File Formats for Long-Term Preservation and Transfer to the Provincial Archives

This document has been created to provide guidance on file formats that are suitable for the long term preservation of electronic records by institutions of the Government of Saskatchewan, as well as for the transfer of electronic records to the Provincial Archives.

Section 24(5)(b) of *The Archives and Public Records Management Act* requires that government institutions protect and maintain all public (government) records that are in their custody or under their control so that those records are:

(i)   usable and accessible;

(ii)   transferable;

(iii) legible and understandable;

(iv)   maintained in a format and in a condition that ensures their retention and preservation in accordance with the applicable approved records schedule; and

(v)   if applicable, capable of being transferred to the Provincial Archives of Saskatchewan in a format acceptable to the Provincial Archives of Saskatchewan.

In order for an institution's electronic records to meet these requirements, the institution needs to carefully consider the file formats in which these records are created and/or saved. This is particularly important when the records in question are assigned long-term or permanent retention periods in the institution's applicable retention schedule.

File formats, especially those that are proprietary, are often undocumented and may become obsolete over time. Without consideration to the file formats used, or to the need to migrate electronic records into new versions of software, government institutions risk losing valuable information as records become inaccessible.

The file formats recommended in this document are for the preservation of long term records and are either non-proprietary national or international standards or are *de facto* industry standard file formats. In addition to the use of standard file formats, government institutions should also ensure that official electronic records are retained on a backed up network drive (such as the Information Technology Department's G Drive), or in an Electronic Document and Records Management System (EDRMS) if one has been implemented by the institution. Official electronic records should not be retained on external storage media, such as CDs or USB flash drives, as these may be easily lost or damaged, or may degrade over time and thus result in the loss of government records.

The Provincial Archives advises that in situations in which the electronic record will be retained by the institution for longer than 10 years, the records be created in one of the file formats identified for long-term storage in this document (see table on pgs. 3-4). If records with long-term or permanent retention periods cannot be created in one of the recommended file formats, the Archives advises that in order to better facilitate long term preservation, the records be migrated into one of these formats shortly after their creation or after their active

business use has ceased. When determining whether or not a record will be retained by your institution for 10 years or longer, remember to consider the duration of time that any condition specified in the retention period of the records schedule is likely to last. For example, a policy document with a retention period of "after policy has been revised or is obsolete + 6 years" may be retained by the institution for longer than 10 years, depending on how long it takes for the condition to be met.

Once records have met their required retention periods, government institutions must submit a request for their disposal to the Provincial Archives. During the disposal process, the Archives determines which government records have historical value and will be transferred to its Permanent Collection; approximately 5% of all government records are selected by the Archives.

The table below identifies recommended file formats for records with long-term retention requirements. These file formats are mandatory requirements for electronic records transferred to the Archives by government institutions. If records selected for transfer by the Archives do not exist in one of these formats at the time of their selection, the government institution must migrate the records into a required format prior to their transfer, or contact the Records Processing and Preservation Services program of the Provincial Archives to determine alternate arrangements.

In addition, when transferring electronic records to the Provincial Archives along with the files themselves, the institution must include a csv (Comma Separated Value) file of metadata for each series of records to be transferred. This provides contextual information about the records that is invaluable to their understanding over time. The seven mandatory metadata fields required for digital records at folder and file level are listed below. The Archives also accepts any additional metadata that will add context to the records. Please note: Much of this information will have been captured by the disposal request form and so can be transcribed from that document to the metadata form.

1. Title
This is the folder or file name of the record(s).

2. Identifier
This is not a system-generated ID number but the file path represented as a URI (Uniform Resource Identifier); it supplies context for the record(s).

3. Creation Date
This is the last edit date when the record(s) was completed as an official record(s). This is not the date the record(s) was copied to its current location. For records generated in a Windows operating system the date last modified is often the most accurate date to capture.

4. Folder or file
Specify whether the digital object is a folder or a file.

5. Checksum

A checksum is a mathematical way of describing a set of data. If the checksum on one set of data matches the checksum on another set of data, the two are absolutely identical. Checksums can be generated by tools utilized by the Provincial Archives to facilitate the secure transfer of electronic records. If needed the Provincial Archives can provide checksum tools to facilitate this process. The institution will need to send checksums for the record(s) and for the csv metadata file itself. These checksums are used to ensure that the transfer is correct and complete.

6. Copyright

Identify any copyright conditions related to the record(s).

7. Privacy status

Identify if there are any privacy concerns relating to the record(s), or if the record(s) contain any confidential or sensitive information.

Table of Recommended File Formats for Long-Term Preservation and Transfer

Content Type	Recommended for Long-Term Storage – Required for Transfer to Archives	Acceptable for Short-Term Storage (less than 10 years following creation)
Text	• PDF for long-term preservation (PDF/A) • Open Document Format (ODF) • Extensible Hypertext Markup Language (XHTML) • Extensible Markup Language (XML) • Hypertext Markup Language (HTML) • Style Sheets (CSS) • Multipurpose Internet Mail Extensions (MIME) • Rich Text Format (RTF) • Standard General Markup Language(SGML) • Text (plain text)	• Microsoft Office including: Word Document Format (DOC/DOCX), Excel Spreadsheet Format (XLS/XLSX), PowerPoint Presentation Format (PPT/PPTX) • Portable Document Format (PDF)
Still Images	• Joint Photographic Experts Group (JPEG 2000) (JP2) (uncompressed) • Tagged Image File Format (TIFF) • TIFF – GeoTIFF • Portable Network Graphics (PNG) • Scalable Vector Graphic (SVG)	• Encapsulated Postscript (EPS) • Graphics Interchange Format (GIF) • Joint Photographic Experts Group (JPEG)

Audio	• Broadcast Wave Format (BWF) • Free Lossless Audio Codec (FLAC) version 1.1.2 • Waveform Audio Format (WAV)	• Audio Interchange File Format (AIFF) • Mpeg-1 layer-3, Mpeg-2 layer-3 (MP3) • Mpeg-4 aac – Advanced Audio Coding (AAC/M4A) • Windows Media Audio (WMA) • Mpeg-TS (TS)
Web Archiving	• Internet Archive Format (ARC) • Web Archive Format (WARC) • MIME HTML (.MHT) • PDF	
Digital Video	• Motion JPEG 2000 • Dirac • Raw bitstream • FFV1	• Mpeg-1, MPEG-2 (*.mpg, *mpeg, wrapped in AVI, MOV) • Mpeg-4 (H.264) (*.mp4, wrapped in AVI, MOV) • Windows Media Video (WMV) • Apple ProRes (MOV) • Mpeg-5 (H.265)
Databases	• Software Independent Archiving of Relational Databases (SIARD) • Delimited Flat file with DDL	• dBase Format (DBF) • Structured Query Language export (SQL)
Spreadsheets	• Comma-separated values (CSV) • OpenDocument Spreadsheet (ODS)	• Microsoft Excel
Computer Aided Design / Technical Drawing	• Drawing Interchange File Format/Data eXchange Format (DXF)	• Computer Graphics Metafile (CGM)

For further information or questions, please contact:

Provincial Archives of Saskatchewan
Information Management Services
Inquiry Line: (306) 787-0734
recordhelp@archives.gov.sk.ca