1. Borealis Institutional Collection Curation Framework Template

Rev. 2024-12-20
Authors: Michael Steeleworthy and Erin Clary. Policy Task Group, Dataverse North Expert Group
Reviewers: Y.G. Rancourt and Robyn Stobbs. Preservation Expert Group, Digital Research Alliance of Canada

This Institutional Collection Curation Framework is intended to assist Borealis administrators and their sponsoring institutions (e.g., academic libraries) as they develop or refine their data curation service. It introduces different elements of curation (e.g., file formats, metadata, data quality) and asks the reader to consider what effect they may have on their curation practice. Many aspects of this framework will intersect with, or be dependent on, your institution's existing policies, practices, resources, and stakeholders. These should all be reviewed as you evaluate what criteria are in scope for your curation service, and you may need to consider whether additional commitments or operational adjustments will be required to provide your preferred level of service.

For considerations about the deposit and appraisal process, please see the Borealis Institutional Collection Deposit & Appraisal Guidelines.

Documentation based on this template may support your responses to the following 2023-2025 CTS requirements in the Application Template: R08 Deposit & Appraisal, R10 Quality Assurance, and R13 Reuse.

1.1. Important Reminder

All of the institutional policies and procedures for your institutional collection in Borealis should work together in order to make managing your collection easier. Ensuring all of your policies and procedures work together is an important step in their development. Please see this chart for a visual representation of how institutional Borealis policies and procedures can flow together to cover all the required/needed elements.

1.2. Introduction: Curation as a Practice

Data curation is the active management of research data as it is created, maintained, used, archived, shared, and reused[1]. It is a set of processes - both automated and mediated by professionals - that can add value, improve a dataset's FAIRness[2] (its findability, interoperability, accessibility and reusability) and prepare it for long-term preservation.

The policies and strategic goals that govern your institutional collection in Borealis are often manifested in your data curation service. How your curation service operates, e.g., the amount of service you provide and the scope of the curation activities you perform, will be dependent on the mission and resourcing capabilities of your organization, and the functionalities of the platform on which the repository is housed. Libraries must determine their reasons for curating data, their capacity to implement a curation service, and what elements of curation will best suit their operating parameters.

This framework introduces a number of curation elements a library may consider to include in its curation practice[3]. These are supported by the concept of curation service scenarios (e.g., unmediated, semi-mediated, and mediated) and the assignment of curation tasks to one of three levels, as summarized in the *Dataverse Curation Guide* *(DVCG)*[4]. The *Institutional Collection Curation Framework Template* uses the DVCG's levels of curation below. Its service scenarios also follow the DVCG's scenarios, but with minor adaptations to their descriptions to accommodate data curation activities with little or no communication from the data originator (data rescues, open data). These service scenarios and curation levels may help your library determine its service model as it defines and benchmarks the steps in its curation practice:

Curation Service Scenarios

Unmediated Curation	There is no intervention from the RDM service. The researcher creates their own collection and datasets in Borealis, uploads their data files,

[1] Portage Network Curation Expert Group. (2020). Primer - Curation. Zenodo. https://doi.org/10.5281/zenodo.4001004.
[2] Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18.
[3] For more information on data curation activities, please see Johnston, L. R., Carlson, J., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R., & Stewart, C. (2016). *Definitions of Data Curation Activities used by the Data Curation Network*. http://conservancy.umn.edu/handle/11299/188638.
[4] Cooper, A., Steeleworthy, M., Paquette-Bigras, È., Clary, E., MacPherson, E., Gillis, L., Wilson, L., & Brodeur, J. (2021). *Dataverse Curation Guide*. https://doi.org/10.5281/zenodo.5579820. Et en français https://zenodo.org/record/5579827.

	adds metadata, and publishes the collection/datasets. The RDM service relies on Borealis' automated curation activities and may provide training or documentation to support deposits.
Semi-Mediated Curation	Researchers are able to create datasets through automatic permissions based on affiliations or through manually added permissions from administrators. After uploading files and adding metadata, the researcher submits datasets for review by admins from the RDM service. Admins receive notifications when datasets are submitted for review, and can either return the dataset to authors with requested changes or publish the dataset.
Mediated Curation	The RDM service creates the collection (or dataset) in Borealis and the data is curated by the library and published once approved by the researcher, or in some cases (e.g., data rescues, open data curation), by the RDM service itself.

