Scholars Portal Dataverse upgrade to version 5.8

Article de blog français ici

Scholars Portal Dataverse upgraded to version 5.8 on January 31. This upgrade brings exciting new features, including:

- Anonymous review
- File-level embargoes
- Curation labels (now available!)
- Set language of metadata record (French support) - Currently disabled until further notice
- Improved identification of GeoJSON and Beyond 20/20 file formats
- New geospatial data previewer for GeoJSON files
- Improved notifications and messaging
- New APIs
- Bug Fixes

To test these features, feel free to use our sandbox environment.

Full details about the new features and changes can be found in the IQSS GitHub Release pages (English only):

- 5.8
- 5.7
- 5.6
- 5.5
- 5.4.1
- 5.4
- 5.3
- 5.2

Please reach out if you have any questions or feedback.

Anonymous review

What is it?

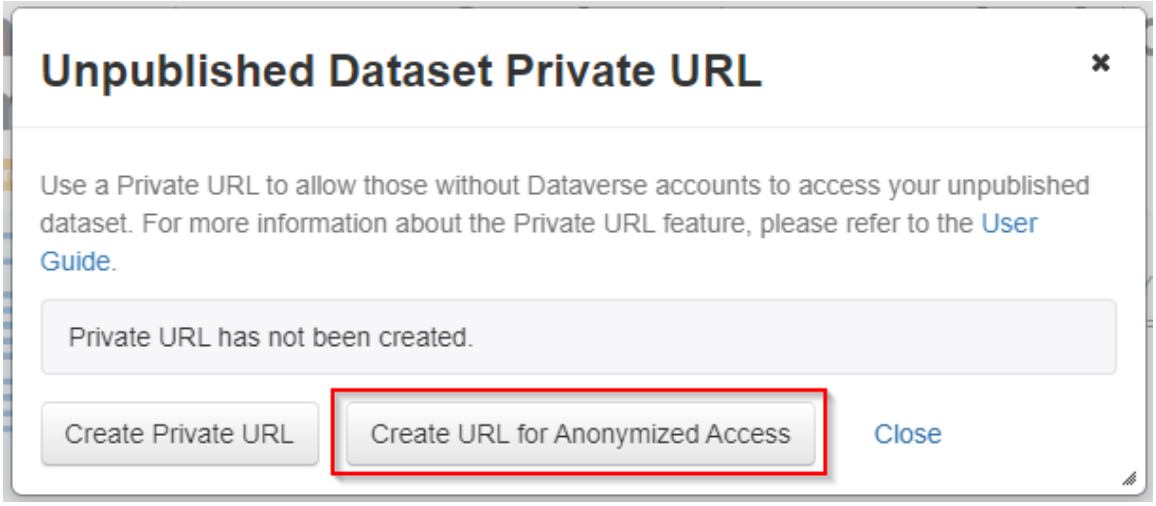
Private URLs provide access to unpublished datasets without the need to log into an account. Users can create a **URL for anonymized access**, which conceals metadata fields that contain identifying information about the author.

What is the use-case?

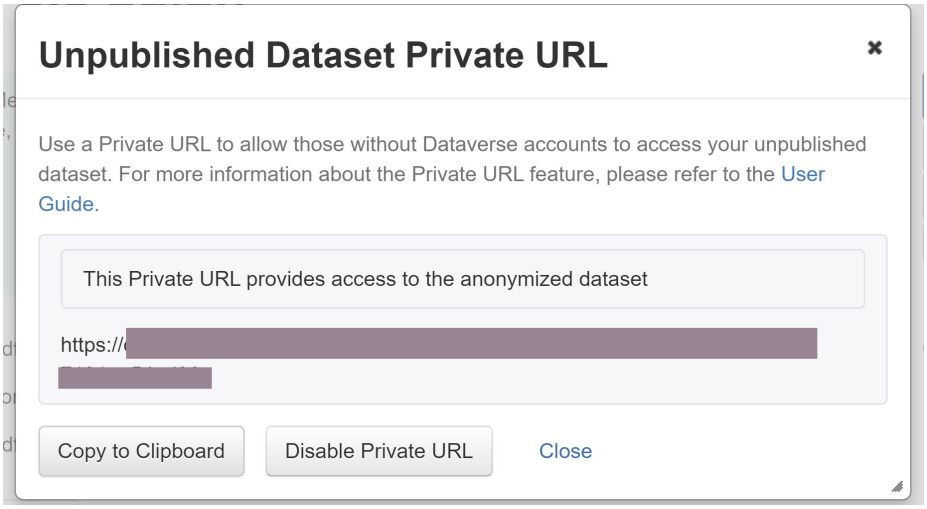
URLs for anonymized access have been a desired feature by the community to support the blind peer-review process.

