Borealis Administrator Guide

Introduction

The Borealis Administrator Guide is intended to support institutional administrators who manage institutional collections as part of the [Borealis](#) service. The guide was designed to be used in conjunction with the [Borealis User Guide](#) with links provided throughout. More in-depth information is provided about the software features and functionality that directly support the activities of institutional administrators, which may include:

- Managing and updating your institutional collection page, including theme (logos, colours) and description
- Establishing institutional policies and guidelines around the use of your institutional collection
- Administering the institutional collection, including determining permissions for affiliated users for depositing and publishing datasets, as well as responding to user questions from your institution
- Managing local services associated with the institutional collection, including data curation activities (if any)
- Providing oversight of your institutional collection, ensuring users at your institution follow the [Borealis Terms of Use](#) when depositing data into your institutional collection, and informing Borealis staff when necessary repository actions are required.

Getting Help

If you are not an administrator of a Borealis collection designated by your institution, please see our [User Guide](#) for guidance on navigating the interface and depositing and using data or [contact support](#) with questions.

As an institutional administrator, there are a number of support options available to assist you in your role. For any technical or administrative questions, contact the Borealis team directly by submitting a completed form from the [Borealis home page](#). Form submissions are monitored daily and responses are usually provided within 2 business days.

You can also email the Borealis Community list-serv. The list-serv is monitored by the Borealis team as well as all other institutional administrators. Use the list-serv if you have a general question you'd like to pose to fellow administrators or to start a discussion about a feature or function. If you are not a member of the listserv, [contact us](#) to be added.

Resources

Refer to the [Support & Additional Resources](#) page in the User Guide for links to a variety of useful documents and websites related to Borealis.

SP Policies and Other Documentation

See the [Borealis About page](#) to review the policies and guidelines that govern the Borealis service.

Community Meetings

Community support for institutional administrators and interested members is facilitated by the Borealis Team and the [Dataverse North Expert Group](#), which is part of the Digital Research Alliance of Canada's Research Data Management Network of Experts.

The Borealis Team includes managers, librarians, software developers, and technical systems administrators. The Borealis Team is responsible for the overall operation and maintenance of the repository.

The Dataverse North Expert Group brings together Dataverse providers, libraries, and administrators to coordinate local and national training, working groups, support services, outreach strategies, promotions, and community development.

The Community Facilitation Team is composed of members of the Borealis Team, the Dataverse North Chair, and staff from the Digital Research Alliance of Canada. Engagement with the community is made possible through regular monthly virtual meetings and the email listserv.

Join our monthly community meetings for updates about development work, platform upgrades, projects, and community news. These meetings will provide lots of opportunities to ask questions and participate in discussions. [Contact us](#) for more information or to suggest agenda topics.

Supporting Institutional Users

As your institution's Borealis administrator, you are the first level of support for users from your institution. There are two main ways that users can contact you or your institutional administrator team through the interface:

- When a user clicks the **Contact** button on your institutional collection page, their message is sent to the email address(es) listed in your contact information of your [institutional collection](#).

- When a user clicks the **Support** button on the main Borealis menu (available on all Borealis pages), their inquiry is sent to the Borealis Community Listserv (currently with over 100 members) to be triaged.

As your institution's Borealis administrator, you should monitor incoming end user emails on the listserv for any from members of your institution. Institutional affiliation is normally included in the email subject line. When an incoming email is from a member of your institution, it is expected that you respond to that email/user within 2 business days. When responding, you can either cc the listserv, so that all institutional administrators can see the response given, or let the listserv know you've answered the user off-list.

The Borealis team will make efforts to connect with the institutional administrators. If there is no response to an end user's email within 2 business days, Borealis will answer the email (cc the institutional contacts) and attempt to assist the end user. If the question is related to institutional-specific policies or procedures, we may only be able to forward the inquiry to your institutional email account to ask for further assistance.

Note: As an institutional administrator, in order to see the same things as a 'regular' user, log out of your account or [create a separate standard account](#) for yourself that has no permissions or roles assigned.

Dataverse Terminology

Many terms and features that appear in the Borealis interface are described throughout this guide, and apply to one or more of the following:

- A **collection** (also known as a **Dataverse collection**) is a container in which one or more **sub-collections** and/or one or more datasets can be stored. Each institution has their own **institutional collection**. In previous versions of the Dataverse software, collections were referred to as **dataverses** and sub-collections as **sub dataverses**.
- A **dataset** is a container in which one or more files, as well as the associated metadata for those files, can be stored.
- The **Dataverse** platform refers to the open-source software that the Borealis repository runs on. It is maintained by the [Institute for Quantitative Social Sciences](#) (IQSS) at Harvard University.
- **Borealis** or the **repository** refers to the Dataverse instance hosted by Scholars Portal to support open discovery, management, sharing, and preservation of Canadian research data. This version is in production and is distinct from our [Demo repository](#) for testing.

- An **account** refers to a [user account](#) for accessing the full holdings and features of the Borealis repository.

Since Borealis is the new name of the platform, should I add the name Borealis to the name of my institutional collection?
To ensure a consistent and clear user experience in the Borealis interface, especially the carousel, it is very important that the name of your institution is included in the name of your institutional collection in order to orient users. Including Borealis in the name of your collection would cause confusion. The Borealis team recommends maintaining existing naming conventions for the name of your institutional collection.

If you would like to include "Borealis," consider updating your description via the **[Edit > General Information](#) menu**, for example, "This University Dataverse Collection is part of Borealis, the Canadian Dataverse Repository, and provides a research data repository for affiliated researchers…"

The Demo Repository

In addition to the **Production** version of Borealis, the **[Demo](#) repository** is available to anyone to test or try out the features and functions of the Dataverse software as a sandbox.

The Demo repository is a completely separate installation with key differences and details to keep in mind:
- Accounts are not connected between Production and Demo.
  - If your institution uses [standard accounts](#), users will need to create a new account on Demo.
- Demo should not be used for storage of real datasets as the storage may be wiped periodically.
- DOIs created as part of dataset citations are not minted and will not resolve.
- Datasets cannot be migrated from Demo to Production.
- Permissions settings in Demo may be different compared to production:
  - Unlike in Production, the root collection in Demo is open for any user to create collections and datasets for instructional or example purposes.
  - If you would like assistance with your institutional collection in demo, [contact the Borealis team](#).

User Accounts

Types of User Accounts

Borealis provides two types of [User Accounts](#): a Standard account and an Institutional account.

- [Standard accounts](#) are available to anyone with an email address, including those with an institutional affiliation within the repository.
- [Institutional accounts](#) (also known as Shibboleth accounts) allow users to login using the same username and password that they use to access their institutions online services (e.g., single sign-on). Users can select their institution from the dropdown list on the [Log In](#) page and follow the prompts to enter their institution's credentials. For more details, see [Implementing Institutional Login](#).

Email affiliation with Standard Account

Standard accounts must have a unique username and email address. Users who are from an affiliated institution should use their institutional email address to create their standard account and [verify their email address](#) in order to ensure:

- They will be redirected to the institutional collection upon login
- They will receive automatic [permissions](#) for your institutional collection based on affiliation group (if configured)

Non-affiliated users will be directed to the Borealis root collection and would need to have manual permissions applied to deposit data.

Implementing Institutional Login

Institutional accounts are only available to organizations who are registered with the [Canadian Access Federation](#) (CAF) via [Canarie](#) and have configured both [Federated Identity Management](#) (FIM) and the [Research and Scholarship (R&S) Entity Category](#) with CAF.

In some cases, organizations are registered with the CAF and have configured FIM, but lack the [R&S entity category](#) that is required for institutional/Shibboleth accounts. Adding the R&S entity category, in these situations, can be done at no extra cost. However, you will need to [contact your IT department for assistance](#).

Once configured with CAF, Scholars Portal receives an automated notification. That institution will then be set up for institutional accounts and the main institutional administrators will be notified.

Converting User Accounts

Can a user switch from a standard to an institutional account?

Yes. If the institution has set up Institutional Accounts, any user with a standard account that uses an affiliated institutional email address can switch to an institutional account. The email address in the standard account must match the address provided by Shibboleth, otherwise a new account is created.

Can a user switch from an institutional to a standard account?

Yes. Contact the Borealis team if a user with an institutional account would like to switch to a standard account, using the same institutional email address. Each email address can only be linked to one account (standard or institutional) in Borealis.

How do I ensure access for a user who is leaving my institution, but still needs to access their collection and/or dataset?

The user would need to create a standard account using a different email address (i.e., not their institutional email). Once the new account is created, contact the Borealis team to request for their institutional account to be merged into their new standard account, which will retain the appropriate roles and permissions on their data in the new account (and delete the old account).

If the user moves to another institution that uses institutional accounts, they would need to create a new institutional account and request for their accounts to be merged into the new account (and delete the old account).

How do I request accounts be merged?

To have accounts merged, contact the Borealis team and include the usernames of the accounts and which account the user wants to keep.

The account that is not being kept will be merged into the other account. All the permissions and roles associated with the old account will be transferred to the permanent account. The process of merging accounts deletes the old account from the repository.

Updating and Managing Account Information

Can a user change the email address associated with their account?

No, once an account is set up with a specific email address, the email address cannot be changed. Instead, the user would need to create a new account with a different email address, and request the two accounts to be merged.

How does a user change their password?

For institutional accounts, a user will need to follow your institution's process for changing a password for an institutional account. Institutional account users cannot change their password in Borealis.

For standard accounts, a user can [change their password](#) using the **Edit Account** > **Password** button under the **Account Information** tab, or by clicking **Forgot Your Password?** on the login page.

What other information can a user change?

 A standard account user can [change their password](#), their given and family names, and their position. They **cannot** change their email address, affiliation, or username.

Institutional account users cannot change any account information.

Finding & Exploring Data

Basic and Advanced Searching

Why do I get different results when I search Borealis when I'm logged in versus when I'm not logged in?

When someone uses the public-facing Borealis website without logging in, the only collections and datasets they can see are those that are published and publicly available. Therefore, the search results displayed will only include published collections and datasets.

When someone uses Borealis when logged in, they may also see unpublished or draft collections or datasets for which they have access. Therefore, the search results displayed may also include those unpublished collections and datasets.

If you share the URL for a search query with someone who has different permissions than you, they may see different results. For example, as the administrator for your institutional collection, you will be able to see all unpublished collections and datasets within your institutional collection. But a researcher at your institution will only be able to see unpublished collections and datasets for which they have access.

Previewing Files

Why am I only sometimes prompted to agree to the terms of use before I'm able to preview (or download) a file?

When there is text entered into a [dataset's **Terms of Use** field](), a pop-up window will appear asking the user to accept those terms before they preview or download any of the files within that dataset. The **Terms of Use** field may be empty if the dataset has a [CC0 Public Domain license](), or if the dataset owner did not enter the **Terms of Use** for a custom license.

When **Terms of Use** have been entered, they will appear to every user, every time they preview or download a file. Users with permissions/roles to edit the dataset will not see the terms when downloading files.

Downloading & Citing Data

What is the maximum amount of data a user can download at a time?

Users of Borealis cannot download more than 1,000 files or 5 GB of data at a time, whichever limit is reached first. Attempting to download more than this amount will result in an error message.

Keep in mind, however, that even downloads smaller than the limits may cause problems if the user's internet connection speed is less than optimal. Users working off-campus will likely experience slower downloads than those working on campus.