Levels of Curation

Level 1	The minimum steps required to successfully publish in Borealis and make the dataset findable, e.g., the dataset has been submitted to the proper collection and required metadata fields are accurate.
Level 2	Activities that enhance the discoverability of datasets and help ensure their usability over time. e.g., recommended metadata fields are populated and the dataset includes sufficient documentation to allow a user with a similar background to understand the dataset and open and use the files.
Level 3	Intensive curation actions intended to prepare datasets for preservation and improve the chances that data and code can be used to reproduce or replicate an associated study. For example, supporting documentation is enhanced, the content of files and code are reviewed, and data files are transformed into formats suitable for long-term preservation.

1.3. Curation Framework Criteria for your Library

The curation elements presented below may be relevant to your library as it develops or refines its curation practice. Note that many of these questions will require you to consider your repository's terms of use, collection policy, depositor agreements, and preservation policy.

Curation activities should be transparent, and should take place with input from the depositor. If certain activities will be done automatically (e.g., file format transformation), consider stating that explicitly in public facing documentation.

1.3.1. Relevance of the dataset to the collection

"How will you appraise a dataset's fit with your collection?"

To determine the relevance of a dataset to your data repository, you must consider your collection policy and procedures. These may include policies for research data, general collections, archives, or scoping criteria for complementary repositories in your organization, as well as available services and resources external to your organization. Ideally, the scope and breadth of the data repository's holdings will be determined prior to initiating the repository service in order to develop consistent practices and manageable data collections.

As with many collection policies, prioritization enables the library to focus on particular themes or align its collections with a broader service mandate. Your prioritization criteria may emphasize datasets associated with a publication, original datasets of intrinsic value, or data associated with a particular research area. Relevance may also be determined by other factors such as the dataset's association to the institution, grant status, or historical value.

The future value of a dataset is difficult to predict. It is linked not only to the data's associated research and reproducibility, but also the technical and organizational systems used to store, provide access to, and preserve the data[5]. Therefore, it may be useful to consider the present research landscape in decision-making: How valuable are these data in communicating current knowledge and perspectives? Are there external metrics (e.g., citations, grants, awards, access statistics) or disciplinary norms (e.g., retention periods) available that can support determinations of value and their effect on the data's relevance to your collection? Be particularly mindful of data that assists in representing research about or by under-documented or marginalized peoples, or which documents any shifts in a field of practice.

These priority areas and deposit guidelines should be disclosed in your library's collection policy, data collection policy, terms of use, preservation policy, or other strategic documents and published in a place accessible to potential depositors.

Please refer to our [Institutional Borealis Collection Policy Template]() for more information.

1.3.2. Metadata and Documentation

"To what extent will metadata and documentation be curated?:

Metadata enrichment is perhaps the most evident and understood curation activity. Borealis provides robust metadata enrichment functionalities, much of which is described in the *Dataverse North Metadata Best Practices Guide*[6], the previously mentioned *Dataverse Curation Guide*, and Borealis User Guide[7]. Your Library must determine its minimum standards and best practices for enriching metadata and documentation. These decisions will depend partly on operational considerations such as resourcing capabilities, your curation service model (e.g., fully mediated vs. partially mediated), and the depth or level of metadata and documentation enrichment. For example, one library may limit its metadata

---

[5] Lavoie, B. (2012). Sustainable research data. In G. Pryor (Ed.), *Managing Research Data* (pp. 67-82). Facet. doi:10.29085/9781856048910.005

[6] Bascik, T., Boisvert, P., Cooper, A., Gagnon, M., Goodwin, M., Huck, J., Leahey, A., Stathis, K., & Steeleworthy, M. (2021). *Dataverse North Metadata Best Practices Guide v 3.0*. https://doi.org/10.5281/zenodo.5576411. Et en français https://doi.org/10.5281/zenodo.5576430.