How do I access this feature?

This feature is available to dataset admins 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.”

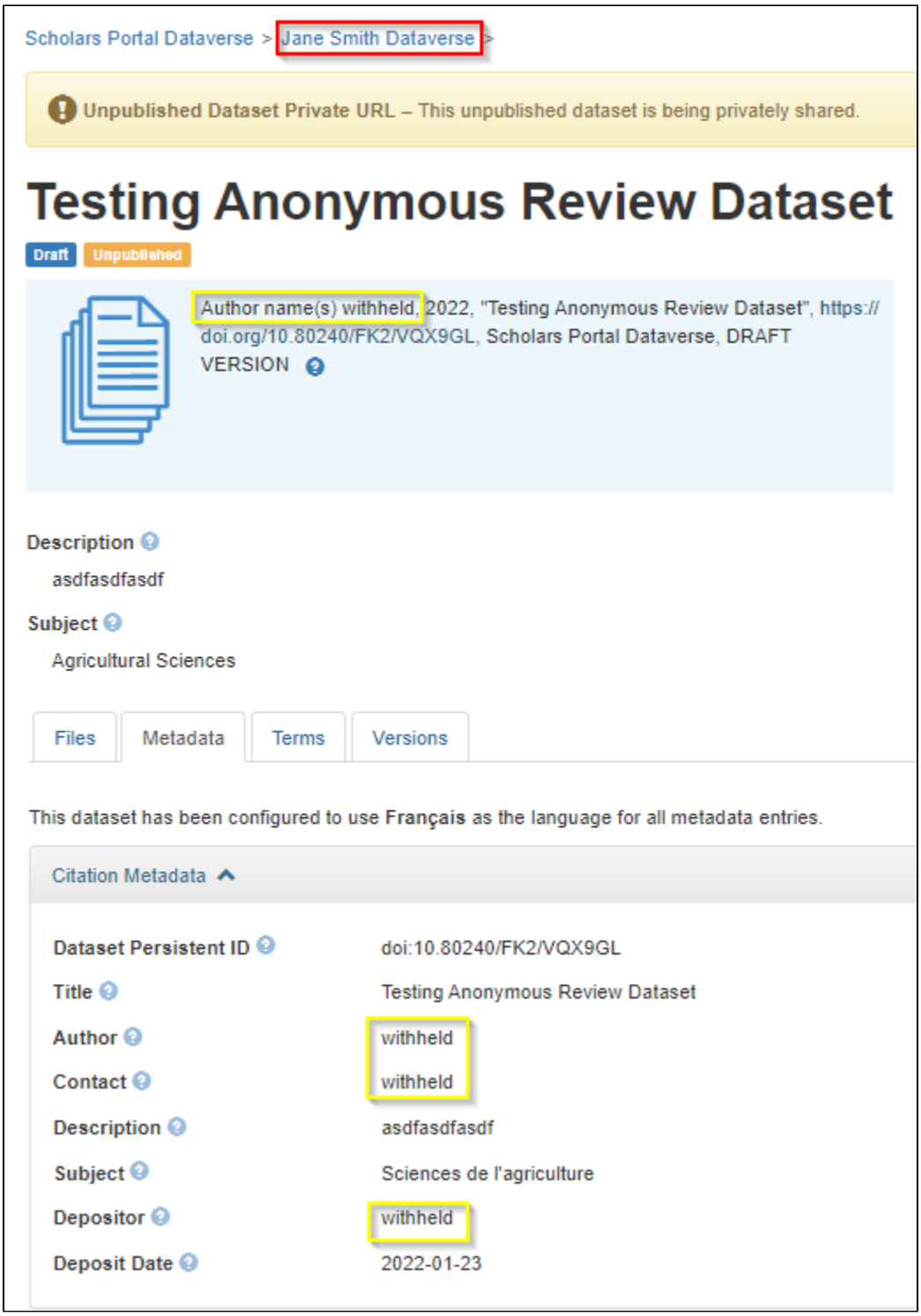


In the next window, click “Copy to Clipboard” to copy the URL to provide access to the anonymized dataset. Once the review process is complete, either select “Disable Private URL” or publish the dataset.



The anonymized dataset has fields withheld such as Author name, Depositor name, and Contact name (shown with yellow boxes). Other fields withheld include Producer, Production Place, and Distributor.

Note: Our testing revealed that the parent Dataverse collection(s) are shown in the dataset page breadcrumbs (shown with a red box at the top of the page), which might reveal some information about the researcher or research group.