If an entire dataset exceeds these size limits, an error message will appear in the **Access Dataset** menu directing you to select and download files from the files table.
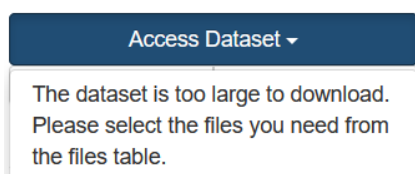
*Alt-text: The error message that appears in the **Access Dataset** menu when the dataset is too large to download.*
Filename: Downloading_Restricted_MoreThanOneFile-1 / Downloading_Restricted_MoreThanOneFile-1-fr

A warning message will appear above the files table if the set of files selected exceeds either download limit. The message will indicate the number of files selected for download, the size of the selected files, and the zip limit for a single download. This information may help you determine how much data you can include in a single ZIP file.
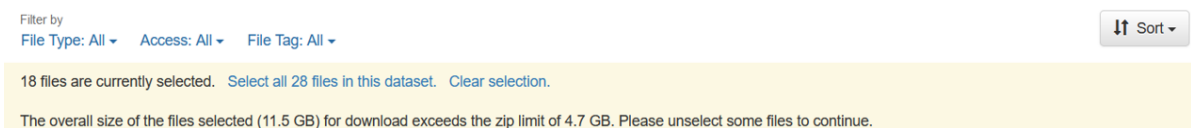
Alt-text: A yellow warning banner appears above the files table if the amount of data selected for download exceeds the limits.
Filename: Downloading_Restricted_MoreThanOneFile-2 / Downloading_Restricted_MoreThanOneFile-2-fr

The warning does not prevent users from clicking the **Download** button. However, clicking the button will result in an error message stating that the user is attempting to download too much data at once. The message will direct users to either reduce the amount of data selected for download or to use the Data Access API for programmatic access to the files.

Download Options ✖

❗ The files selected are too large to download as a ZIP.

You can select individual files that are below the 4.7 GB download limit from the files table, or use the Data Access API for programmatic access to the files.

*Alt-text: If too much data is selected, clicking the **Download** button will result in an error message.*

Filename: Downloading_Restricted_MoreThanOneFile-3 / Downloading_Restricted_MoreThanOneFile-3-fr

Criteria for Depositing Data

Borealis Terms of Use

The [Terms of Use](#) for Borealis is an important document that outlines the various terms and conditions all users and depositors must follow in order to use the repository. The Terms of Use are displayed when users [create a new account](#). They can also be reviewed on Borealis at any time, for example before users [make a deposit](#) or [publish a dataset](#).

Among other things, some important items to remember from the Terms of Use include:

- Borealis [does not allow any confidential or sensitive information](#) to be uploaded and stored in the repository.
- Any data must be anonymized or de-identified before it is uploaded to the repository.
- If any identifying information is stored on Borealis, it should be accompanied by an explanation of the consent requested by the research project and/or the consent provided by the research subjects.

What do I do if a user from my institution has contacted me to say they've violated the Terms of Use?

The [User Guide](#) notes that if a user determines they have violated the Terms of Use in any way, they should contact their institutional administrator immediately. If you receive such a notification, please [contact the Borealis team](#) immediately and explain the situation that's occurred.

In some cases, the violation may be easily remedied (e.g., a user shared their password with someone, or they realized they uploaded a file with personal information). In other cases, the violation may require that the user's account be cancelled and/or their access restricted. Accounts that have been cancelled cannot be reinstated.

Institutional Policies, Guidelines or Agreements

Where do I find information about policies and guidelines related to Borealis?

The [Borealis About page](#) provides details about the Borealis service, including the mission statement, governance and organization, designated community, and policy documentation.

- [Preservation Plan](): The preservation plan outlines the objectives, roles and responsibilities, strategies, and actions for preserving the digital files uploaded by users and stored in the repository.
- [Privacy Statement](): This privacy statement explains what data we collect through the use of Borealis, how this information is treated, and for what purpose.
- [Technology Infrastructure and Security Information](): The Technology Infrastructure and Security Information document outlines general information, technical infrastructure, application security, and storage and backup details.
- [Terms of Use](): The Terms of Use governs all use of the Borealis service.

For additional questions or inquiries, please [contact us]() for assistance.

What sorts of policies and procedures might it be useful for my institution to implement in order to help manage our institutional collection?

The [Dataverse North]() Policy Working Group developed [policy templates]() (shared CC-BY) to assist academic libraries in developing a comprehensive policy framework for their institutional collections. This local policy layer is designed to complement broader policies, such as the Borealis [Terms of Use](), that apply to all collections within Borealis.

Such policies and procedures that may be helpful to develop include (but are not limited to):

- **Collections Development Policy or Guidelines** -- which might include guidance on what types of research data will and will not be kept in your institutional collection. Such a document may align with your library's collections development policy or your library's strategic priorities.
- **Deposit Guidelines** -- which might include limitations or restrictions on the file formats allowed for deposit, procedures regarding the way in which research data must be submitted for deposit, and requirements as to the amount and type of documentation required to accompany a deposit. Such guidelines may be particularly important if you receive a large number of requests for deposit, as they could be designed to streamline the work required by yourself or other library staff.
- **Depositor Agreement** -- which might include a set of terms and conditions determined between the library and the researcher which outlines the roles and responsibilities of each party. Such an agreement might include restrictions on who can view/download the data, how long the data will be kept for, and what is and isn't allowed to happen to the data.
- **Curation Policy or Procedure** -- which might include the type of review you or the library staff go through whenever a researcher deposits data into your institutional

collection. Tasks might include a review of the data being deposited, a review of the documentation included with the data, re-running any code or analyses included with the data, etc. Such procedures may be linked to your deposit guidelines.

- **Deaccession Policy or Guidelines** -- which might include guidance around when a dataset or collection is or can be deaccessioned or even deleted. Such guidance may be linked to your collections development and/or preservation documents.
- **Preservation Policy or Guidelines** -- which might include guidance on where research data is stored for the long-term, including other repositories the data may be transferred to or how the data might be transformed or reformatted. Such guidance may be linked to your collections development document.

Can my institution create our own Terms of Use for Borealis?

Yes. Any participating institution can create policies or procedures that go above and beyond the terms and conditions outlined in the [Borealis Terms of Use](). Such policies and procedures could include:

- Additional limitations on who from your institution can create a collection and/or dataset.
- Additional limitations on the types of projects and/or research data accepted for publication in the repository.
- Restrictions on who can be assigned specific permissions and roles within your institutional collection.
- Additional procedures on the curation and review of datasets before they're uploaded or published in the repository.
- Restrictions or limitations on the file formats accepted for deposit.
- Requirements as to the amount and type of documentation that must be included in a dataset.
- Additional requirements as to the minimum metadata required for a dataset and/or file.
- Additional agreements as to the way in which research data from your institution is preserved, deaccessioned, and/or managed.

Participating institutions, however, cannot create policies or procedures that conflict with the [Borealis Terms of Use.]()

Is there a limit to how much data a researcher can deposit in Borealis?

Borealis does not restrict how much data an individual researcher can deposit. However, there are a number of factors to consider with researchers who want to deposit large

datasets to determine if their dataset is best suited for Borealis or whether another repository might be better suited (e.g., Federated Research Data Repository (FRDR) or a disciplinary repository). A comparison table of repository features is provided at the end of this section.

We recommend working with researchers to determine the expected users of the data, their potential technical capabilities, and anticipated use cases (e.g., will users want to download the entire dataset or just portions of it) when deciding on the most appropriate deposit location.

Storage: Each participating institution has a base storage allocation and can increase the storage at the OLRC rate. Therefore, It is up to the institution to determine whether datasets should be deposited based on institutional collection policies and guidelines.

Uploading data:
- Currently, the maximum file size for Borealis is 3 GB. Files larger than this cannot be uploaded into Borealis.
- For large datasets, researchers will likely want to use APIs or the command line bulk uploader, DVUploader, to deposit their data. If the connection stalls, you can restart the tool and it will continue to upload (Note: the Borealis team has not tested the DVUploader beyond 35-50 GB).
- We recommend discussing technical capabilities and the state of the dataset (including all of its documentation) with researchers when determining where to deposit data.

Downloading data:
- For large datasets, users will not be able to download all files in the dataset at once through the browser or using the Data Access API due to the zip bundle limit of 5 GB.
- Users may select files to download in chunks of 5 GB (an error message in the interface will inform users when their selection exceeds the download limit).
- A command line tool like wget can also be used to download all the files at once, however the file hierarchy will not be maintained.
- The Borealis team has developed a Python script to download the entire dataset with the file hierarchy intact, however this script has only been tested with datasets up to 35 GB.
- For large datasets, we recommend including instructions in the dataset documentation for these methods so that users are aware of how to access the data. Please note that this approach may be a barrier for some users to obtain data.

	Disciplinary Repository	Borealis	FRDR
Scope	Specialist (re3data.org provides a registry of research data repositories)	Generalist	Generalist
Storage	Capacity will vary Storage likely outside of Canada Built to handle specialized datasets	Total storage available varies by institution Canadian storage	Supports storage of big data Canadian storage
Uploading data	Eligibility, pricing, repository functionality will vary May only accept certain file types	Max file size of 3 GB API and command line options for bulk uploading File-level restrictions and versioning available Open to researchers at participating Canadian institutions	No max file size Globus transfer tool for bulk uploading of large files Only supports open data No versioning available Open to faculty at Canadian institutions
Downloading data	Repository functionality and access requirements will vary	Max zip bundle limit of 5 GB Command line and Python script options to download full dataset	Globus transfer tool for bulk downloading of large files

Can sensitive data be deposited in Borealis?

There are a number of factors to consider with researchers who want to deposit sensitive or confidential data in Borealis. This includes but is not limited to datasets with data on human subject participants, vulnerable populations, Indigenous communities, at-risk species, politically sensitive topics, and data restricted by third-party licensing agreements.

The sharing of sensitive and confidential data in Borealis is addressed in the following section of the Terms of Use:

- By using the Service, you confirm that all **your Content does not contain information that could directly or indirectly identify a subject, except where the release of such identifying information has no potential for constituting an unwarranted invasion of privacy and/or breach of confidentiality**.

You may also wish to consult with the relevant institutional policies and procedures, including Research Ethics Board guidelines.

If you and the researcher determine that their dataset meets this requirement, Borealis could be used to deposit and share.

Some Borealis features that may be helpful to consider include:
- [Custom Terms of Use section](#) to define access requirements for a dataset (see example custom agreements in the [Appendix](#)).
- [Restricted files](#)
- [Guestbooks](#) for datasets
- Dataset documentation can include a sample of the consent form to help secondary users understand what participants have agreed to for data sharing and reuse

For further guidance on sensitive data sharing on Borealis, please see our Sensitive Data Guide (forthcoming).

Collections

Managing Your Institutional Collection

Every institution that participates in Borealis is provided with their own institutional collection. This institutional collection is overseen and managed by institutional administrators. As an institutional administrator, you are provided with the **Admin** role for your institutional collection. The **Admin** role allows you to do the following within your institutional collection:

- Edit your institutional collection.
- Manage permissions and groups for sub-collections and datasets.
- Add and edit sub-collections.
- Add and edit datasets (and files).
- View unpublished (draft) sub-collections and datasets.
- Publish sub-collections and datasets.
- Delete unpublished and published sub-collections and files.
- Delete unpublished datasets and deaccession published datasets.

