Borealis preservation action plan

Description

Borealis serves as the dedicated repository for research data within the Libraries, powered by the versatile platform of Dataverse and supported by the robust infrastructure of Scholars Portal. Designed to be universally applicable across all academic fields, Dataverse accepts a broad spectrum of data types and formats. Securely housed on Canadian servers, the repository ensures the safekeeping of all deposited files within a protected digital space. Depositors are granted the agency to share their content with the public, select individuals, or maintain its confidentiality, according to their discretion

Unit plan owner

Advanced Research Services

Key staff

Data Curation Librarian

Digital Preservation Librarian

Science and Data Literacy Librarian

URL

https://borealisdata.ca/dataverse/uvic

Content groups

Research data generated by UVic researchers and curated by the Libraries

Monitoring designated community needs

Monitoring the evolving needs of the designated community within the context of preserving research data at UVic Libraries, according to the OAIS (Open Archival Information System) model, involves several key practices:

- Regular communication, engagement, and feedback: The Data Curation Librarian ensures ongoing communication with designated stakeholders, including research project leads and computing personnel. This involves active participation in forums like the UVic Libraries Dataverse Advisory Group, which comprises researchers, librarians, and computing specialists. Additionally, the Data Curation Librarian engages in university-wide committees such as the Research Advisory Council (RAC).
- Repository analysis: The Data Curation Librarian analyses repository usage patterns to anticipate future needs, using the Borealis metrics dashboard.

- Collaboration with researchers: As part of their duties, the Data Curation Librarian regularly collaborates on research projects. This is facilitated by the UVic Libraries' Grants and Awards Librarian position, which maintains the [Library Services for Grant-Funded Research Projects service](#), enabling researchers to include RDM and digital preservation as part of their funding proposals. The Data Curation Librarian and Science and Data Literacy Librarian also provide regular workshops on data management planning, data analysis, visualization, and mapping via the [UVic Libraries' Digital Scholarship Commons](#), enabling them stay informed about participants' evolving research interests, methodologies, and data requirements.
- Attend workshops and conferences and participate in professional associations: The Data Curation Librarian and Digital Preservation Librarian participate in workshops and conferences attended by members of the designated community. These events provide opportunities to learn about emerging technologies, methodologies, and challenges faced by researchers and educators. These positions also participate in several national organizations that connect them with researchers and technology specialists, including via the Canadian Association of Research Libraries (CARL), the Canadian Research Knowledge Network (CRKN), and the Digital Research Alliance of Canada (The Alliance).
- Adaptation and Flexibility: Preservation strategies, metadata standards, access policies, and other aspects of data management are reviewed on an annual basis based on feedback and the evolving requirements as articulate by members of the Dataverse Advisory Group.

Preservation planning

On an annual basis, the Digital Preservation Librarian meets with the Data Curation Librarian to review preservation planning configurations for our [Archivematica](#) preservation system. This includes:

1. Evaluate all open and proprietary data formats within UVic's Borealis datasets using the Borealis Metrics Dashboard, which catalogs file types across collections and their respective dataset locations, each with a unique DOI.
2. Upon completion of the above assessment, refine and update the file normalization procedures for preservation within Archivematica, ensuring compatibility and longevity of archived materials.
3. If identified file formats pose a high risk of obsolescence, the Digital Preservation Librarian will initiate the re-ingestion of associated Archival Information Packages (AIPs) based on the revised normalization procedures established in step 2.
4. Leveraging data from Borealis usage reports, which detail file types and their dataset affiliations, AIPs corresponding to specific Borealis datasets can be readily accessed via their assigned DOIs for reprocessing under the updated normalization guidelines.