For more details about anonymous review, see the Advanced User Guide (English only).

File-level embargoes

What is it?

Embargoes make files inaccessible after a dataset is published until the embargo period end date. The effect of adding an embargo is similar to making a file restricted, meaning that file previews and the ability to download is prevented until the embargo period is over. Requests for file access cannot be made during this time.

Users configure a specific embargo on an individual file 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.

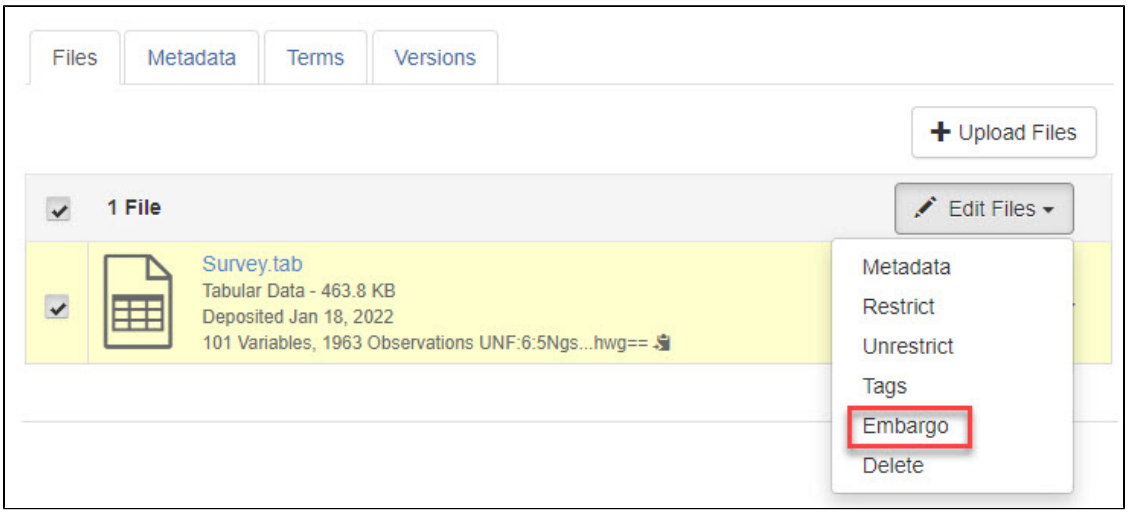
Embargoes can only be set, changed, or removed before a file has been published.

What is the use-case?