The **Editing a Collection** section of the User Guide describes the step-by-step processes you can use to manage and edit your institutional collection. All of these processes can be accessed by clicking the **Edit** button in the top, right-hand corner of the main page for your institutional collection.
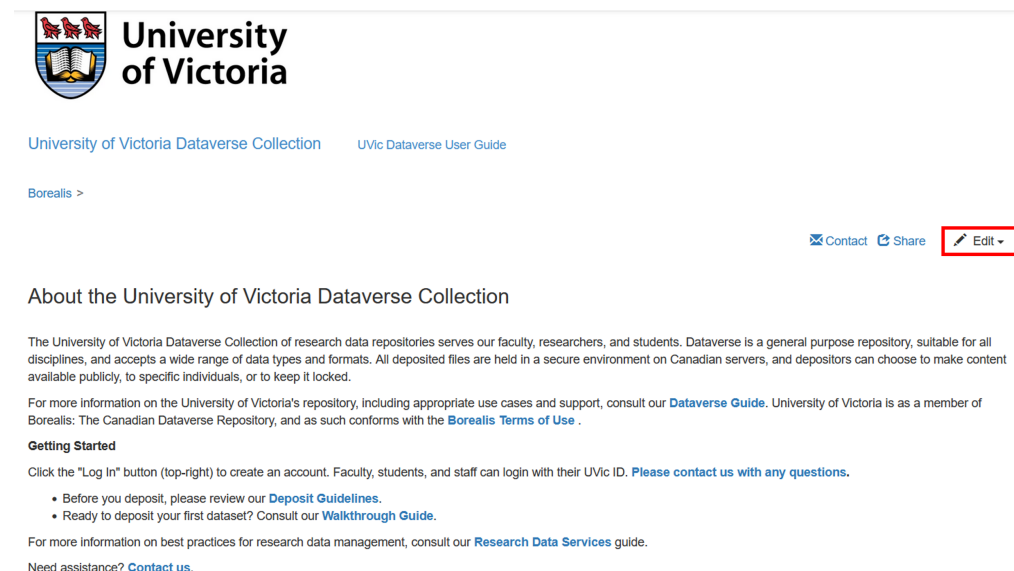
*Alt-text: An example of the main landing page for an institutional collection with the **Edit** button highlighted. Filename: Collections_ManagingInstitutional-1 / Collections_ManagingInstitutional-1-fr*

When clicked, the **Edit** button will display a drop-down menu with items you can edit or change.
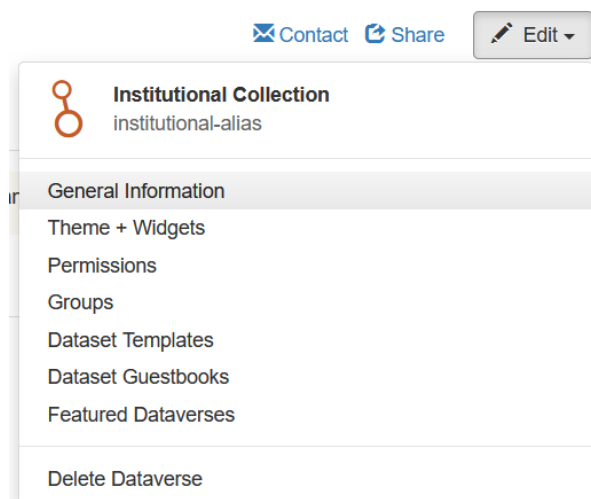
*Alt-text: The clicked **Edit** button on a collection page, with the drop-down menu options displayed.*
Filename: Collections_Managing / Collections_Managing-fr

General Information

Under the collection's **Edit > General Information,** there are a number of fields available to edit and update. If you require a change to the name or identifier for your institutional collection, please [contact us](#).

Email Field(s)

The **Email** field should include one or more email addresses for institutional administrators. These are the email addresses messages are sent to if a user clicks on the **Contact** link to the left of the **Edit** button. While any email address can be used in these fields, you may want to consider using a generic email address or email alias so you don't have to change this field when staff changes.

Description Field

The **Description** field can contain text and some HTML tags. Any text entered in this field will be displayed on your institutional collection's main page, under the **Contact** and **Share** links and above the **Search this dataverse …** field.

Your description is important to orient users to your institutional collection. Therefore, it is extremely helpful to include some or all of the following information in your **Description** field:

- A brief explanation of your institutional collection, including who can deposit.
- A website link to your library's RDM-related material, or a link to the website that contains all your [Dataverse-related policies and procedures](#).
- A generic email address which researchers can contact if they have any questions.

Themes + Widgets

Every collection starts with a default theme, however you may want to customize your institutional collection. Users may also customize their sub-collections to inherit the theme of the institutional collection or to better highlight the associated projects.

Under the collection's **Edit > Themes +and Widgets** menu**,** the look and feel, or the theme, of a collection can be customized with the addition of a logo, text and background colours, taglines and websites, and a footer image. Some configurations, such as adding the institutional logo, will be set up by the Borealis team when your institutional collection is created.

For logos and footer images, supported file types are JPG, TIF, or PNG and should be no larger than 500 KB. The maximum display size for an image file in a theme is 940 pixels wide by 120 pixels high.

For more details, see [Themes and Widgets](#).

Metadata Fields

Under the collection's **Edit > Dataset Templates** menu, the list of available templates will be listed. By default, the checkbox for *Use metadata fields from Borealis* will be checked, which means that all available metadata schemas and fields are available for user entry. This also means that only some of the fields under **Citation Metadata** -- (Dataset) Title, Author Name, Contact Email, Description, and Subject -- will be required or mandatory for all users. None of the other fields under **Citation Metadata** will be required, nor will any fields from any of the other five (5) metadata schemas.

Note: If you deselect "Use metadata fields from Borealis," you will not be able to use the [CC licensed dataset templates](#) in your collection or sub-collections.

If you would like to change which fields are available and/or required for all users of your institutional collection, uncheck the box beside "Use metadata fields from Borealis." You can then go into each metadata schema section, select the fields you want displayed, and select whether those fields will be required or optional for users. Once saved, your selections will apply to all datasets created within your institutional collection.

The **Metadata Fields** can be changed or edited for all collections, including sub-collections within your institutional collection. This means that an administrator of a sub-collection can [change what metadata fields](#) are required for datasets within that sub-collection. Changes to the required metadata fields will only apply to new sub-collections and datasets and will

not affect existing sub-collections and datasets. If you do not want sub-collection administrators to change **Metadata Fields**, you will either need to ensure they're not given the **Admin** role and/or are informed of what they are and aren't allowed to change within their sub-collection.

Browse/Search Facets

On the main page for Borealis, all institutional collections, and sub-collections, there is a set of facets or filters listed on the left-hand side of the page. When the checkbox beside *Use browse/search facets from Borealis* is checked, the default facets/filters -- Author Name, Subject, Keyword Term, Deposit Date, Time Period Covered Start, and Time Period Covered End -- will be displayed on your institutional collection's main page.

If you would prefer to display a different set of facets/filters, you'll need to uncheck the box for *Use browse/search facets from Borealis*. Once unchecked, you'll be able to add and remove metadata fields from the **Selected** box on the right.

Once saved, the **Selected** metadata fields will be used as facets/filters on the main page of your institutional collection, as well as any sub-collection.

Dataset Templates

If you select **Dataset Templates** from the **Edit** button's drop-down menu, you'll be able to make changes to the default dataset template used for all new datasets in your institutional collection.

Default Dataset Template

When the checkbox beside *Include Templates from Borealis* is checked, you'll have six (6) dataset template options listed in the chart on this page. Five (5) of these template options are related to Creative Commons licenses (CC BY-NC-SA, CC BY, CC BY-SA, CC BY-NC-ND, and CC BY-NC). The sixth template option is for a *Custom License.* In addition, the repository software's default dataset template option is the CC0 Public Domain license.

When the CC0 license is used for datasets, no terms and conditions can be applied to the use of the data and the **Terms of Use** field can remain empty. However, if the CC0 license is not applied, the specific terms and conditions of the use of the data in the dataset must be provided in the **Terms of Use** field.

Borealis provides dataset templates for the other types of Creative Commons licenses, as well as a custom license dataset template. When any of the Creative Commons license dataset templates are used, text is automatically entered into the **Terms of Use** field. When

the custom license dataset template is used, the **Terms of Use** field remains empty until the user enters text.

If you would prefer to have a default dataset template other than the Public Domain license, you can click the **Make Default** button beside any of the six (6) templates of your choice.

When the checkbox beside *Include Templates from Borealis* is unchecked, the list of dataset template options on this page will disappear. If you do not add any new dataset templates, the only option available to users will be the Public Domain license template.

New Dataset Templates

Regardless of whether the checkbox beside *Include Templates from Borealis* is checked or unchecked, you do have the ability to create a new dataset template for your institutional collection.

*Note: Once you create a new dataset template, it is not automatically set to be the default. If you want to use your new dataset template as the default template, you will need to click the **Make Default** button.*

Guestbooks

Guestbooks are created and managed at the collection level. Researchers who wish to use a guestbook for their dataset located directly within an institutional collection would need to work with the institutional collection administrator to set-up the guestbook and to obtain the responses. If more than one dataset uses the same guestbook, all responses will be grouped together.

The **Dataset Guestbooks** page in your institutional collection allows you to create a new guestbook or download/view the responses to a dataset guestbook.

Deleting a Collection

If a collection does not contain any datasets or sub-collections, it can be deleted by clicking **Edit** on the collection page, and selecting **Delete Dataverse**.

If a collection is not empty, then you will first need to either move the datasets or sub-collections to another collection, or delete them.

To move a collection, contact the Borealis team. In your request, please indicate:
- The url of the collection in its current location.

- The url of the collection that you would like the collection to be moved to.

If the collection being moved has a guestbook, template, metadata block, link, or featured sub-collection that is not compatible with the destination collection, the settings of the collection being moved may be removed.

Implementing a Collection for a Journal, Association, or Multi-Institutional Project

In addition to institutional collections, Borealis is able to host collections for journals, professional associations, or multi-institutional projects that require a separate promotional and or multi-institutional space. These collections would not be sub-collections within an institutional collection, but would rather be at the 'same level' as an institutional collection.

Each journal, association, or multi-institutional project collection must be sponsored by an existing institution within Borealis. That sponsoring institution will then provide the same type of administrative support to the journal, association, or multi-institutional project collection as they do their own institutional collection. Or, the sponsoring institution may decide to train their sponsored group to do this administrative support themselves, with the sponsoring institution being available for second level support.

To request the set up of such a collection, please contact us with the following information:
- The name of the sponsored organization.
- The names and email addresses of the main contacts within that organization.
- The name and email address of the main contact that organization has at your institution.
- The name the sponsored organization would like to give their new collection.
- An image file (preferably in PNG format) of the logo the sponsored organization would like to use in their collection.

Initial Storage Allocation

Each new journal, association, or multi-institutional project collection will be provided with an initial storage allocation of 100 GB. If that collection requires more storage space in the future, it would need to be requested, and possibly paid for, by the sponsoring institution.

Agreements or MOUs

The sponsoring and sponsored organizations must complete a written agreement or memorandum of understanding that outlines roles and responsibilities of both organizations. This is especially important for two reasons:

1. Such an agreement could provide details of a succession plan that would outline what happens to the data if the journal, association, or multi-institutional project ceases to exist or can no longer manage the collection.
2. If the sponsored organization wishes to apply for CoreTrustSeal Certification for their collection, they will need to have such an agreement documented for more than one of the requirements.

An agreement/MOU template is available upon request.

Anonymous Review

For users interested in using Borealis to support an anonymous peer review process, the anonymous review feature generates private URLs that provide access to unpublished datasets without logging in. Users can create a **URL for anonymized access**, which conceals metadata fields that contain identifying information about the author.

The following fields are masked:
- Author name
- Depositor name
- Contact name
- Producer
- Production Place
- Distributor

This feature is available to dataset administrators and curators. Under the **Edit dataset** menu on a dataset page, select **Private URL**. In the pop-up window, select **Create URL for Anonymized Access**. In the next window, select **Copy to Clipboard** to copy the URL for sharing. Reviewers are able to view and download files in the dataset using this link.
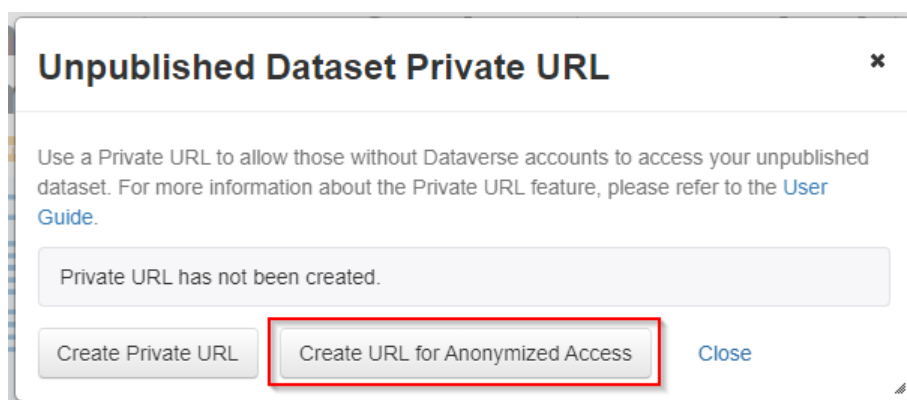
*Alt-text: The **Unpublished Dataset Private URL** window, with the **Create URL for Anonymized Access** button to generate a private URL for anonymous review highlighted.*
Filename: Collections_Anonymous / Collections_Anonymous-fr

Once the review process is complete, return to **Edit Dataset > Private URL** and select **Disable Private URL** to deactivate the link, or publish the dataset.

