the source record, although that may be adequate in some cases; in others it may compromise the evidential value of the records.

Compression

Compression is a means of reducing the size of a digital image for storage or transmission. Consider whether not the evidential value of the file will be compromised if the removal of information during the compression process would affect the records.

Post-Digitization Records Management

Quality Control

It is essential to define, document, and implement quality control procedures. Quality control (QC) helps to ensure that the digital copy is a true and accurate copy of its paper-based original. This is critical to being able to assert that the digital copies have integrity and authenticity.

Quality control should be performed by both the vendor digitizing the records (if applicable) and the program area.

QC procedures should address:

- verification that digital output matches quality of non-digital source record input
- extent and frequency of digital image sampling
- criteria for checking image quality
- frequency of and criteria for checks on file naming and indexing
- acceptable variations from normal procedures
- processes for re-digitizing

Quality checks should include:

- confirmation that the number of physical items scanned is equal to the number of images produced
- de-skewing of pages and removal of blank pages
- validation that indexing is correct
- verification of the readability of output images.
  - For this important step, it is recommended that the "Acceptable Quality Limit" be explored when assessing quality control and verify the digitized product. There are a number of online calculators that will assist in calculating the samples required to meet ISO 2859 quality assurance standard.

Automatic de-speckle, significant cropping, and other enhancements should **not** be used unless previously agreed upon prior to digitization.

QC procedures must be carried out before digital images are accepted into a business process and before the destruction of non-digital source records can be considered, if authorized by an approved records schedule.

Paper Source Records

If the records are designated as having long-term value and will be transferred to the Archives of Manitoba after they have reached their disposition date, the paper source records must be kept, maintained in their original order, and transferred to the Archives. The preparation and digitization processes must not damage the records.

If the records are scheduled to be destroyed, the paper source records must not be disposed of without authorization, and any applicable records schedules will need to be updated to include a provision for managing the paper source records and the digital records.

Note: Paper source records are **not** transitory records and cannot be transferred under a non-filed office paper schedule.

Contact the Archives of Manitoba to ensure you have properly planned to meet your requirements. In order to facilitate discussion, complete the Business Needs Analysis.

Digital Records

Digital records need to be managed to ensure their continued existence. Technology-dependent digital records are inherently vulnerable to hardware, software, and media obsolescence. Digitized records may need migration, conversion, or other strategies to preserve them for as long as they are needed. In addition, program areas will need to assess the authenticity and integrity of the digital record against risks, including legal admissibility, operational needs, and audit scrutiny. See Policy 2 in the Recordkeeping Framework for Departments and Agencies: Policies and Requirements for more information about the ongoing management of the digitized records.

Government Records Office, Archives of Manitoba
T: 204-945-3971 | E: gro@gov.mb.ca
Visit our web site to learn more about Government Recordkeeping

---

Digitizing Records

Appendix A – Business Needs Analysis

The Business Needs Analysis has been provided as a separate document in order to facilitate completion and submission to the Government Records Office for review. You can find the document at this address:

https://www.gov.mb.ca/chc/archives/gro/recordkeeping/docs/digitizing_records_appendix_a.pdf.

Appendix B – References

Alberta, *Digitization Guideline.* https://imtpolicy.sp.alberta.ca/guidelines/Pages/Digitization-Guideline.aspx

Archives New Zealand, *Destruction of Source Information after Digitization.* https://archives.govt.nz/manage-information/resources-and-guides/disposal/destruction-of-source-information

British Columbia, *Digitizing Government Information Standard,* Last revised April 2020. https://www2.gov.bc.ca/assets/gov/government/services-for-government-and-broader-public-sector/information-technology-services/standards-files/digitizing-government-information-standard.pdf

CAN/CGSB-72.34-2017 National Standard of Canada. *Electronic Records as Documentary Evidence.* This standard is available at no cost from the Canadian General Standards Board: http://www.publications.gc.ca/pub?id=9.839939&sl=0

ISO 13008:2022 *Information and documentation - Digital records conversion and migration process.* https://www.iso.org/standard/52326.html

ISO/TR 13028:2010 *Information and Documentation – Information and documentation - Implementation guidelines for digitization of records.* https://www.iso.org/standard/52391.html

ISO 15489-1:2016 *Information and documentation – Records management – part 1: Concepts and principles.* https://www.iso.org/standard/62542.html

New Brunswick, *Digitization Standard* (2013). https://archives.gnb.ca/CIM/Standards/en-CA

New South Wales, *Digitisation* https://staterecords.nsw.gov.au/recordkeeping/advice/digitisation

Queensland Government, *Digitise Records.* https://www.forgov.qld.gov.au/digitise-records

See also: Fact Sheet on Recordkeeping Standards