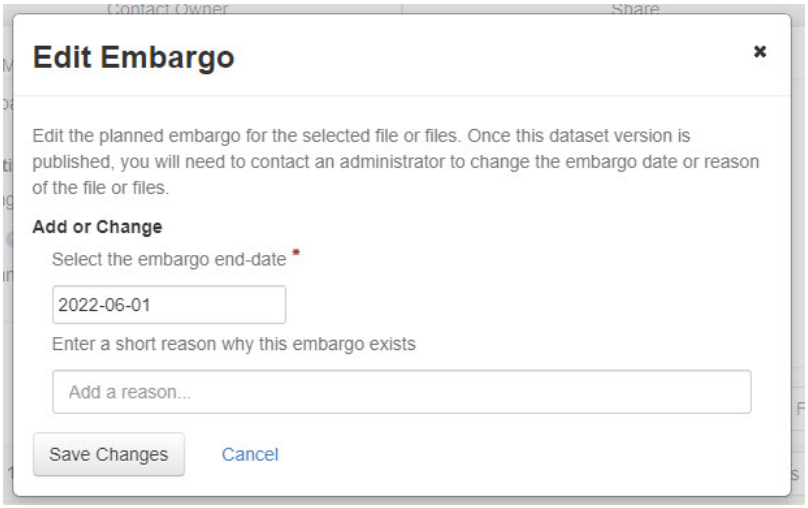
Embargoes allow users time to publish the dataset to make the metadata available, but the files are not accessible until a specified release date for reasons such as article publication plans or agreements with funders or industry partners.

How do I access this feature?

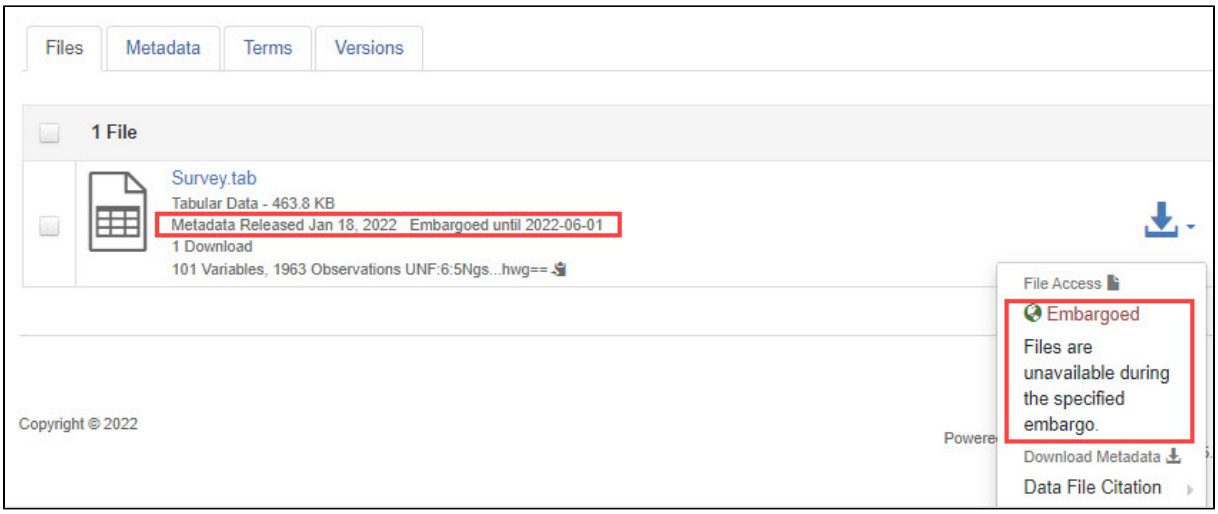
This feature is available to dataset admins, curators, and contributors. Select the file(s), and under the “Edit Files” menu, select “Embargo” (shown with a red box).



In the pop-up window, type in the end-date (YYYY-MM-DD) or use the calendar widget to select the end-date of the embargo. Enter in a short explanation for the reason for the embargo in the text box. Select “Save Changes” to set the embargo.



Once the dataset has been published, the embargo period is displayed in the file metadata and under the file access menu (shown with red boxes). The embargo cannot be changed after the dataset is published.