Important Note: the parent collection(s) for the dataset will appear in the dataset page breadcrumbs, which may reveal some information about the researcher or research group. If this presents obstacles to your peer review process, please contact us to set up a double blind review collection.

*Alt-text: The names of the parent collection(s) for the dataset still appear in the dataset page breadcrumbs when the dataset is accessed using the **URL for anonymized access**. This may reveal some information about the researcher or research group.*
Filename: Collections_Anonymous-1 / Collections_Anonymous-1-fr

See the Advanced User Guide to learn more about anonymous review.

Datasets

As described in the [Permissions](#) section, there are three main approaches to managing dataset deposits in your institutional collection. Depending on the permissions you have configured in your institutional collection, users may not be able to create their own datasets ([Mediated](#)). Users may be able to create a dataset, but they may or may not be able to publish on their own without your approval ([Unmediated](#) vs.[Semi-mediated](#)).

Creating a New Dataset

Please refer to the **[Creating a Dataset](#) section** in the User Guide for step-by-step instructions on how to create a new dataset within your institutional collection or sub-collection.

If you have set up your institutional collection such that users are **not** able to create their own datasets, you will need to assign the proper users to the newly created dataset in order for them to be able to add data and files. If you do not wish those users to be able to publish the dataset without your permission or review, you will likely want to [assign them the **Contributor** role](#).

With the **Contributor** role, users will be able to add files to the dataset, edit the dataset's metadata, and submit the dataset for review before publishing.

Reviewing a Dataset

As part of the [semi-mediated](#) approach, users would be able to create a dataset, but not have the ability to publish on their own. When users click the "Submit for Review" button, you will receive a notification for you and your team to review.

Notifications will appear by email and the **Notifications** section of your Borealis account. A link in the notification will direct you to the dataset in question. Note that the notification will not indicate if the submission is a new dataset or a new version resulting from a change to a published dataset. View the **Versions** tab in the dataset page to determine the status of the submission.

For the review process, consider consulting the [Digital Research Alliance of Canada's Curation Guide](#) along with your institutional policies to ensure the dataset is structured and documented according to best practices. You and your team may also wish to develop a checklist or rubric based on your policies to ensure a consistent, comprehensive, and unbiased review. Sample checklists can also be found in the [Curation Guide](#).

You may also consider developing publicly available documentation about your data curation policies and practices. This will help researchers at your institution understand the review process and steps they can take to conform their data prior to submission. This documentation will also support your institution's CoreTrustSeal application, if you choose to seek certification.

After reviewing the dataset, you can choose to either "Publish" the dataset or "Return to Author." If the dataset is returned, the user will be notified that changes are needed before the dataset can be submitted for review again. The interface currently does not provide a way to explain **why** the dataset is being returned. It is highly recommended that you contact the user (via email) to provide more details **prior** to clicking "Return to Author." Alternatively, you may use the API to provide the user with a reason for return. For information about using the API, please see the API Guide.

Once the dataset is published, the user will be notified that the dataset has been published. If you are publishing a new version of a dataset, we recommend selecting *Minor Release* if only the metadata has been changed and *Major Release* for substantive changes to files and/or content.

Moving a Dataset or Collection

If you or a user would like to move a dataset from one collection to another (or a collection from one location to another), please contact the Borealis team to request the move. In your request, indicate:

- The doi of the dataset being moved OR the URL of the collection to be moved.
- The url of the collection where the dataset is currently located.
- The url of the destination collection.

If a guestbook is associated with the dataset or the dataset is linked to another collection, the guestbook and link may be removed if they are not compatible with the collection that the dataset is being moved to.

Files

File Requirements or Standards

Borealis supports a variety of file types and formats, including tabular data files, research code, shapefiles, Flexible Image Transport System files, and compressed files. Certain file types are supported by additional functionality, which can include downloading in different formats, previews, file-level metadata preservation, and file-level data citation.

Please refer to the **Adding Files to a Dataset** section in the User Guide for more detailed discussion of file requirements, standards, and limitations.

Tabular Data Files

Tabular data uploaded to Borealis in certain file types – Stata, SPSS, R, Excel (.xlsx), and CSV – will be ingested and converted to non-proprietary tab-delimited files (.tab) that can be used by any tabular software. This tabular data format allows users to visualize and explore data using the Data Explorer – a statistical data exploration application integrated in the Dataverse platform.

Some file types or data formatting may result in incomplete .tab files or ingest errors that prevent .tab files from being generated at all. For more information about supported tabular file types and data formatting requirements, see the Advanced User Guide.

Tabular file ingest

Uningesting files

If you have a tabular file in a dataset that has been ingested but you would prefer it to appear in its original format, contact the Borealis team to uningest the file.

File-level Embargoes
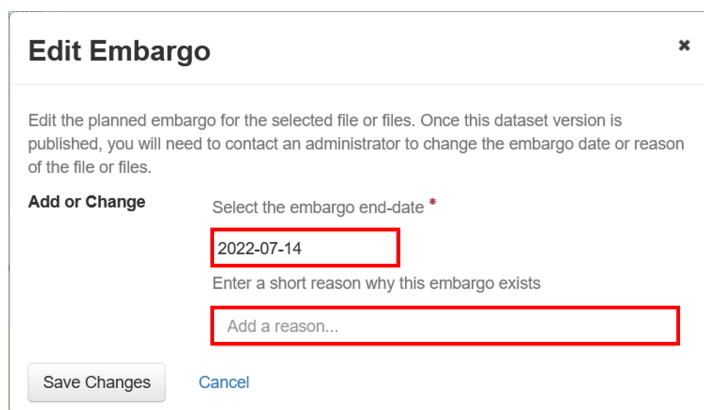
Embargoes make files inaccessible after a dataset is published until the embargo period end date. Embargoes may be applied for various reasons such as article publication plans or agreements with funders or industry partners.

The effect of adding an embargo is similar to making a file restricted, meaning that file previews and downloads and requests for access are prevented until the embargo period is over. File metadata is available during this period.

Embargoes are configured at the file level and can be applied to individual files or a set of selected files. Therefore, in order to set a dataset-level embargo, users would need to set the same embargo date on all files in the dataset.

Embargoes can only be set, changed, or removed before a file has been published.

Users with **Administrator**, **Curator**, and **Contributor** roles may access the embargo feature. To configure an embargo, select the relevant files and navigate to **Edit Files > Embargo**. In the pop-up window, type in the end-date using the YYYY-MM-DD format or use the calendar widget to select the end-date of the embargo and enter a short explanation for the embargo in the text box. Select **Save Changes** to set the embargo.

*Alt-text: The **Edit Embargo** pop-up window for adding or editing the embargo end-date and reason for the embargo for selected files.*
Filename: Files_Embargo / Files_Embargo-fr

Once the dataset has been published, the embargo period is displayed in the file metadata and in the file access menu. The embargo cannot be changed after the dataset is published.
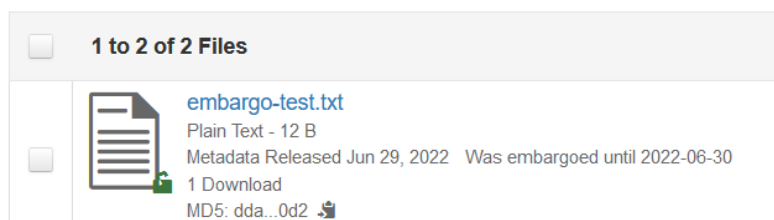
Alt-text: An example of a file within a dataset displaying the date that file metadata was released and the end date for the embargo period.
Filename: Files_Embargo-1 / Files_Embargo-1-fr

For more details about embargoes, see the [Advanced User Guide](#).

Managing Access to Restricted Files

Individuals who wish to access files in a published dataset that are restricted – either permanently or due to an embargo -- may be granted access following approval from the data depositor or creator. Only users with the **Contributor Plus**, **Curator**, or **Administrator** roles may enable access to restricted files.

Users may request access in two ways:

- Selecting the **Request Access** button
  - If a user selects the **Request Access** button, an email notification will be sent to users of the dataset with the required roles. Depending on the roles and permissions established for your institutional collection, you and your team may be required to facilitate this process or depositors may be able to directly manage requests.
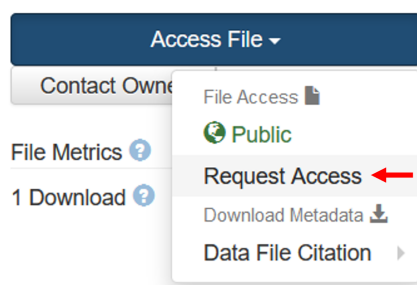
*Alt-text: The **Request Access** option in the **Access File** drop-down menu.*
Filename: AccessRestricted_RequestAccess-1 / AccessRestricted_RequestAccess-1-fr

  - If the data depositor or creator agrees to grant access, go to the **Edit** menu on the dataset page and select **Permissions > File**. In the **Requests** section, select **Grant** for each approved request.
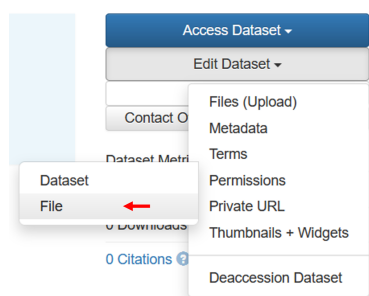
*Alt-text: The menu option to edit **Permissions** for a **File** in the **Access Dataset** drop-down menu.*
Filename: AccessRestricted_RequestAccess-2 / AccessRestricted_RequestAccess-2-fr
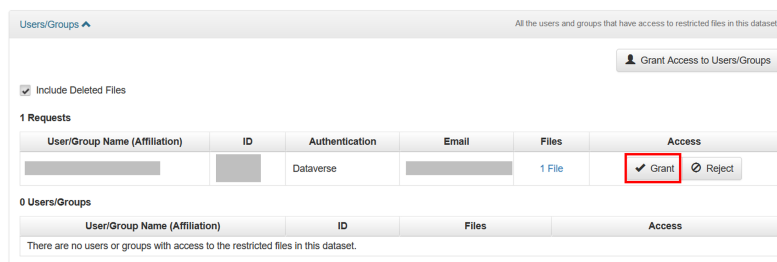
*Alt-text: **Grant** or **Reject** requests for access in the **Users/Groups** section of the Permissions page.*

Filename: AccessRestricted_RequestAccess-3 / AccessRestricted_RequestAccess-3-fr

○ Users must have a Borealis account to access this pathway.
- Requests submitted outside of the Borealis interface
  ○ Requests can also be received outside of the Borealis interface – for example, if the requestor selects **Contact** to send an email to users with the required roles. If access is approved, the requestor must first create a Borealis account to receive access.
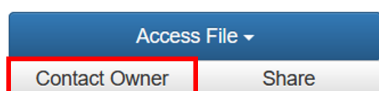
*Alt-text: The **Contact Owner** button on a file page.*

Filename: AccessRestricted_ContactOwner-1 / AccessRestricted_ContactOwner-1-fr

  ○ Once their account is created, go to the **Edit** menu on the dataset page and select **Permissions > File**. In the **Restricted Files** section, select **Assign Access** to open the **Grant File Access** dialogue. In the window, use the **Users/Groups** field to search for and add the user. Select the checkbox(es) next to each file that the user should have access to, then click **Grant**.