[7] Scholars Portal. (2022). *Borealis User Guide. https://learn.scholarsportal.info/all-guides/borealis/*

curation activities to improving citation-related metadata, and another may benchmark its practice to the enhancement of metadata elements that improve discoverability and reproducibility by requiring keywords, a dataset description, open documentation, variable definitions, etc.

To maintain consistency in practice, it is recommended that libraries develop local best practices to guide their metadata work. The Dataverse North Metadata Best Practices Guide can be used as a starting point as it provides information about Borealis' required fields and guidance on its metadata conventions.

To determine how metadata and documentation enrichment may be incorporated into its curation practice, libraries might consider 1 of 3 levels:

1. Metadata and documentation are not curated. Depositors are responsible for providing complete, accurate information that describes files, their content, and use.
2. At the discretion of the data repository, metadata and documentation may be curated for completeness and accuracy. Enrichment may be benchmarked to the repository's required level of curation (e.g., Dataset is publishable, Dataset's discoverability and usability is enhanced; Dataset's preservation capabilities are enhanced).
3. Metadata and documentation will be enriched to meet the repository's required level of curation. The repository will not publish data until all required metadata fields are completed and key documentation is provided.

Criteria may also be defined to apply different levels to different sub-collections or datasets.

1.3.3. Data Integrity and the Chain of Custody

"To what extent is the chain of custody important to your library?"

Borealis automatically records provenance metadata related to the ingest process, including the name and email of the person who submitted the data and the date of submission, but your library will need to determine the importance of a dataset's chain of custody prior to deposit, as well as the integrity of the files. Recording a dataset's chain of custody can help ensure that changes are authorized and justified and support transparency and accountability.

During the data deposit process for a mediated curation service, for instance, this may include developing preferred or required methods for data transfer to the repository to mitigate concerns about dataset integrity, completeness, and provenance. Examples include requiring a depositor to upload directly into the Borealis system, mediated transfer to repository staff through cloud-based folders, secure file transfer systems, or even email.

Mandating a specific transfer method can sometimes bring the benefit of best practices but it may not be feasible for datasets of every size, and it might come at the cost of user annoyance. For example, a file transfer convention such as BagIt would systematize the collection of critical metadata at the point of transfer but can potentially create a speed bump perceived as unnecessary by a depositor who is unaware of the importance of metadata and fixity to file transfer and curation.

Once data is received, curation activities by staff and depositors may also result in intentional changes to enhance discovery, access, and use of datasets. A particular point of focus may be documentation review to ensure relevant contextual information, such as software requirements, third-party data sources, licenses and agreements, and access restrictions are provided. Your library may maintain a curation log to record the nature and reason for changes made prior to publication. Your library may choose to record curation activities for individual files in Borealis through provenance files or metadata, or to maintain logs in an external system.

To determine how data integrity and chain of custody may be incorporated into its curation practice, libraries might consider 1 of 3 levels:

1. Data Integrity and Chain of Custody will not be assessed in the curation process. The repository service will curate files without questioning file fixity or provenance.
2. At the discretion of the data repository service, datasets may be assessed for integrity and chain of custody. If this assessment results in any changes to datasets, curation logs are created and retained.
3. All datasets must be deposited with the repository via the XYZ model of transfer. All datasets will be assessed for integrity and chain of custody. All assessments will be recorded in curation logs, whether or not changes are made.

1.3.4. Data Quality

“To what extent will data quality be assessed by your library?”

The extent to which a library will assess the quality of data it receives is dependent on its subject expertise in the research domain and its resource capacity. Often, but not always, institutional repositories will receive data that has already undergone rigorous cleaning, quality control, and analysis procedures. Institutional data repositories may collaborate with research groups to ensure that deposited datasets are of a high quality prior to the deposit of data with the repository itself, thereby linking data quality to the work of subject experts and saving time in the curation process. Such an operational decision would acknowledge the subject-related limits of the repository but also provide more time for curation practices related to access, discovery, or preservation.