For more details about embargoes, see the Advanced User Guide (English only).

Curation labels (now available!)

What is it?

Curation labels can be applied to a draft dataset version to indicate its status as part of a curation process. Currently, the standard set has been enabled for institutional Dataverse collections, which includes the following labels:

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

Labels are visible on the Dataset page and in Dataverse collection listings/search results to dataset admins and curators. Internally, the labels have no effect, and once the dataset is published any existing label will be removed. A reporting API call can be used by admins to obtain a list of datasets and their curation statuses.

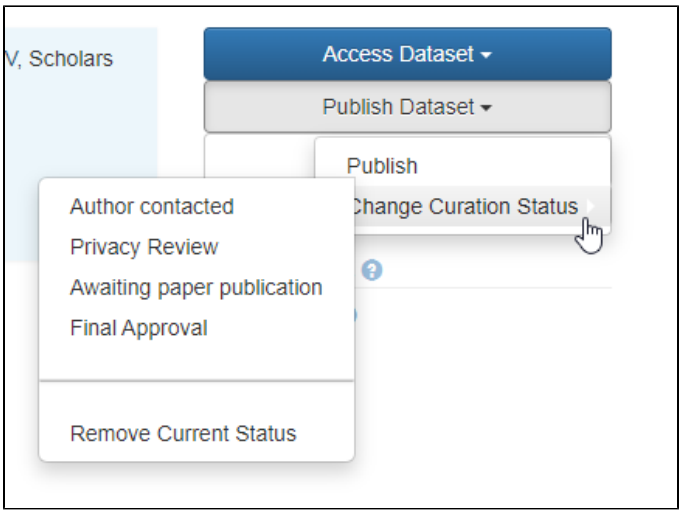
The Scholars Portal team can work with institutional Dataverse collection admins to enable the curation label list for their Dataverse collection.

What is the use-case?

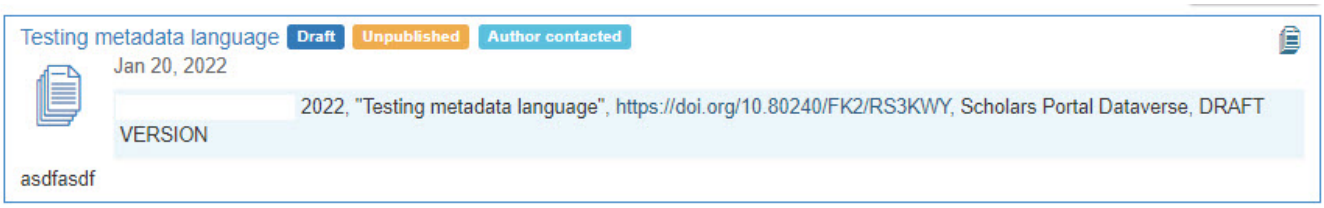
The labels can be used by Dataverse collection administrators as part of their curation process in order to more easily identify the status of datasets in their collection(s).

How do I access this feature?

This feature is available to dataset admins and curators (those who are able to publish datasets). Under the “Publish Dataset” menu, go to “Change Curation Status” and then select the desired label or remove the current status.



The labels are viewable from the Dataset collection page as coloured tags.

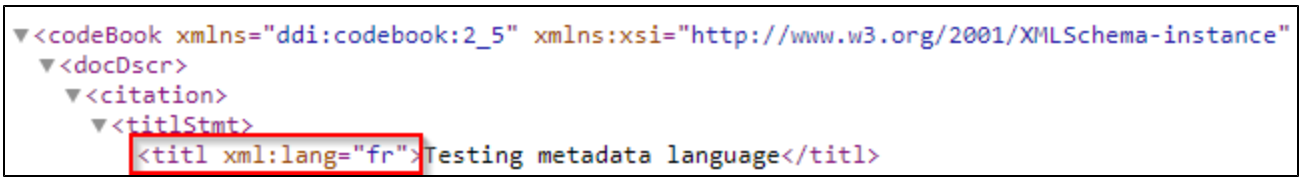


Set language of metadata record (French support) - Currently disabled until further notice

What is it?