*Alt-text: Grant specific users or user groups access to restricted files by selecting the **Assign Access** button in the **Restricted Files** section of the Permissions page.*

Filename: AccessRestricted_ContactOwner-2 / AccessRestricted_ContactOwner-2-fr
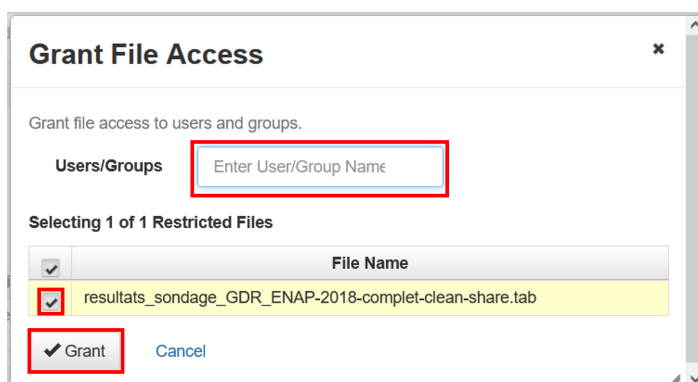
*Alt-text: The **Grant File Access** pop-up window with the **Users/Groups** field to select users*

and user groups and checkboxes to select restricted files to grant access to.
Filename: AccessRestricted_ContactOwner-2 / AccessRestricted_ContactOwner-2-fr

Publishing Data

Data Curation Functions

The Dataverse software provides an extensive number of curation-related features and functions that can help ensure users deposit [FAIR](#) (findable, accessible, interoperable, and reusable) data. The majority of these features are built into the Dataverse software, which means they replace the need to perform these tasks manually. However, while these features may be built-in, some of them have to be turned on or set up before they can be used in your institutional collection.

From a Dataverse perspective, the following features can be used to help make research data FAIR:

- Metadata
- Access - Permissions and Groups
- Dataset Templates
- Dataset Guestbooks
- Licenses or Terms of Use
- System-generated Citations
- Restrictions and Terms of Access
- Private URLs
- File Formats and File Ingestion
- Persistent Identifiers
- Version Tracking
- Tombstone Records (for deaccessioned datasets)

Creating a Collection

The following items can be selected/entered when a collection is created, but can also be edited *after* the collection is created.

1. **Required Citation Metadata Fields**:
   a. Dataverse (Collection) Name
   b. Identifier (URL)
   c. Category
   d. Email
   e. Citation Metadata required fields for datasets - determined at the system level, cannot be made optional.
2. **Optional Citation Metadata Fields:**
   a. Affiliation
   b. Description

c. Citation Metadata optional fields for datasets - can be made required for all datasets within the collection.

d. Other metadata fields for Geospatial, Social Science and Humanities, Astronomy and Astrophysics, Life Sciences, and Journal metadata schemas - can be made required for all datasets within the collection.

e. Browse/Search Facets

See the [Publishing Data](#) section of the user guide for more details.

Editing a Collection

The following items can be altered once a collection is created.

1. **Metadata Fields**:
   a. All metadata fields, required or optional.
2. **Permissions**:
   a. Default/inherited permissions from the parent collection.
   b. User-based permissions and assigned roles.
3. **Groups**:
   a. Default/inherited groups from the parent collection.
   b. Creation of group(s) specific to this collection.
4. **Dataset Templates**:
   a. Default/inherited dataset templates from the parent collection.
   b. Selecting a default dataset template for this collection.
   c. Creation of a new dataset template specific to this collection.
5. **Dataset Guestbooks**:
   a. Creating guestbooks for datasets within this collection.
   b. Viewing guestbook entries.

For **Permissions**, note that some roles may allow users to make changes to required versus optional fields and items in a specific collection. This means, depending on the role assigned, some users may be able to circumvent requirements setup by institutional administrators. See the [Permissions](#) section of the guide for more details.

Creating a Dataset

The following items can be selected/entered when a dataset is created, but can also be edited *after* the dataset is created.

1. **Dataset Template**:
   a. Selection of the dataset template to be used for the new dataset. Note that you cannot change the template for a dataset once the dataset has been created, but many of the fields within the template can be edited and new fields added. Changes to a template will not be applied to existing datasets

using that template. If a dataset template is deleted, datasets created using that template are also not affected.

    i.   System-level default dataset template includes a CC0 (Public Domain) license.

    ii.   Five standard dataset templates are available which each include a Creative Commons license.

    iii.   One standard dataset template is available which includes the ability to provide a custom license.

2. **Required Citation Metadata Fields**:
   a. (Dataset) Title
   b. Author Name
   c. Contact Email Address
   d. (Dataset) Description
   e. Subject

3. **Optional Citation Metadata Fields**:
   a. Author Affiliation
   b. Author Identifier Scheme and Identifier
   c. Contact Name
   d. Contact Affiliation
   e. Description (Effective) Date
   f. Subjects
   g. Keywords
   h. Related Publications
   i. Notes
   j. Depositor
   k. Deposit Date

4. **Other Required Metadata Fields**:
   a. Any metadata fields (from Geospatial, Social Science and Humanities, Astronomy and Astrophysics, Life Sciences, and Journal metadata schemas) which are set as required at the collection level.
      i. Note that a metadata field from one of the listed schemas that is left as optional will not appear for entry during dataset creation. Instead, the field can be entered after the dataset is created.

5. **System-generated Citation**:
   a. Automatically created when a new dataset is created. Will be updated automatically if citation fields are changed.

It is recommended that undergraduate and graduate students include a faculty sponsor on their datasets to facilitate long-term access and management.

How Dataset Templates affect the Licenses and Terms of Use Fields

The dataset template used to create a new dataset directly influences which license and **Terms of Use** fields are available and/or include content. As six of the seven standard dataset templates include the name of the Creative Commons license in the title, it is not recommended that those templates be used if the license may be changed later. The Custom License dataset template is the most flexible template for future changes.

When a Creative Commons license dataset template other than the default one for CC0 Public Domain is used to create a dataset, the **Waiver** field is set to "No" and the **Terms of Use** field is pre-populated with text associated with that specific CC license.
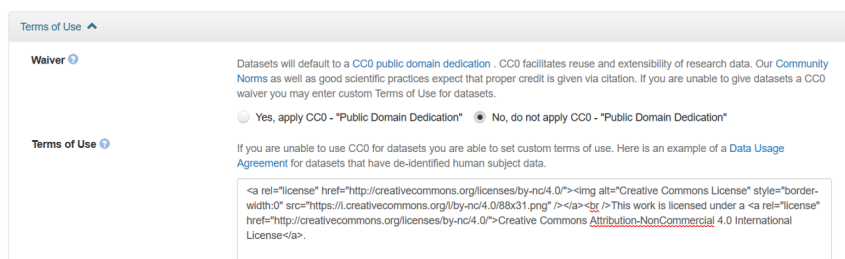
*Alt-text: The editable **Terms** tab for a dataset that was created with the CC BY-NC 4.0 dataset template. The **Waiver** field is automatically set to "No" and text populates the **Terms of Use** field.*
Filename: Publishing_Template-3 / Publishing_Template-3-fr
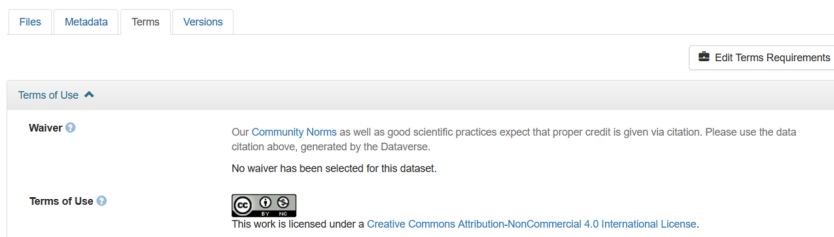
*Alt-text: The public view of the **Terms** tab for a dataset that was created with the CC BY-NC 4.0 dataset template. The tab indicates that "No waiver has been selected for this dataset" and text about the CC BY-NC 4.0 license (including the graphic) is displayed.*
Filename: Publishing_Template-4 / Publishing_Template-4-fr

When the custom license dataset template is used, the **Waiver** field is set to "No," but the **Terms of Use** field is blank. If the custom license dataset template is used, information about the terms of use or license must be entered manually by a user. If the **Terms of Use** field is left blank, there is technically no license in effect for that dataset.

Once a dataset is created with any dataset template, the **Waiver** and **Terms of Use** fields can be altered. However, Creative Commons licenses cannot be revoked. If a dataset is published with a CC0 license, and that license is later changed to a more restrictive license (e.g., CC BY-NC 4.0), users who already accessed the dataset when it had the CC0 license are not required to change how they use the dataset. Therefore, it is vital the proper license

and terms of use are entered for a dataset before it is published. If in doubt, opt for a more restrictive license, so that any changes made to the license in the future will make the dataset more accessible.

Custom Terms of Use

Selecting the custom license dataset template allows users to define a variety of licenses, agreements, and other restrictions that users must agree to before receiving full access to the dataset.

Defining terms of use will provide secondary users with a clear understanding about whether and how they can access data, what they can and cannot do with the data, and any additional requirements they must fulfil when accessing and reusing data.

Custom terms may be used in conjunction with other features, including file-level restrictions and guestbooks, to implement restrictions.

The following fields are available in the interface for users to define conditions for access and use:

- **Terms of Use:** how data can be used once downloaded
- **Confidentiality Declaration:** whether a confidentiality declaration needs to be signed to access the dataset
- **Special Permissions:** any special permissions required to access a resource and how to obtain those permissions
- **Restrictions:** any restrictions on access to or use of the collection, such as privacy certification or distribution restrictions, that are applied by the author, producer, or disseminator of the data collection. This includes if the data are restricted to only a certain type of user, such as for academic researchers only
- **Citation Requirements:** special and/or explicit citation requirements for data to be cited properly in articles or other publications that are based on analysis of the data
- **Depositor Requirements:** information regarding user responsibility for informing dataset depositors, authors, or curators of their use of data by providing citations to the published work or providing copies of the manuscripts
- **Conditions:** any additional information that will assist the user in understanding the access and use conditions of the dataset
- **Disclaimer:** information regarding responsibility for uses of the dataset

Alternately, users may create a custom data use or licensing agreement to address these and any other conditions. The agreement can be entered in its entirety in the **Terms of Use** field. Users may wish to consult their institutional legal and ethics offices when drawing up custom agreements.

These custom terms may also be replicated or adapted in the **Terms of Access** field for restricted files in the dataset.

See the [Appendix](#) for example custom terms of use agreements for:
- Datasets containing de-identified human subject data
- Datasets containing identifiable or quasi-identifiable data that may only be reused for academic research purposes

Editing a Dataset

The following items can be altered once a dataset has been created.

1. **Metadata Fields**:
   a. All metadata fields, required or optional.
2. **Fields in the Terms of Use Section**:
   a. Terms of Use (field)
   b. Confidentiality Declaration
   c. Special Permissions
   d. Restrictions (description)
   e. Citation Requirements
   f. Depositor Requirements
   g. Conditions
   h. Disclaimer
3. **Fields in the Restricted Files + Terms of Access Section**:
   a. Terms of Access (field)
   b. Toggle to enable or disable the requirement to request access.
   c. Data Access Place
   d. Original Archive
   e. Availability Status
   f. Contact for Access
   g. Size of Collection
   h. Study Completion
4. **Guestbooks**:
   a. Option to select and preview a guestbook for the dataset. Dataset administrators can select to include a guestbook if they want, but they must select from a list of pre-developed guestbooks. They cannot create a new guestbook.
5. **Permissions**:
   a. Default/inherited permissions from the parent collection.
   b. User-based permissions and assigned roles for this dataset.

6. **Private URL**:
   a. Create or disable a private URL for unpublished datasets.

Curation Labels

Curation labels can be applied to a draft dataset version to indicate its status as part of a curation process. The standard set includes the following curation labels:

- Author contacted
- Privacy review
- Awaiting paper publication
- Final approval

Labels are visible as colored tags on the dataset page and in collection listings/search results to dataset admins and curators. Administrators receive notifications when the curation status labels are changed. Internally, the labels have no effect, and any existing label is automatically removed when the dataset is published. Administrators can use a reporting API call to obtain a list of datasets and their curation statuses.