To determine how data quality might be assessed during the curation process, libraries might consider 1 of 3 levels:

1. Data quality is not assessed during curation. The depositor is honour-bound to deposit data and documentation that can be understood by others in the field, and is sufficiently prepared for (re)use.
2. At the discretion of the repository service, data quality may be assessed during curation. A data quality assessment may occur depending on resource capability, value of the dataset, and subject expertise within the curation team. A data quality assessment may result in recommendations or requirements for data cleaning to complete the ingest process.
3. All data will be assessed during curation. A data quality assessment may result in recommendations or requirements for data cleaning to complete the ingest process.

1.3.5. File Formats

"What are your library's conventions for accepting file formats?"

The library must determine the extent to which it will allow, transform, or reject closed and proprietary file formats in the repository. The formats you agree to ingest will have both operational and strategic implications, especially if your repository is committed to long-term preservation for reuse and reproducibility. Accepting only open formats will facilitate preservation and improve the possibility that items in the repository's collection remain available for use in the long-term; however, this may increase the length of the curation process, add burden to curators and depositors, and reduce demand for repository services. Conversely, a repository service that allows depositors to share proprietary or closed file formats may reduce the burden on depositors or curators at the front end, but at the risk of stewarding data that becomes unusable over time. Some flexibility may be required, and understanding what formats are widely used in various disciplines, and whether freeware to extract, transform, or visualize data in that format is under active development, can help you evaluate whether there is likely to be continued support for a closed file type in the near term.

If you will transform files, you may need to work with the depositor at the point of ingest to minimize the risk that information is lost in the transformation process. This may also be necessary if you do not have access to a software that can be used to view and export content in an alternate format. Publishing a list of preferred formats is advisable as it acknowledges your library's commitment to the support of these formats while also signalling that other formats may not receive the same level of support. If you do not have a published list of preferred formats, the Library of Congress Sustainability of Digital Formats[8] chart is a good resource to track the effects of time and technological change on file format accessibility.

Libraries using Borealis should understand how it handles tabular data (e.g., spreadsheets, SPSS data files) and Microsoft Excel files when considering the impact of file formats on their curation practice. Upon upload, Borealis transforms tabular data, including XLS, SAV, CSV, and others into non-proprietary tabular text data files (TAB), and creates citation-related metadata and DDI variable-level metadata. This tabular ingest process, while providing a preservation-friendly action, also enables discovery and presentation capabilities with the Data Explorer, an integrated web application for exploration and curation of variables in Dataverse. Only the first tab or sheet within a spreadsheet file will be ingested and displayed in the TAB file. Therefore, multiple-sheet Microsoft Excel files and Excel files that contain formatting will lose data and context in this TAB transformation process. Borealis does maintain all original copies of tabular data within the system for access and download, but curators must be aware of this TAB transformation process and the resulting appearance of a TAB file in the dataset.

To determine the impact that file formats may have on the curation process, libraries might consider 1 of 3 levels:

---

[8] Library of Congress. (2004). Sustainability of Digital Formats: Planning for Library of Congress Collections. https://www.loc.gov/preservation/digital/formats/.

1. The repository service does not curate file formats, and may or may not provide guidance to depositors on preferred file formats. If guidance is provided, the depositor is expected to submit data in preferred formats, but deposits that do not adhere will still be ingested. No transformation to open formats will occur.
2. At the discretion of the repository service, proprietary, closed, or common file formats may be accepted. Format transformation may take place during curation, while maintaining original formats for future access.
3. The repository service accepts only open file formats or files that can be transformed to open formats. All files will be checked for openness during the curation process. Format transformation may take place during curation, while maintaining original formats for future access. Depositors may be asked to replace non-open files with open copies.