Dataverse collection admins can specify at the Dataverse collection level which language will be used when entering metadata for new datasets. Once the language has been set, a message appears at the top of the metadata editing page (e.g., “Please enter metadata in the Français language, or change the metadata language choice in the parent dataverse before proceeding”). A message appears at the top of a metadata record (e.g., “This dataset has been configured to use Français as the language for all metadata entries") as shown in the screenshot below.

Additionally, the specified language appears in the DDI Metadata Export (as shown in red).

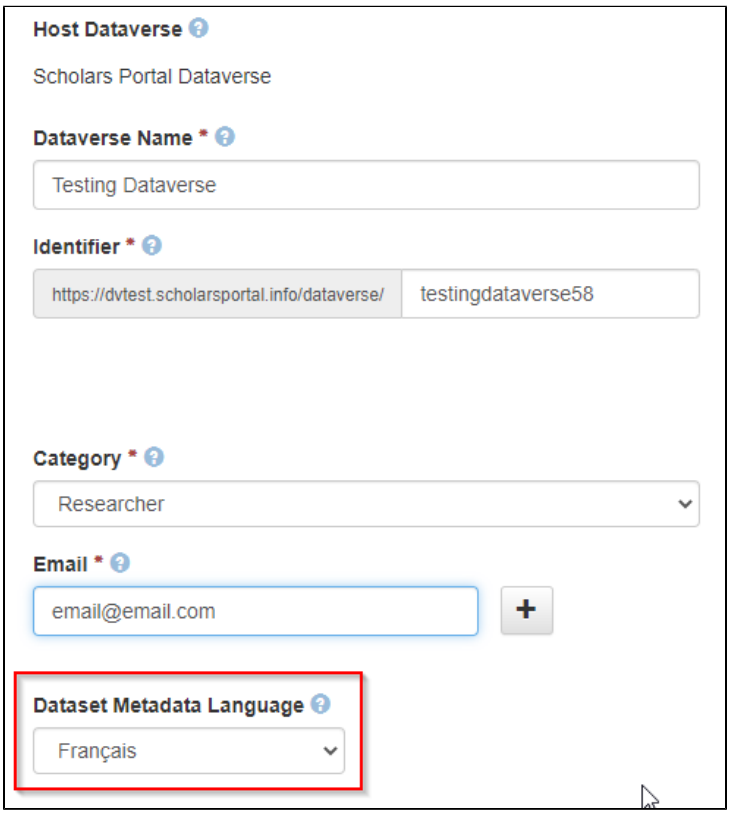


What is the use-case?

This feature provides internationalization support so that Dataverse collection admins can set the metadata language for new datasets and as a result the appropriate XML language tags are added to the DDI XML output. While it does not enforce using a specific language in the metadata, the setting indicates the requested language when editing and again when metadata is displayed.

How do I access this feature?

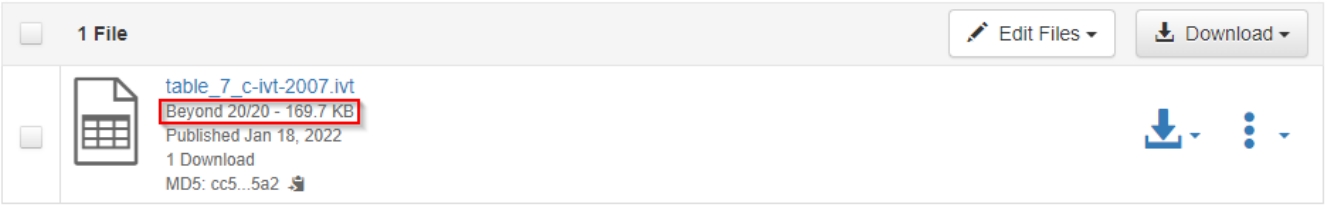
Note: Currently disabled. Dataverse collection admins can configure this feature. On the Dataverse collection page, go to the “Edit” menu and select “General Information.” Under “Dataset Metadata Language” select English or French. Sub-collections will inherit this setting, but it can be modified by the admin.