Dataset administrators and curators can access the labels through the **Publish Dataset menu.** Under **Change Curation Status**, select the desired label or **Remove Current Status.**
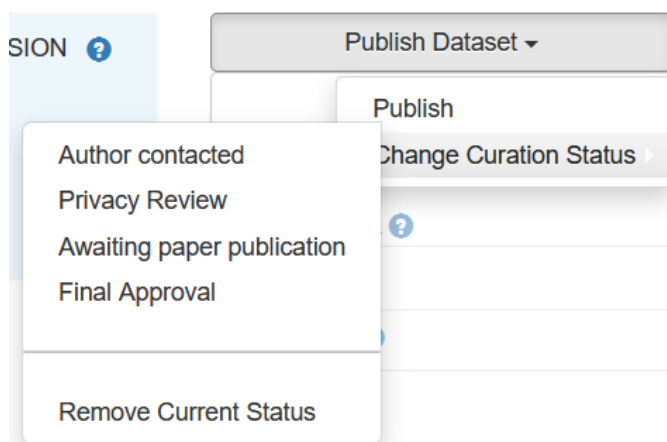
Alt-text: T*he **Publish Dataset** menu with t*he Change Curation Status feature highlighted and the curation labels and Remove Current Status options visible.
Filename: Publishing_Curation / Publishing_Curation-fr

Adding Files to a Dataset

The following items can be selected/entered when a file is uploaded, but can also be edited *after* the file is uploaded.

1. **Required Metadata Fields**:
   a. File Name

2. **Optional Metadata Fields**:
   a. File Path
   b. (File) Description
   c. Provenance File
   d. Provenance Description
   e. File Tags
3. **File Restrictions**:
   a. Option to select one or more files to be set as restricted once uploaded.
4. **Terms of Access**:
   a. If one or more files are set to be restricted once uploaded, the Terms of Access field will be available for entry during the file upload process.
5. **System-generated Citation**:
   a. Automatically created when a new file is uploaded. Will be updated automatically if metadata fields are changed.

Editing Files in a Dataset

The following items can be altered once a file has been uploaded to a dataset.
1. **Metadata Fields**:
   a. All required and optional metadata fields.
2. **File Restrictions**:
   a. Option to select one or more files to be set as restricted.
   b. Option to select one or more currently restricted files and set them as unrestricted.
3. **Terms of Access**:
   a. If one or more files are selected to be restricted, the Terms of Access field will be available for entry when the **Save Changes** button is clicked.

For more details about editing files in a dataset, see the User Guide.

Publishing a Dataset

The following items are activated or created when a dataset is published.
1. **Persistent Identifier**:
   a. A DOI is automatically reserved when the dataset is first saved. The DOI is minted through the DataCite DOI service and will resolve once the dataset is published.
2. **Public Access**:
   a. Unrestricted files within the dataset become public.
3. **Metadata**:
   a. Dataset metadata becomes searchable within Borealis.

b. Metadata used within the dataset becomes [harvestable by discovery platforms](#).

4. **Version Tracking**:
   a. Changes made to the dataset or the files or metadata within the dataset will be tracked and, when the changes are published, [a new version of the dataset will be issued](#).

Replacing Files in a Published Dataset

Once a dataset is published, the **Edit > Replace** feature can be used to update or revise existing files in the dataset. Replacing a file also generates a new dataset draft version, and this feature creates an explicit link between the previous and current versions of the dataset to track the history of the file across versions of the dataset.
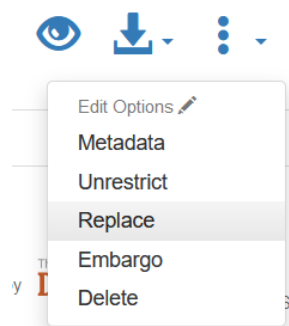
*Alt-text: The **File Options** menu with the **Replace** function highlighted.*
Filename: Publishing_ReplacingFiles-1 / Publishing_ReplacingFiles-1-fr

Files with the same checksum, or algorithmically assigned identifier, may be uploaded to a single dataset. If a user attempts to replace a file with a file that has the same MD5 checksum as another file in the dataset, a warning message will be displayed. If a user attempts to replace a file with a file that has the same checksum, however, an error message will appear and the file will not be replaced.

Deaccessioning a Dataset

[Deaccessioning](#) a dataset (or a version of a dataset) is available to users with the **Admin** or **Curator** role. Given the nature of this serious action, deaccessioning should only occur if there is a legal or valid reason for the dataset to no longer be accessible to the public.

Institutional administrators may be required to deaccession datasets if they are alerted to issues including:

- Violations of the [Borealis Terms of Use](#)
- Copyright violation
- Legal requirements and proven violations

- Research misconduct (e.g., falsification of data)
- Confidentiality

Please review the User Guide for step-by-step instructions on how to deaccession a dataset.

The following items are changed or updated when a published dataset (or a version of a dataset) has been deaccessioned.

1. **Public Access**:
   a. Files within the dataset are no longer accessible to the public. Administrators of the dataset can continue to view and download the files from a deaccessioned dataset.
2. **Dataset Record**:
   a. A tombstone record which includes the DOI, basic metadata, and citation for the dataset remains in Borealis.
   b. Metadata remains searchable and harvestable.

Metadata Harvesting

The Dataverse platform supports open APIs and the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) to allow published collections and datasets and their associated metadata to be harvested by other digital repositories and search engines for global data discovery.

Metadata in Borealis is regularly collected and published by FRDR, Google Dataset Search, Datacite (datasets published with DOIs only), Web of Science Data Citation Index, Mendeley Data (datasets published with DOIs only, via Datacite), Harvard Dataverse, and OpenAire (via FRDR).

Note that datasets indexed by Google will be discoverable through Dataset Search, a search engine dedicated to dataset discovery. Researchers are unlikely to find datasets from Borealis in Google Scholar, which focuses on text-based research outputs.

Borealis also works with the Shared Access Research Ecosystem (SHARE) to integrate public datasets into open web discovery services. The SHARE notification system is a higher education-based initiative that strengthens the effort to identify, discover, and track research output.

Metrics

Metrics in Borealis are available at the collection and dataset levels through a custom Scholars Portal application. A monthly email alert is sent to the listserv when new data is available. Metrics provide a useful snapshot of repository usage and reach, however it is important to note that metrics alone are not comprehensive measures of impact and engagement.

Metrics are currently available in English only.

Collection Metrics

The **Metrics** button in the main Borealis search interface directs you to a [dashboard](#) for monthly statistics for the Borealis repository as a whole and for top-level collections within it. Use the **Select Dataverse Collection** dropdown to see metrics for your institutional collection.

Alt-text: The header of the Borealis search interface, with the **Metrics** *link highlighted.*
Filename: Publishing_CollectionMetrics / Publishing_CollectionMetrics-fr

The dashboard displays the most recent 12-month period by default. Historical reports can be accessed by selecting **Previous Months**. The dashboard also include visualizations for the following metrics:

- Monthly downloads
- Number of datasets in the sub-collections of the parent collection
- Size in GB of sub-collections of the parent collection
- Range and distribution of file types identified in the collection
- Range and distribution of subject categories present in the collection

If you would like to save a copy of the data, use the **Download** feature to download an Excel spreadsheet with data from the last year, totals for the last year, and all time totals. The different sheets and fields available in the file are described in the first sheet of the file for FAQ.

You can filter the spreadsheet to include only data from your institutional collection, and perform any other kinds of calculations that are helpful for analysis. Some of the other sheets in the file include:

- Downloads by dataverse and dataset
- File types
- Subjects
- Count of users by affiliation

Dataset Metrics

Dataset Metrics are found on the dataset page, below the **Access Dataset** menu.

Dataset Metrics ❓

15 Views ❓

3 Downloads ❓

0 Citations ❓

*Alt-text: The **Dataset Metrics** table on a dataset page, showing the number of **Views**, **Downloads**, and **Citations** for the dataset.*
Filename: Publishing_DatasetMetrics / Publishing_DatasetMetrics-fr

Depending on permissions, users may be able to see the following metrics in this section:

- **Views**: Includes dataset views, aggregated file views, and file downloads. Dataset views are only visible to users who are logged in.
- **Downloads**: Each file downloaded is counted as 1, and added to the total download count. The total number of downloads is visible to users who are logged in and those who are not.
- **Citations**: Number of citations for the dataset that are retrieved from Crossref via DataCite using Make Data Count standards, see the [Advanced Guide](#) for more details. Users may select the metric to see a list of citation URLs.

Metrics API

An API to retrieve metrics is also available through the Dataverse software. These API calls can be run directly in the browser and don't require command-line knowledge. They can return data in JSON format or CSV, depending on which command you're using.

Through these APIs, you can get details about your collections, datasets, files, and download counts. Use your institutional alias to filter the data for your institutional collection. If you are interested in using the Metrics API, please review the [documentation](#) to see what commands are available to filter your results. For an overview of what APIs are and how they work, please see the [Dataverse API section](#) of the guide.

Note: The API currently is not working for download metrics, and the Borealis team is following up with the Harvard team to resolve the issue.

Permissions

As the administrator of your institutional collection, you are able to assign roles and define permissions for users of your collection.

The **Permissions** section of the User Guide contains details about Roles, Groups, Collection Permissions and Dataset Permissions. We recommend that you review these sections prior to reviewing the additional information in this guide.

Roles

Default Roles

Under **Edit > Permissions,** the **Roles** section lists all the default roles available for selection in the collection and what tasks each role is allowed to perform. These roles can be assigned to individual users or user groups, and more than one role can be assigned. The roles you assign to your institutional collection depends on the institutional policies and procedures you've created for your collection.

Creating a New Role

Additional roles with different permissions may be created to meet the needs of your institution and users. Contact the Borealis team to have a custom role setup for your collection.

Groups

Creating and Editing an Admin Group

Permissions within an institutional collection can be managed by individual users or by groups. To manage the administrative permissions for support contacts at your institution, you should have been set up with an "Admin Group." An Admin Group will streamline the process to add or remove individual administrators from your institutional collection, as you won't need to manually update access to each sub-collection that exists within your institutional collection.

Setting up an Admin Group is the same as Creating a Group. The main difference is the role assigned to the group, which would be the **Admin** role.

Going forward, whenever a user is added or removed from an Admin Group, the permissions in all sub-collections are automatically updated to reflect the current membership of that group. No other changes are required.

Once an Admin Group is set up, you'll also want to delete existing institutional administrators from the list of permissions. Click **Remove Assigned Role** in the rows for each individual administrator you want to remove from the sub-collection.

Note: Adding and removing permissions will generate notification emails to all users whose permissions have changed. Changes to groups will not generate notification emails.

Other Group Types

IP Groups (&IP)

Each institution has an IP group that is made up of users that access Borealis within a specific IP range (e.g., while on campus or using a VPN). The use case for the IP Group is to restrict access to files using the "File Downloader" role to users within the IP Group. Within the dataset permissions menu under "Assign Roles to Groups/Users", search for an IP group by entering the institution name, or enter "&ip" to see all of the IP groups. Select "File downloader" for the role and click "Save changes."

Affiliation Groups (&aff)

Affiliation groups are made up of user accounts that are using an email domain associated with an institution, and that have also verified their account. For example, the Queen's University Affiliation Group is made up of users who have a verified account with an e-mail address ending in "@queensu.ca". An account is considered verified if the user has verified their email address, usually done during account creation. For recommendations on using affiliation groups, see [Using Affiliations and Shibboleth Groups for Permissions](#).

Shibboleth Groups (&shib)

Shibboleth groups are available for any institution that has [enabled single sign-on](#) for institutional accounts. This group contains all user accounts associated with the institution that are using this form of login. In order to be in this group, a user with a standard account must convert their account to use institutional login. For recommendations on using Shibboleth groups, see [Using Affiliations and Shibboleth Groups for Permissions](#).

Custom Groups (&explicit)

Any user that can manage permissions in a collection can create their own group of users accounts, and then assign permissions to this group rather than to individual users.

Institutional Collection Permissions

Overview of permissions settings

Borealis has flexible permissions settings, which means the settings are customizable depending on the policies and workflows of your individual institution.