It is recommended that libraries consider the DCN Data Curation Primers to standardize file format handling during the curation process.

1.3.6. File Organization and Naming Conventions

"What are your Library's File/Folder Hierarchy and Naming Conventions?"

As with file formats, the library must determine if it has preferred or required file organization and file naming conventions, and weigh the benefits of standardizing this information against the risk of added burden to depositors and curators. Ideally, the storage of digital information should be approached in a structured, transparent, and predictable manner to ensure that it is easy to assess the completeness of the data, to restore it in cases of loss or error, and to promote ease of access by both humans and machines.

Consistency in your approach to folder structure and naming conventions is all that is required to achieve some level of control and transparency, but there are also specifications such as the Oxford Common File Layout (OCFL) that support software-independent access to file storage.

File organization and naming conventions can be dependent on the repository's conventions, the depositor's preferences, associated code that calls on a static file, or a publication that already cites a file name. To determine the impact that file organization and naming conventions can have on the curation process, libraries might consider 1 of 3 levels:

1. The repository service will not consider file organization or naming conventions in the curation process.
2. At the discretion of the repository service, file organization and naming conventions may be considered within the curation process; changes may be recommended or required prior to the completion of the ingest.
3. All datasets' file organization and naming conventions will be included in the curation process. Changes may be recommended or required prior to the completion of the ingest.

Libraries should consider publishing public facing guidance to alert depositors of their expectations, and how these expectations will be enforced (e.g., uOttawa guidance: File naming and organization of data / Nommage de fichiers et gestion de versions).

1.3.7. Data Sensitivity, Copyright, and Terms of Use (Risk and Rights Management)

"How will your library manage sensitive data, copyright, and related issues?"

Your library's curation activities for sensitive data should be limited to Borealis' ability to securely handle such data. Your repository service and its curators should ensure that depositors do not upload data that contain identifiable human subject information or information that is otherwise deemed sensitive or confidential. Similarly, a repository service must ensure that data subject to copyright, content that belongs to a third party, or data that might be in contravention to the Borealis Terms of Use are not uploaded, or have procedures in place to mitigate and resolve such issues.

Prior to the inception of its repository service, the library should determine where responsibility lies to ensure that published contents are neither sensitive nor in violation of the service provider's terms of use (e.g., Borealis Terms of Use) or local terms of use. Libraries may develop terms of use that place responsibility with the depositor at the time of upload, which may shorten the curation process, but which will require downstream remediation if a dataset is found to be sensitive or in violation of the terms of use. Alternatively, the library could develop a more involved curation process that builds in additional steps to ensure the depositor and dataset is in compliance with these terms. These steps may include:

1. Depositors should agree to your repository terms of use or depositor agreement before beginning a new submission.
2. At the discretion of the repository service, a curator may review documentation or inspect files to ensure contents do not violate the repository terms of use. This review may occur when the curator determines that a confirmation of the presence (or removal) of sensitive information or information in contravention of the terms of use is required (e.g., based on study title or description, the file format, or based on the discipline).
3. At the discretion of the library service, a curator may request and review an unsigned participant consent form or research agreement to confirm that data can be published and shared.
4. Published content found to be in violation of the repository terms of use will be subject to an established process to mediate takedown requests, restrictions, or withdrawals.

Libraries should consider the *Joint FORCE11 & COPE Research Data Publishing Ethics Working Group Recommendations*[9] for best practices on handling ethical cases related to the sharing and publication of research data.

1.3.8. Embargoed Data

"*To what extent will your library accept and manage embargoed data?"*

---

[9] Puebla, I., Lowenberg, D., & Force11 Research Data Publishing Ethics WG. (2021). *Joint FORCE11 & COPE Research Data Publishing Ethics Working Group Recommendations*. Zenodo. https://doi.org/10.5281/zenodo.5391293

Establishing guidance on how your repository handles embargoes will provide consistency to your curation practice and help manage depositor service expectations. While Borealis offers automated embargo functionality that reduces the curator's need to manage future release dates, your repository service must determine if embargoes will be part of its practice. A library may choose to model its data repository as a fully open collection, much like its general collection, or it may choose to allow time-limited embargoes to meet depositor or publisher requirements.