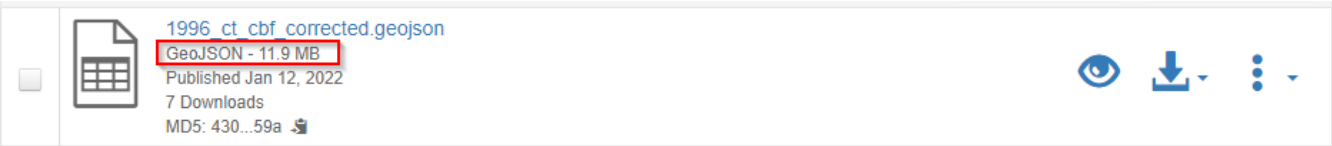


Improved identification of GeoJSON and Beyond 20/20 file formats

What is it?

Dataverse recognizes and identifies many different research data file formats and types. In this new release, Scholars Portal contributed an improvement to the Dataverse repository codebase to incorporate new identification support for the GeoJSON and Beyond 20/20 data file formats. This ensures that these files are correctly identified in the system and displayed in the interface’s file-level metadata. Previously, when these file formats were deposited, the Dataverse system displayed an ‘Unknown’ label for the formats.





What is the use-case?

Support for improved and expanded file format identification in the Dataverse system enables data managers and users to properly identify file formats and take the care needed to access and reuse them. Some file formats may become obsolete over time, or require specialized software which may have specific versions for reuse. It is therefore important that these formats are documented alongside deposited data files and associated metadata. In the future, if file formats are documented correctly, someone accessing them will know how to open them and hopefully reuse them.

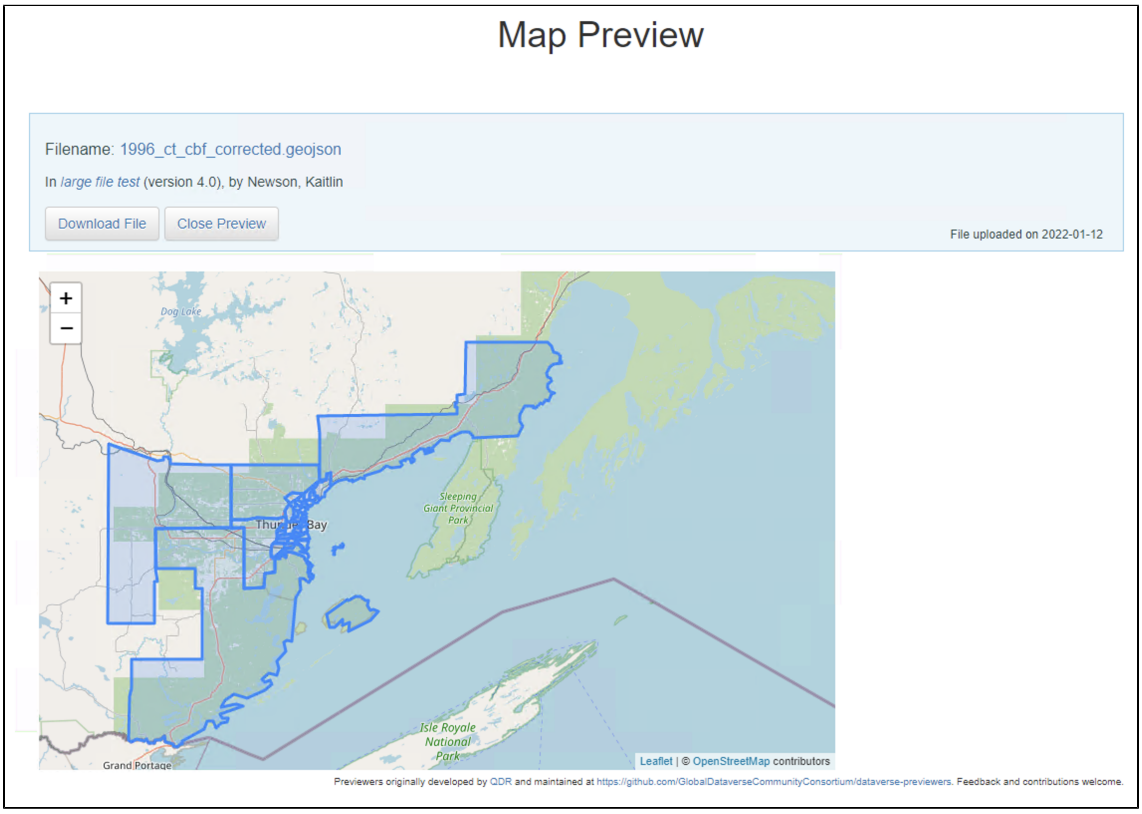
How do I access this feature?