From your institutional collection page, through the **Edit > Permissions** option, under the **Edit Access** menu, you can decide the requirements that determine 1) which users can add datasets and/or sub-collections to your institutional collection, and 2) what default permissions they'll be granted on the datasets that they are able to create within your institutional collection.

Edit Access   ✕

Who can add to this dataverse?
- ⦿ Anyone adding to this dataverse needs to be given access
- ○ Anyone with a Dataverse account can add sub dataverses
- ○ Anyone with a Dataverse account can add datasets
- ○ Anyone with a Dataverse account can add sub dataverses and datasets

When a user adds a new dataset to this dataverse, which role should be automatically assigned to them on that dataset?
- ⦿ Contributor - Edit metadata, upload files, and edit files, edit Terms, Guestbook, Submit datasets for review
- ○ Curator - Edit metadata, upload files, and edit files, edit Terms, Guestbook, File Restrictions (Files Access + Use), Edit Permissions/Assign Roles + Publish

Save Changes   Cancel

*Alt-Text: The **Edit Access** window that appears when you want to edit the options in the **Permissions** section.*

Filename: Permissions_PermissionsSection_Step1 / Autorisations_8

1. The first question "Who can add to this Dataverse" allows you to determine how open your institutional collection is to new additions – you can set whether or not **the entire user base (all logged in users)** has the ability to add datasets or sub-collections.
2. The second question on this page allows you to choose the default role (and thus the permissions) granted to users who can add a dataset to your institutional collection. The role you select will be automatically granted to any user who creates a dataset in your institutional collection, for that dataset, at the moment that he or she creates it. The role the user is given determines his or her permissions for the dataset they've created.
   a. The key difference between the two roles is that **Curators** can publish their own datasets, while **Contributors** must submit the dataset to be reviewed before publication.

Based on these settings, there are three main approaches to managing your collection:

Mediated

- Selecting "Anyone adding to this Dataverse needs to be given access" restricts users by default from uploading data directly to the institutional collection. This option would mean that users would need to contact your institution and admins would need to [create sub-collections](#) or [datasets](#) on behalf of your affiliated users and apply appropriate permissions.
  - With this option, the selection between the **Contributor** or **Curator** default roles is irrelevant, since only administrators will be able to create and publish new sub-collections and datasets.

Semi-mediated (most-common)

- Selecting "Anyone with a Dataverse account can add datasets" with the default role of **Contributor** provides a semi-mediated deposit. This is the most common setting in Borealis. Users can create a dataset within your institutional collection, but they need to [submit datasets for review](#) and have it published by an administrator on your collection. This setting does not allow users to create sub-collections. Therefore, an administrator would need to set-up any required sub-collections and apply permissions on behalf of users manually.
- If you would like only affiliated users to be able to create a dataset and submit for review, follow the steps [Using Affiliation and Shibboleth Groups for Permissions](#).

Unmediated

- There are a few permutations of the permissions settings that provide an unmediated approach to deposit, which allows users to create and publish sub-collections and datasets without review. **We do not recommend a completely open deposit**, as any user with an account could upload malicious or inappropriate content that goes against the [Terms of Use](#). If a more open deposit model is needed at your institution, **we strongly suggest** that you follow the steps [Using Affiliation and Shibboleth Groups for Permissions](#) to restrict deposit to only affiliated users.
- Note that if users are able to create sub-collections within your collection, they will become an Admin on those sub-collections and be able to create and publish datasets on their own. As an administrator on the institutional collection, your admin role will be inherited onto any sub-collections and datasets created.
- For any unmediated type deposit, institutional administrators are expected to closely monitor all activity and deposits to ensure that they follow the Borealis [Terms of Use](#).

Using Affiliation and Shibboleth Groups for Permissions

If you would like to restrict deposits to users with your institutional affiliation, use your [affiliation group](#) or [Shibboleth group](#) to do so.

Semi-mediated using affiliation and Shibboleth groups

This option will only allow affiliated users to submit datasets for review:

- Under **Permissions**, select **"Edit Access."**
  - For the first question, select "Anyone with a Dataverse account needs to be given access."
  - Ensure **Contributor** is selected under "which role should be automatically assigned to them on that dataset."
- Under **Users/Groups**, select "Assign Roles to Users/Groups."
  - Search for the affiliation group or Shibboleth group for your institution and select the appropriate group.
- Assign the appropriate role to the group:
  - In the "Assign Roles to Users/Groups" window, select **Dataset Creator** for the role.
- Review the [semi-mediated](#) deposit description above for expectations of Admins.

*Alt-Text: The **Permissions** window with the setting "Anyone adding to this dataverse needs to be given access" and the University of Waterloo shibboleth group is given the role of Dataset Creator.*

Filename: Permissions_Institutional_Semi / needs screenshot in french (filename once screenshot is captured: Permissions_Institutional_Semi-fr)

Unmediated using affiliation and Shibboleth groups

To allow only affiliated users to create and publish collections or collections and datasets:

- Under **Permissions**, select "Edit Access."
  - For the first question, select "Anyone with a Dataverse account needs to be given access."

- ○ For "which role should be automatically assigned to them on that dataset," in the "Edit Access" window, select **Contributor** or **Curator** role
  - ■ The **Contributor** role allows users to create draft datasets and upload files within your institutional collection. Users will still need to [submit datasets for review](#) and have them published by you or your team.
  - ■ The **Curator** role allows users to publish their own datasets in your institutional collection. Review the [unmediated](#) deposit description for expectations of Admins.
- Under **Users/Groups**, select "Assign Roles to Users/Groups."
  - ○ Search for the affiliation group or Shibboleth group for your institution and select the appropriate group.
- Assign the appropriate role to the group:
  - ○ In the "Assign Roles to Users/Groups" window, select:
    - ■ **Dataverse Creator** for the role OR
    - ■ **Dataverse + Dataset Creator** for the role

Managing deposits in your institutional collection

As an institutional collection administrator, you are expected to monitor the activity within your institutional collection. To assist with this process, admins will receive notifications via email and in the platform for various actions (depending on permissions settings), such as:

- When a user creates a dataset within your institutional collection
- When a user submits a dataset for review
- When access has been requested for a restricted file

If you are not receiving email notifications, verify the permissions to confirm that you are an administrator for the collection or sub-collection.

Preservation

Scholars Portal provides minimum level preservation activities for files deposited in Borealis to ensure materials are preserved at the bit-level. This level of preservation prioritizes retaining the intellectual content of files in storage through regular fixity checking as well as distributed storage of copies in multiple locations to protect against data loss or corruption. These activities do not guarantee the future usability or accessibility of files based on their content or format.

Bit-level preservation is provided for all deposited data files, including those associated with draft or restricted datasets and different versions of files. It does not cover files automatically generated by the Dataverse platform, including derivatives, thumbnails, and citation metadata files.

See the Borealis [Preservation Plan](#) for full details on how Scholars Portal provides bit-level preservation for the digital files uploaded by users and stored in the repository.

File Fixity and Authenticity

The Dataverse platform includes various functions to guarantee the authenticity and integrity of deposited data over time, from ingest to archival storage to access and use. When users upload files to Borealis, each file is assigned an MD5 checksum, an algorithmically generated identifier, which enables verification of file fixity over time.

For tabular data files, citation metadata as well as DDI (Data Documentation Initiative) variable-level metadata are also created during the ingest process. Variable-level metadata can be exported as an XML file or in a DDI-formatted HTML codebook. Derivative tabular formats (tab-delimited or RData) created by the Dataverse software are assigned a [UNF](#) (Universal Numeric Fingerprint) instead of an MD5 checksum. The UNF is generated based on the semantic content of the file and independent of the format or location of the content within the file and the storage format.

File format identification also occurs at time of ingest, resulting in the display of the file's MIME type in the repository's database and interface. File formats are identified using an internal service that identifies tabular data files and the [JHOVE](#) tool for file format validation. If these are unsuccessful, the software will attempt to identify the file format based on file extension. JHOVE outputs for the well-formedness and validity of a particular file are not recorded.

Additional Preservation Activities

Additional preservation activities may be conducted by institutional administrators: exporting BagIt-formatted packages and integrating Dataverse with Archivematica.

Scholars Portal can provide technical support to institutions interested in additional processing and export of datasets in their institutional collection. The Borealis team is unable to provide guidance for the appraisal and selection of datasets from your institutional collection that should be eligible for additional preservation activities.

For all selected datasets, verify that the user has obtained the necessary rights and privileges to process and store dataset files independently from Borealis.

BagIt Export

Datasets from your institutional collection may be exported as [BagIt](#)-formatted packages, often called Bags. Each package contains the user-uploaded files (except in the case of tabular data uploads, where only the converted .tab version is retained in the bag), and metadata in the form of a JSON-LD serialized OAI-ORE map file and DataCite XML file. BagIt packages produced by Borealis conform with the Research Data Alliance [BagIt profile](#).

You may choose to deposit your Bags to a file system location or space in DuraCloud. Datasets are exported using an [API call](#) and may be conducted by you or your team. If you would like assistance exporting BagIt packages, [contact the Borealis team](#) to coordinate the export.

To learn more about the benefits and challenges of preserving datasets from your institutional collection using BagIt exports, follow the [community discussion](#) or [contact us](#).

Integrating Dataverse and Archivematica

[Archivematica](#) is an open source, standards-based processing tool for creating well-formed packages for preservation storage. The [Archivematica integration](#) in the Dataverse platform allows you to select and process datasets from Borealis for long-term preservation and management. Only the most current version of a dataset will be processed and stored by the Archivematica workflow.

This integration creates independent preservation packages of selected datasets, with additional workflow functions such as signature-based file format identification, file format validation, characterization, and normalization to verify its authenticity and fixity.

There are a few options, if your institution is interested in establishing connections between your institutional collection and an Archivematica processing workflow.

- Scholars Portal offers the [Permafrost](#) service which offers a suite of tools, training, and resources for Ontario Council of University Libraries (OCUL) to process digital objects for long term preservation and access using the preservation storage in the [Ontario Library Research Cloud](#).
- The Council of Prairie and Pacific University Libraries (COPPUL) offers [Archivematica-as-a-service](#) in collaboration with Artefactual Systems and the University of British Columbia (UBC), using UBC's EduCloud Server Service.

See the Borealis [Preservation Plan](#) for more information about these activities and [contact the Borealis team](#) if you are interested in setting up these services.

CoreTrustSeal Certification

[CoreTrustSeal](#) is an international, community-based, non-governmental, and non-profit organization promoting sustainable and trustworthy data infrastructures. A CoreTrustSeal Certified Repository is a repository service that has been recognized as sustainable, transparent, and trustworthy. Requirements for certification are based on the [OAIS (Open Archival Information System) Reference Model](#) for preserving and making available digital information.

The [certification process](#) involves a self-assessment and external review of compliance with and documentation for a set of defined best practices for processes, policies, and infrastructure.

Applying for CTS Certification

You may wish to apply to certify your individual institution's repository instance or a collection within it. Certification demonstrates your institution's commitment to the reliability and durability of research data in your repository, and the transparency of policies and procedures.

Whether or not your team submits an application, the required self-assessment is useful for identifying areas for improvement and provides a useful benchmark on which to compare yourself to other repositories and to your repository in the future. As funding agencies have already started to require research data to be published in digital repositories, certification prepares your repository to meet additional criteria about such repositories that are introduced in future.

Features and integrations for the Dataverse software offer many key technical supports for meeting CTS requirements, however the software is only a small part of the certification process. The larger focus is on developing and documenting processes, policies, and curatorial expertise.

The Borealis team is able to provide in-depth assistance with the CTS application. This includes developing and documenting processes, policies, and infrastructure that would be required for a CTS application from member institutions.

To help you meet the evidentiary requirements outlined by the CTS application, the following policies and procedures governing Borealis are available:
- [About Us](#) page: describes the service, governance, designated community, and resources/support available.

- Preservation Plan: outlines the objectives, roles and responsibilities, strategies, and actions for preserving the digital files uploaded by users and stored in the repository.
- Privacy Statement: explains what data we collect through the use of Borealis, how this information is treated, and for what purpose.
- Technology Infrastructure and Security Information: outlines general information, technical infrastructure, application security, and storage and backup details.
- Terms of Use: governs all use of the Borealis service.