All versions of Borealis since v5.8 allow curators to create time-limited embargoes at the file level. Curators should understand that once a file marked for embargo is published, it is impossible to shorten or lift the embargo; all public access is delayed until the embargo ends.

To determine whether embargoes will be part of the curation process, libraries might consider 1 of 2 levels:

1. The repository service does not accept embargoes
2. At the discretion of the curator and in consultation with the depositor, an embargo period may be granted. This period will be determined by the repository to meet the depositor's needs while maintaining consistency in practice.

1.3.9. Traditional Knowledge and Data Collected with Indigenous Partners

"How will your curation practices respect Indigenous data sovereignty?"

A growing number of policy instruments affirm the moral rights of Indigenous peoples to data sovereignty, including data and knowledge creation, ownership, and stewardship, as well as the imperative that settler-researchers and their organizations acknowledge and respect these rights. Instruments such as the United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP)[10], the Truth and Reconciliation Commission's *Calls to Action*[11], Tri-Agency's Research Data Management Policy[12], the FNIGC's *First Nations Principles of OCAP®*[13] (Ownership, Control, Access, Possession) and the Global Indigenous Data Alliance's *CARE Principles*[14] (Collective Benefit, Ability to Control, Responsibility, Ethics), emphasize the importance of honouring the community's control over the collection, analysis, (re)use, storage, archiving, and access to their data.

Regardless of the kind of curation practice you provide, your repository service must be prepared to manage Indigenous research data in the curation process. In keeping with the Tri-Agency Research Data Management Policy, grant-eligible post-secondary institutions -

---

[10] UN General Assembly. (2007, October 2.) United Nations Declaration on the Rights of Indigenous *Peoples*. A/RES/61/295. https://undocs.org/A/RES/61/295
[11] Truth and Reconciliation Commission of Canada, 2012. (2015.) *Truth and Reconciliation Commission of Canada: Calls to Action*. https://publications.gc.ca/site/eng/9.801236/publication.html
[12] Government of Canada. (2021, March 15). *Tri-Agency Research Data Management Policy*. Retrieved March 2, 2022, from https://www.science.gc.ca/eic/site/063.nsf/eng/h_97610.html
[13] First Nations Information Governance Centre. (n.d.) *The First Nations Principles of OCAP®*. https://fnigc.ca/ocap-training/.
[14] Global Indigenous Data Alliance. (2019.) *CARE Principles of Indigenous Data Governance*. https://www.gida-global.org/care

and therefore their libraries - should recognize that "data created in the context of research by and with First Nations, Métis, and Inuit communities, collectives and organizations will be managed according to principles developed and approved by those communities, collectives and organizations, and in partnership with them"[15]. At a minimum, your service should inquire about the formal and informal agreements between researchers and Indigenous partners that govern data sharing, possession, access, and use, and you may choose to review any relevant documentation before the data are ingested. In some instances, these agreements may alter your curation process (e.g., how metadata is enriched, what documentation is required and where the data may be stored) and downstream access and use provisions.

It is recommended that data curators and repository service managers in Canada enrol in FNIGC's *First Nations Principles of OCAP* training course. Libraries and curators should also consult with their Research Offices and Indigenous Affairs Offices for advice and counsel, and operation in the spirit of reconciliation with their Indigenous research partners.

1.3.10. Software and Code

"To what extent will your library curate software and code associated with the data?"

The extent to which your library will curate software or code associated with the dataset is dependent on your curation team's subject expertise, its understanding of programmatic languages or application-specific syntax, and the extent to which reproducibility or reusability is the goal of the repository. Reproducibility refers to the reanalysis of data to confirm previous research[16]. While standards for reproducibility can shift between subject domains, it usually requires data, fully described code, and documentation on collection, instrumentation, and processing.