When depositing a GeoJSON or Beyond 20/20 file in Dataverse, the system will automatically recognize and identify it correctly, displaying the label in the interface's file-level metadata. This identification further supports certain curation activities performed by the Dataverse repository system, such as file format normalization and file preview.

New geospatial data previewer for GeoJSON files

What is it?

Dataverse supports file preview for certain file formats, including tabular data files such as Excel and SPSS, and for images, videos, etc. With this release, Scholars Portal has expanded the file previewer framework to support geospatial data file preview. Utilizing the information stored within GeoJSON files, Dataverse can now preview spatial attributes and display coordinates such as points and boundaries on a map-based visualization previewer.



What is the use-case?

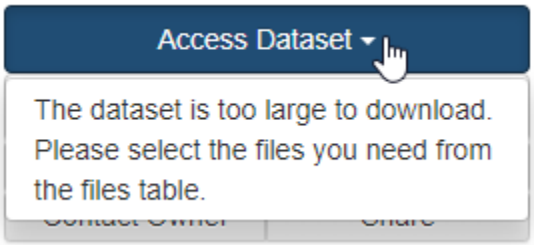
Researchers with geospatial datasets often rely on map-based visualizations to analyze and communicate research findings and results. With this new geospatial preview support, anyone accessing GeoJSON files in Dataverse will have the option to preview the data on a map to show results without downloading or opening the data in specialized software or tools. Furthermore, this map-based preview display can be used for visualization and analysis purposes, or, for linking to from within a researcher's paper or website for communication and data dissemination purposes.

How do I access this feature?

Any authorized user can deposit GeoJSON files into Dataverse. The new map-based file previewer will automatically display after the file uploads and the Dataset is created.

Improved notifications and messaging

- Users will see a warning message when attempting to download files as a zip package above Scholars Portal Dataverse's size limit (above 5 GB). Previously users would need to open the manifest.txt file in the zip package to determine which files were not downloaded. The warning message appears under "Access dataset" when all files in the dataset are selected at once.



- New notifications will be sent via the UI and as emails when a new dataset is created within a Dataverse collection for admins and curators.

New APIs

- New APIs will assist Scholars Portal staff to better manage users and determine their permissions/roles and activity.
- Users can now access expanded and new metrics through a new API.
- A new API to retrieve Guestbook responses has been added. This makes it easier to retrieve the records for large guestbooks and also makes it easier to integrate with external systems.
- A new file access API offers HTML/crawlable access view of the folders and files within a dataset.

Bug Fixes

- Unauthorized users will no longer be able to access summary statistics that appear in exported metadata for restricted files.
- Removal of files from physical storage for a deleted draft
- Search with accented characters now accepted e.g. (á, à, â, ç, é, è, ê, ë, í, ó, ö, ú, ù, û, ü…)
- And many more!

Please reach out if you have any questions or feedback.