Dataverse API

Using the API

How you use an API can depend on your operating system, permissions on your local machine, and other tools you are familiar with. There are many ways that you can work with APIs, including using your built-in command line interface, with programming languages like Python, or through applications for working with APIs like Postman.

Most of the examples in the official documentation use a tool called **curl**, which can be used in your machine's command-line environment (shell). To use curl:

- In Windows, you can download and install Git for Windows, which contains a tool called GitBash.
- In Windows 10, you may be able to enable the Windows Subsystem for Linux.
- On Mac/Linux, the default "Terminal" application can be used.

See the Library Carpentry Unix Shell Setup page for some additional setup steps and tips.

To learn more about working with the command line, see the Programming Historian "Introduction to the Bash Command Line" lesson.

Once you have an environment set up, a quick way to test if it is working correctly is with an API command like the following, which retrieves information about a collection (in this case, the root collection on the demo site):

```
curl https://demo.borealisdata.ca//api/dataverses/dv
```

After typing this command and pressing enter, you should see JSON-formatted data with details about the collection, like the description and the date it was created.

The following guidance provides step-by-step walkthroughs for common API commands. For comprehensive documentation and other commands, review the API Guide.

Testing the API

It's generally a good idea to test out the API commands you're using first, especially if you are new to using APIs or are not sure how a command works. To get started, we suggest you try out your API commands in the demo environment first, especially if those commands can't be reversed (e.g. deleting content). An API command run on the command line will not usually present you with an option to "confirm" your action, so we **strongly recommend** testing it first!

Getting Your API Token

To complete many tasks with the Dataverse API, you will need to have your account's API token. To retrieve your API token in Borealis:

1. Log into Borealis
2. Click your username in the upper right corner
3. Select "API Token"
4. Copy the API Token on this page, or click "Create Token" if one has not already been generated and copy it.

API tokens should always be stored in a secure way, like a password, as they give complete access to all of the data in your account.

Viewing JSON in Your Browser

Many API commands involve viewing JSON-formatted information directly in your web browser. While some browsers will format this in a readable way by default, some will not. If you find the JSON hard to read, consider installing a JSON viewing browser extension (e.g. JSON Formatter for Chrome).

SuperUser Commands

Some API commands can only be run by a super user. This means that this command has to be run by someone from the Borealis team. If you have a command like this that you want to run, contact us.

Finding a Dataset ID

For API commands related to datasets, you may need to get the ID of the dataset, which is different from the dataset's persistent identifier. To find the ID of a dataset:

1. Find the permanent identifier (DOI or Handle) for your dataset
2. Use the following API command directly in your browser, replacing the BOREALISURL and DATASETIDENTIFIER with your own:
   https://BOREALISURL/api/datasets/:persistentId/?persistentId=DATASETIDENTIFIER
   1. For example, for this dataset, we would use the following:
      https://borealisdata.ca/api/datasets/:persistentId/?persistentId=doi:10.23685/1H9TOV
3. The dataset ID is the first ID in the JSON output.
   1. In the JSON that is shown, we see the data ID for this dataset is 97164.

Finding a Dataverse ID

For API commands related to dataverse collections, you may need the alias of the collection. This is the identifier that is in the URL of your collection.

For example, for the UBC Dataverse, which is available at https://borealisdata.ca/dataverse/ubc, the ID of the collection would be "ubc".

Some Common & Useful API Commands

Find the date a collection was created

To get more details about a specific collection, such as the date it was created, you can enter the following API call as a URL in a web browser, replacing the $ALIAS in the URL with the alias of the collection:

https://borealisdata.ca/api/dataverses/$ALIAS

For example, if we wanted to find more information about the "ottawa" collection, we would use the following URL:

https://borealisdata.ca/api/dataverses/ottawa

If the collection has not been published, you will need to add your API token to the URL:

https://borealisdata.ca/api/dataverses/$ALIAS?key=$API_TOKEN

Change a citation date

1. Find the dataset id for the dataset with the citation date you want to update using the persistent identifier from your dataset (DOI or Handle).
2. Ensure that the date you want to change the citation date to is reflected in your metadata in a date field, e.g. in the deposit date field.
3. Find the name of the field you want to use as the citation date (e.g., for the deposit date you would use the field name of "dateOfDeposit," for production date, use "productionDate"). In the following example, the field is called 'dateOfDeposit' under the metadata blocks:
   - https://demo.borealisdata.ca/api/datasets/:persistentId/?persistentId=doi:10.5072/FK2/QLQNAP
4. Enter the following command, replacing the dataset ID, API key, and name of the metadata field being used to replace the citation date:

   ```
   curl -X PUT
   ```

```
"https://borealisdata.ca/api/datasets/$datasetID/citationdate?key=$a
piKey" --data "dateOfDeposit"
```

Get the size of a collection

To get the size of a collection which you have admin access to, paste the following command in your web browser, replacing the API Token and the collection ID (i.e., the DATAVERSE_ID) that is found in the URL of a collection:

https://borealisdata.ca/api/dataverses/DATAVERSE_ID/storagesize?key=API_KEY

To get this through the command line, use the following command:

```
curl -H "X-Dataverse-key: API_TOKEN"
https://borealisdata.ca/api/dataverses/DATAVERSE_ALIAS/storagesize
```

- For example: curl -H "X-Dataverse-key: API_TOKEN"
  https://borealisdata.ca/api/dataverses/queens/storagesize

This API call will return a total size in bytes. You can convert it to MB/GB using an online tool like this one.

Bulk Uploading

When used in a script, API commands can be combined or used within loop structures to bulk upload data, such as multiple files within a folder. One example of this can be seen in this script, which can be used to create multiple datasets at a time within a collection.

Bulk Downloading a Larger Dataset

If you are downloading a dataset which has 5GB of data or more, you will not be able to download all of the files at once through the user interface. One way around this is to use a tool called wget in your command-line in order to download all of the files in the dataset to a local folder on your machine. In some Windows environments, you may need to install wget before being able to use this command.

First, you will want to create a new folder for the dataset you're downloading, and navigate to this directory in your terminal.

If the dataset has any restricted files, you will need to retrieve your API token in order to download them (see more details below). If no API token is used, then this command will only download the files you have access to.

You will also need the identifier (DOI or handle) for the dataset you want to download. This should be formatted as either (for example) "doi:10.5683/SP3/OHVUDH" or "hdl:10864/10120".

Depending on how large the dataset is, this may take a while to run. [More information about this command is available in the documentation.](#)

To download all of the files in a dataset, use the following command, replacing the IDENTIFIER with your own:

```
wget -r -e robots=off -nH --cut-dirs=3 --content-disposition
"https://borealisdata.ca/api/datasets/:persistentId/dirindex?persistentId=
IDENTIFIER"
```

Download restricted data

If the dataset has any restricted files, you will need to be given access to the files and retrieve your API token in order to download them. To add your API key to the command, use the following formatting, replacing the IDENTIFIER and API-KEY with your own:

```
wget -r -e robots=off -nH --cut-dirs=3 --header "X-Dataverse-key: API-KEY"
--content-disposition
"https://borealisdata.ca/api/datasets/:persistentId/dirindex?persistentId=
IDENTIFIER"
```

Download a folder

If the dataset has a file structure in place, this command can be adjusted to download a specific folder. To download a folder, add the "folder" parameter to the URL and replace the IDENTIFIER and FOLDER-NAME with your own:

```
wget -r -e robots=off -nH --cut-dirs=3 --content-disposition
"https://borealisdata.ca/api/datasets/:persistentId/dirindex?persistentId=
IDENTIFIER&folder=FOLDER-NAME"
```

If the folders are nested, use the complete path to the folder, e.g. "FOLDER1/FOLDER2/FOLDER3".

Download a version

If the dataset has more than one version, and you don't want to download the latest published version, you can add the version to your URL, such as "1.0". To download a version, add the "version" parameter to the URL and replace the IDENTIFIER and VERSION-NUM with your own:

```
wget -r -e robots=off -nH --cut-dirs=3 --content-disposition
"https://borealisdata.ca/api/datasets/:persistentId/dirindex?persistentId=
IDENTIFIER&version=VERSION-NUM"
```

Data Migration

You can use the Dataverse API to import datasets that do not have existing DOIs from your institutional holdings into Borealis. Use the following script to upload these datasets as unpublished items in Borealis.

Borealis accepts metadata for ingest in the following standards:
- Dublin Core
- DDI (Data Documentation Initiative Codebook 2.5)
- DDI HTML Codebook (A more human-readable, HTML version of the DDI Codebook 2.5 metadata export)
- DataCite 4
- JSON (native Dataverse Software format)
- OAI_ORE
- OpenAIRE
- Schema.org JSON-LD

Dataverse JSON is the preferred format for metadata uploaded for migration. A full list of Borealis fields in JSON format and an example record in JSON format for migration are available for your reference. This list also maps all the fields to elements in various standards, which may be useful for creating the JSON.

Once your metadata and all the files (or DOIs of the datasets to be updated) are ready, you can use the APIs to either batch upload datasets or replace metadata.

We recommend that institutional administrators provide support for metadata formatting and ingest processes to streamline the migration.

Scholars Portal can provide limited support to institutions interested in migrating existing datasets into Borealis. Scholars Portal does not currently support file-level DOIs due to additional work and ongoing costs and maintenance required.

If you would like support, contact the Borealis team with the following information:

- The number of datasets being migrated
- The total size of the data being migrated

- ○ Indicate if any individual files are larger than 3GB
- Whether the data have DOIs that need to be imported
    - ○ The Borealis team must run imports for datasets with DOIs as superuser access is required. DOIs will also be transferred to the Scholars Portal Datacite account, and future DOIs for these datasets will be minted with the Borealis prefix.
- Where in Borealis the datasets are being migrated to e.g., institutional collection, sponsored collection, etc.
- Where datasets are being migrated from e.g., an existing Dataverse installation, a different repository, etc.
- The metadata formats currently in use for the datasets and what metadata can be exported from their current location

Ongoing Migration Projects

The following migration projects are under investigation or underway. If you would like to know more about these projects or work with us on a new project, contact us.

- ODESI
- DLI training repository
- CUDO (Carleton University Digital Objects)
- Nordicana (Universite Laval)

Resources

Please refer to the [Support & Additional Resources](#) page in the User Guide for links to a variety of useful documents and websites related to Dataverse.

Resources from the Alliance

- [Dataverse North Expert Group](#) -- a community of practice for libraries using or interested in using the Dataverse repository platform in Canada.
- [Institutional Research Data Management (RDM) Strategies](#) -- strategies and advice on developing research data management policies at research institutions.
- [Appraisal Guidance for the Preservation of Research Data](#) -- guidance for evaluating research data for long-term preservation in repositories.
- [Curation Guide](#) -- step-by-step instructions for curating new datasets deposited in the Dataverse platform.
- [Institutional Collection Policy Templates](#) -- templates to develop a local policy framework for selecting, depositing, curating, and deaccessioning data sets.

Resources from the Dataverse Project

- [Community Meetings](#) -- annual community meetings for sharing solutions and standards for research data management using the Dataverse software.
- [Community Norms](#) -- guidance and best practices for curating, sharing, and preserving data in the Dataverse software.
- [Dataverse Software Guide for CTS Certification](#) -- explanations and examples of how the Dataverse software and community speak to various CTS requirements.

References

- [Dataverse Project User Guide](#)
- [DataverseNO Curator Guide](#)
- [Texas Data Repository User Documentation](#)

Appendix

Example custom licences and agreements for a dataset's Terms of Use or a file's Terms of Access sections:

- [Sample Data Usage Agreement](#) for datasets containing de-identified human subject data, produced by the Dataverse Project
- [Australian Social Science Data Archive License](#) for use for scientific purpose only for datasets containing identifiable or quasi-identifiable data that may 1) only be reused for academic research purposes, 2) involve a request submission and review process, and 3) provide different licensing for use of data and documentation, developed by the Austrian Social Science Data Archive