A repository that curates research data associated with a publication may enable verification of analyses cited in the article itself. On the other hand, a dataset that holds raw data and its processed counterpart, annotated code that cleaned and processed the raw data, as well as information on data collection and instrumentation moves closer to the benchmark of reproducibility.

Code that is required to run, clean, and process data for analysis should be deposited alongside a dataset to improve verification of data and methods. Software curation has its own complexities owing to its hardware and operating system dependencies. Be aware that software and code may refer to different objects in different fields, and that some fields or studies may not use software or code.

When considering the level to which software and code should be curated, a library might consider 1 of 4 levels:

1. The repository service accepts but does not curate software or code. Depositors are honour-bound to submit open-source software and code or provide documentation that extensively describes data collection and processing methods.

---

[15] https://www.science.gc.ca/eic/site/063.nsf/eng/h_97610.html
[16] Pasquetto, I. V., Randles, B. M., & Borgman, C. L. (2017). On the Reuse of Scientific Data. *Data Science Journal, 16*(8). http://doi.org/10.5334/dsj-2017-008

2. The repository service accepts open-source software and code. At the discretion of the service, software and code in certain datasets will be curated for accuracy and completeness. All software and code from certain fields may be curated, depending on the collection strengths and requirements of the repository service.
3. Any software or code deposited with the repository will be curated for accuracy and completeness.
4. Software and/or code must accompany all data deposited into the repository service. These objects will be curated for accuracy and completeness.

1.3.11. Licenses

"Which licenses will your repository recommend or require?"

The Library should consider which licenses it is willing and able to support. Will you recommend or require that depositors use a particular license? Will depositors be asked to select from a set list of options, or will you allow custom licenses or terms of use?

Curators should familiarize themselves with the preloaded set of Creative Commons licenses in Borealis as well as other licensing instruments (e.g., MIT, GNU GPL). Providing a small set of options may make it easier for curators to recommend a license, easier for depositors to choose a license, and easier for end users to navigate the terms of use for any dataset downloaded from your repository; it may also be easier to manage digital assets in the long-term. However, restricting the set of available licenses may force you to turn some datasets away. Any data provided to the depositor by a third-party source or data derived from previously published data may be subject to specific terms of use, and will require a license that respects those terms of use. Likewise, data that were collected with Indigenous partners or industry partners may be shareable, but only if specific terms or restrictions are applied to the dataset. Software presents another challenge since the Creative Commons licenses that many repositories recommend are not generally appropriate for code and software.[17]

In order to help a depositor select an appropriate license, the curator may need to review the data and documentation, the terms of use for any data or code that may have been derived from, or provided by, a third-party source, and informal and formal research and data sharing agreements. Similar to sensitive data, you will need to consider where responsibility lies to ensure that published contents are not in violation of the service provider's terms of use (i.e., Borealis Terms of Use) or your local terms of use. Libraries may develop terms of use that place responsibility with the depositor at the time of upload.

As you consider what types of license(s) you will allow, and what responsibility the curation service has for working with depositors to review content and select an appropriate license, the following may be helpful:

1. Depositors should agree to your repository terms of use or depositor agreement before beginning a new submission.

[17] Creative Commons recommends against the use of CC licenses for code and software. For more information, see Creative Commons. (n.d.). Frequently Asked Questions: Can I apply a Creative Commons license to software? Last modified November 22, 2021. https://creativecommons.org/faq/#can-i-apply-a-creative-commons-license-to-software.

2. At the discretion of the repository service, a curator may work with the depositor to select an appropriate license for their dataset, either from an existing set of licenses preloaded into the Borealis Repository, or from any number of publicly available licenses.
3. At the discretion of the repository service, curators may review documentation or inspect files to ensure the contents do not obviously violate intellectual property rights or data use agreements. Curators may also work with depositors to properly attribute any third-party sources, and/or to create a custom license that will respect the terms of use of various data sources.
4. Published content found to be in violation of the repository terms of use will be subject to an established process to mediate takedown requests, restrictions, or withdrawals.