Canadian Research Knowledge Network

Réseau canadien de documentation pour la recherche

CRKN TDR Preservation Strategic Framework

Approved by the Preservation and Access Committee: July 26, 2021

About this Document

This document describes CRKN's approach to ensuring the long-term preservation of and access to digital content stored within the Trustworthy Digital Repository (TDR). It is intended to guide the development, revision, and implementation of policies and practices relevant to digital preservation.

This framework is based on the early work of Canadiana.org in developing the TDR and will evolve over time to adapt to the changing digital preservation landscape.

Preservation Strategic Framework

CRKN's strategy for the long-term preservation of digital content within the Canadiana Trustworthy Digital Repository ("the Repository") is based on five core approaches:

1. Process Integrity

Preservation of digital content requires that the Repository processes operate in a predictable and consistent manner. To ensure this, CRKN:

- Maintains documented practices, processes, procedures, specifications, and standards to ensure Repository operators (all staff involved in the administration or operation of the Repository) perform their duties in a consistent and predictable way, and enable review and auditing
- Performs regular internal and external review and updating of all procedures and their associated systems to ensure that processes are relevant, up to date, and being followed. This review process includes an internal audit at least once every two years, and external audits as determined appropriate by the Preservation and Access Committee.

2. Data Integrity

Preservation of digital content requires that the integrity of the underlying digital data be safeguarded and verified. To ensure this, CRKN:

- Maintains a distributed preservation system, which guarantees a copy of each digital object is stored in at least three geographically distinct locations, to ensure that damage at or the loss of any single node will not result in permanent data loss
- Performs continuous automatic fixity checking to detect missing or corrupt data combined with automatic replication and repair of lost or corrupt information to ensure that undetected cumulative failures do not result in loss or corruption
- Reviews and tests the structural metadata of digital objects to ensure that it can be identified and used out of the context of their AIP
- Preserves the structure of digital objects through metadata which can be unambiguously linked to the object
- Uses open file formats for digital objects and metadata which are well supported at the time of adoption, expected to remain supported for a long period of time, and which are expected to have a viable migration path when superseded by newer formats

- Validates and normalizes file formats and contents on ingest to ensure the Repository hosts only objects of known and supported types
- Performs regular environmental scans of the level of support for all preserved file and data formats to detect impending obsolescence and alerting depositors of at-risk formats
- Migrates end-of-life formats to newer ones, where feasible, with the goal of preserving as much of the original structure and content as possible, preserving the migrated content alongside the originals along with documentation of the process used to create the former from the latter

3. Infrastructure Integrity

Preservation of digital content requires that the systems on which the content resides are reliable, secure, and predictable in their operation. To ensure this, CRKN:

- Refreshes systems on an ongoing basis through a process of evergreening ensures infrastructure remains modern and reliable
- Standardizes software and hardware to a limited number of configurations keeps the overall complexity and variability of infrastructure low, thereby simplifying the ability to understand and maintain the infrastructure
- Standardizes and automates deployment and management of software and systems to ensure the Repository is configured and operates in a consistent and predictable way
- Prefers industry standard and open-source components to reduce the risk of dependence on unsupported, dead end, or poorly understood systems and tools
- Cross-trains staff involved in Repository operations to ensure redundancy in all key operations and areas of knowledge provides resilience against loss or turnover in staff and enables cross-checking of tasks
- Manages and monitors access to physical infrastructure to reduce risk of loss or damage, whether intentional or accidental

4. Object Provenance

Preservation of digital content and the provision of useful and trustworthy information to stakeholders requires that objects can be traced back to their sources. To ensure this, CRKN:

- Documents and tests processes for handling data and logging of operations performed on data to provide an audit trail for reviewing, testing, and replicating changes to data and creation of derivatives
- Associates each Archival Information Package (AIP) with a particular depositor and maintains depositor agreements to ensure that objects are linked with their owners
- Creates and maintains metadata describing the location of the original objects from which digitized versions are created to aid in locating original materials for comparison and verification
- Maintains documented, tested, and replicable processes for the creation of DIPs, and logs of DIP creation ensure that the authenticity of Dissemination Information Package (DIP)s can be established
- Records all transformations made to preservation objects combined with retention of copies of previous formats to enable validation and replication of migration activities

5. Information Access

Ongoing access to useful and trustworthy information is the ultimate goal of digital preservation. To ensure this, CRKN:

- Prefers widely used open standards and formats and structural validation of file formats upon ingest to reduce the likelihood of incompatibility with current and future tools and maximize the likelihood that formats can be migrated as and when needed
- Uses a limited number of formats and schemas to reduce the number of use cases and scenarios which must be accounted for and tested to ensure ongoing access
- Provides public access to Repository contents combined with ongoing feedback and communication with users and stakeholders to ensure detection of failures of the Repository to meet the information needs of stakeholders
- Performs continuous access and retrieval tests to verify that access systems work as intended

Appendix A: Glossary of Terms

Archival Information Package (AIP)	A structure consisting of all of the files comprising a preserved digital object and its associated metadata and other control data.
Dissemination Information Package (DIP)	An object or set of objects derived from an AIP for dissemination to a system or end user.

Appendix B: Change Log

2021-07-26	First version