Preservation level

Level 2 (normalization) on ingest with Level 3 (migration) available as needed (see Preservation Planning section above)

Data curation policies encourage deposit of open formats where possible. Normalization for oarur data management guidelines recommend using open, widely recognized file formats when submitting data, to ensure that it remains accessible over time. To preserve the longevity of the files we receive, our system, Archivematica, automatically converts certain types of files into standard formats known for their durability. For instance, various unique image files, like those from Adobe or Canon, are converted to a standard TIFF format. We keep both the original and the standardized

version in a secure digital package, along with detailed information about the files, such as their type and any changes they've gone through. This process also applies to video and audio files. PDF documents are similarly standardized to PDF/A, a format designed for long-term preservation. No matter the type of file, we make sure that we have all the necessary information to identify and manage the formats, which allows us to update them to newer formats in the future if required, by simply reprocessing them in Archivematica and adjusting our format guidelines as needed.

Storage level

Level 3: Level 2 + at least one copy in a geographic location with a different disaster threat than the other copies and at least two copies independently managed. At least one copy on a different storage media type

Borealis ensures the robust preservation of data through a strategic storage approach within the Ontario Libraries Research Cloud by Scholars Portal. Here, three synchronized copies of the data are distributed across various higher education data centres throughout Ontario. To safeguard against data loss, Scholars Portal implement a daily backup protocol to tape, managed by IBM Tivoli Storage Manager (TSM). For a comprehensive outline of their preservation practices, please refer to the Borealis Preservation Plan at Borealis Preservation Plan.

Complementing this, UVic Libraries also facilitate additional redundancy through the Narwhal network based at the University of Victoria (UVic), where data is stored in AIPs after processing through Archivematica. This AIP is backed up at UVic using both TSM and Veeam, with a duplicate off-site backup in Kamloops via the Educloud Backup Service. Moreover, for an extra layer of security, an independent backup is maintained on AWS S3 Deep Glacier storage, ensuring our data remains intact and accessible even under the most unexpected circumstances.

Tool chain

- RDA BagIt for Archiving in Dataverse
- UVic Narwhal (NAS)
- Educloud Backup Service
- AWS S3 Deep Glacier
- Archivematica
- Ontario Libraries Research Cloud

Preservation overview

	Actions	Lead responsible	Notes
1	Materials are deposited in repository with oversight by Data Curation Librarian and with support from Scholars Portal staff	Advanced Research Services	
2	On an annual basis, the Digital Preservation Librarian compiles and sends a list of the latest datasets, along with any updates, to Scholars Portal. In response, Scholars Portal delivers 'Bags' — which are comprehensive packages of content — through the Ontario Library Research Cloud (OLRC). These 'Bags' are then downloaded to an interim	Library Systems	Access to staging area for Archivematica controlled by Library Systems.

	holding space, setting the stage for their subsequent transfer into the Archivematica system for long-term preservation.		
3	Dataverse bags are processed through Archivematica and stored as AIPs on NAS (Narwhal) with geo-remote backup, and on AWS S3 Deep Glacier	Library Systems	Access to Narwhal and S3 controlled by Library Systems.
4	All processes and tracked and documented on OneDrive	Library Systems	

Data model/metadata

- When the BagIt script is invoked, Dataverse sends a copy of each published dataset in the form of BagIt archival Bags that adhere to Research Data Alliance standards, commonly referred to as 'BagPacks'. These packages include all file checksums, ensuring data integrity, the original dataset files alongside their preserved versions, and detailed accompanying metadata that details the dataset, including descriptive information, rights, and version history.
- Archivematica creates Archival Information Packages (AIPs) that contain both the original dataset files and their preservation-ready counterparts. These packages are also bundled with their respective checksums to guarantee authenticity and are complemented by extensive metadata, contained within multiple METS files. This metadata encompasses the PREMIS standard for preservation and the Dublin Core (DC) standard, ensuring a comprehensive record of the dataset's provenance, authenticity, and history.

Documentation

Available on request (contact coreyd@uvic.ca for more information).

Audit process

Random samples of objects are extracted from Narwhal and AWS on annual basis and inspected to ensure that the object and associated metadata can be viewed or played back by an end-user.

Audit log

Audit date	Outcome	Outcome details
	[successful/not successful]	[E.g. "xxx random samples met access goals]

Review and succession criteria

The Digital Preservation Librarian will review this action plan annually with Advanced Research Services and Library Systems.

Review date	Attendees	Notes

Last updated: 9 April June 2024, Corey Davis, Digital Preservation Librarian