

Glossary of Terms for Use with Borealis Institutional Collection Policy Templates

Rev. 2024-12-20

Author: Lisa Goddard. Policy Task Group, Dataverse North Expert Group

Reviewers: Émilie Fortin, Amanda Tomé. Preservation Expert Group, Digital Research Alliance of Canada

This document outlines definitions for terms used in the Borealis institutional collection policy templates. These definitions are based on various sources, including:

- [CTS glossary](#)
- [Borealis Preservation Plan - Definitions](#)
- [CODATA terminology](#)
- [Digital Preservation Coalition \(DPC\)](#)

Appraisal: The process of determining whether a dataset has sufficient long-term archival value to merit acquisition, processing, and preservation.

Authenticity: The digital material is the same as it was when it was first created unless the accompanying metadata indicates any changes.

Bit level preservation: The preservation of the file's original bitstream. It may include maintaining copies and backups, virus checking, fixity-checking, and periodic refreshment to new storage media. Bit level preservation does not guarantee future access.

Borealis, the Canadian Dataverse Repository: The instance of the Dataverse platform as a shared service provided in partnership with Canadian regional academic library consortia, institutions, research organizations, and the Digital Research Alliance of Canada, with technical infrastructure hosted by Scholars Portal and the University of Toronto Libraries.

Collection: A collection is a container for datasets (research data, code, documentation, and metadata) and other sub-collections, which can be set up for individual researchers, departments, journals and organizations (from the Dataverse Project with adaptation). See also Institutional Collection.

Content: Includes, without limitation, any information, data, text, software, scripts, graphics, and interactive features generated, provided, or otherwise made accessible through an Institutional Dataverse Provider.

Content preservation (or Active preservation): Processes to preserve the meaning, usability and functionality of the digital content over time (from the [DPC Rapid Assessment Model](#)).

Curation: The activity of managing and promoting the use of data from their point of creation to ensure that they are fit for contemporary purpose and available for discovery and reuse. For

dynamic datasets this may mean continuous enrichment or updating to keep them fit for purpose. Higher levels of curation will also involve links with annotation and with other published materials (from the CODATA and CTS glossaries).

Curator: An individual who has responsibility for managing a set of content in a university's institutional collection in Borealis. This may include but is not limited to making acceptance decisions on submission, auditing and improving metadata, and migrating the content if necessary.

Dataset: A container for a group of related content and files, defined by a theme or category that reflects what is being measured/observed/monitored. The presentation of the data in the application is enabled through metadata (from CODATA glossary). See the Borealis User Guide for [more information](#).

Deaccession: To remove access to a dataset or files, or to suppress, move, or delete the metadata, dataset, or files. Deaccessioning a dataset or dataset version in Borealis will only remove public access to that dataset. It will not delete the files or metadata from the system.

Depositor: A user who submits a dataset to a collection in Borealis. In most cases, the Depositor will be either an author of the submitted dataset, a researcher collaborating with the author(s), or a librarian.

Designated community: An identified group of potential consumers who should be able to understand a particular set of information from an archive. These consumers may consist of multiple communities, are designated by the archive, and may change over time (from DPC Handbook).

Digital preservation: The series of managed activities necessary to ensure continued access to digital materials, including data and metadata, for as long as necessary. Digital preservation is defined very broadly and refers to all of the actions required to maintain access to digital materials beyond the limits of media failure or technological change (from CTS glossary).

Documentation: Information provided with a dataset to enable understanding and use of the data by others, and establish its provenance, history, and context. This might include codebooks, data dictionaries, ReadMe files, and curation and audit logs.

Ingest: The process of entering data and metadata into a data repository.

Institutional collection: One or more collections in Borealis that are under the administration of a single research institution (e.g., university, academic library, or research organization). Each research institution in Borealis has an Institutional Collection that may contain other sub-collections and datasets.

Integrity: The content of a file remains unchanged over time, with the exception of intentional changes made and documented for preservation purposes.

Mediated curation: The repository service creates the collection (or dataset) on behalf of the depositor in Borealis, and the data is curated by the repository service and published once

approved by the depositor, or in some cases (e.g., data rescues, open data curation), by the repository service itself.

Metadata: Information that describes a resource to sufficiently support management, preservation, and use of that resource over time. This includes significant contextual, historical, and technical information.

Migration: The transformation of a file format to a newer version of the content's existing format or to a different format altogether to ensure ongoing access to data as technology becomes obsolete.

Normalization: The process of converting copies of original files to a non-proprietary, widely-used, and/or preservation-friendly format during ingest.

Reappraisal: The process of re-evaluating a dataset at a later date to assess its continued archival value.

Research data management (RDM): Practices covering the entire lifecycle of the data from planning to long term preservation.

Retention period: The minimum duration for which a repository service commits to store and maintain a dataset. After this period, the dataset should be reappraised to determine next steps.

Semi-mediated curation: The repository service creates a collection in Borealis or starts a dataset deposit and assigns a role to the depositor. The depositor submits data to their collection (or dataset). Depending on local policy, the dataset is either flagged for review by the institutional Borealis administrators, or the depositor requests to have the dataset reviewed by the data management team before or after it is published.

Unmediated curation: There is no intervention from the repository service. The depositor creates their own collections and datasets in Borealis, submits their data and publishes it. The RDM service relies on Borealis' built-in features only.