TDAI Data Commons Gen 1
User Guide
1 Introduction ........................................................................................................... 3
  1.1 Document Summary ......................................................................................... 3
  1.2 Version Control .............................................................................................. 3
2 Capabilities ........................................................................................................... 4
  2.1 Portal Specifications ....................................................................................... 4
3 Browsing Data Commons Projects ...................................................................... 5
  3.1 Portal Overview .............................................................................................. 5
  3.2 Searching Projects .......................................................................................... 5
  3.3 Viewing Projects ............................................................................................. 6
  3.4 Project Action Buttons ..................................................................................... 7
    3.4.1 Download Dataset ..................................................................................... 7
    3.4.2 Edit Project ................................................................................................. 8
    3.4.3 Delete Project ............................................................................................. 8
4 Contributing to the Data Commons ................................................................... 9
  4.1 Publish New Dataset dialog ............................................................................ 9
    4.1.1 Author(s) .................................................................................................. 9
    4.1.2 Title .......................................................................................................... 10
    4.1.3 Keywords .................................................................................................. 10
    4.1.4 Summary .................................................................................................. 10
    4.1.5 Long Description ...................................................................................... 10
    4.1.6 Access Level .............................................................................................. 10
      4.1.6.1 Private Access Level ........................................................................... 10
    4.1.7 Categories .................................................................................................. 10
    4.1.8 Featured Image (deprecated) ..................................................................... 10
    4.1.9 Upload Files ............................................................................................... 10
    4.1.10 Features (deprecated) ............................................................................ 11
  4.2 Dataset Publishing completion ....................................................................... 11
5 Data Commons Policies ....................................................................................... 12
1 Introduction

The TDAI Data Commons Portal is a hub where data may be stored, curated, shared and collaborated on. The first generation of the Data Commons portal (aka Data Commons Gen1) is a web portal that was designed and implemented by a third-party organization that is no longer involved in the project. Currently, Data Commons Gen1 is hosted and maintained by the Ohio Supercomputing Center (OSC). Workarounds were implemented for the most limiting of the initial design issues, and Data Commons Gen1 is available for general use.

1.1 Document Summary

This document is a User Guide to the Data Commons Gen1 portal. It describes the interface and contains procedures on the tasks that a user may wish to accomplish using the Data Commons Gen1 portal.

1.2 Version Control

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>3/12/19</td>
<td>First Draft Started</td>
</tr>
</tbody>
</table>
2 Capabilities

The vision for the Data Commons for it to become the hub of data sharing and collaboration at the Ohio State University and beyond. This will entail not only being a data repository, but also hosting resources for analysis of data within the Data Commons itself. It also means being able to index data from sources around the University to enhance the library of available data resources.

Data Commons Gen1 mainly focuses on the data repository and sharing aspects of the vision of the TDAI Data Commons. Data Commons Gen1 allows users to create and manage projects and share data sets associated with those projects. For data consumers, there search tools and project tags to facilitate finding data sets relevant to one’s interest.

2.1 Portal Specifications

TDAI Data Commons Gen1 is hosted by the Ohio Supercomputing Center. Data storage is also provided by OSC.

The URL for the Data Commons portal is https://datacommons.tdai.osu.edu

For Gen1, the Data Commons portal requires a valid OSU name.# login to access. This means that project access classifications of “Public” and “OSU” are effectively the same.
3 Browsing Data Commons Projects

3.1 Portal Overview

A screen capture of the main page of the TDAI Data Commons Portal is below. The main components of the main page are:

- The navigation bar on the left of the window, including the Dashboard view (the main page), My Datasets (to show only Datasets/projects that you own, Requests, the About page and the Support page)
- Filtering options for determining which projects appear in the projects window
- The projects window that includes the project Title, Summary, Category, and other statistics pertaining to that project
- A link to add a new Dataset (aka Project) to the Data Commons

![Figure 1: Data Commons Project Page](image-url)

3.2 Searching Projects

There are a few ways to search or filter for Projects in the Data Commons:

- Using the search box – Currently the search box will search only keyword or Categories (or tags) against which the contributor or curator has added to the
project. Unfortunately, it does not search against Project Author as it advertises, nor does it search within the Project Summary or Long Descriptions. This search works well if the project has been tagged or categorized appropriately.

- Selecting by Project access classification (Public, Private, OSU or All) – This simply filters the Projects by whether they were set by the contributor or curator as Public, Private or OSU. The “Select none” checkbox in reality is a “Show all” checkbox, as it removes access classification filters.

- Filtering Projects by Category – This is probably the easiest way to search for projects. It relies on the contributor or curator to have added appropriate tags to the library of tags but is otherwise effective.

- Filtering Projects by Feature (deprecated) – Project Feature is a redundant project classification method that is like the project category method and is not being enforced or employed at this time. Contributors or curators may have added a Feature to their project however it is inconsistent across projects so is not a reliable way to discover project datasets.

- Filtering by Published or Uploaded Date (not recommended) – This may be useful if a user has prior knowledge as to the date that a project was published or uploaded. It is otherwise difficult to use given the lack of ability to filter within a date range – only a single date is supported for filtering.

3.3 Viewing Projects

Projects may be viewed by clicking on the Project Title within the search results (or Project List). When a Project has been selected for view, an overlay with the description of the project will appear, along with a button to Download the Dataset.

If the user logged in is the owner of the project, an additional link below the Download button will appear to allow the owner to Edit/Delete. At this time this link is NOT functioning as intended. To Edit or Delete a project, the floating buttons that are alongside the project should be used instead.

The detailed project view will not be populated for a project that has been set to Private, unless the user is the owner of that project. Unfortunately, the detailed project view will also not be populated even if a user is a named collaborator on the project (bug).
3.4 Project Action Buttons

Depending on the level of access a user has been granted for a project, up to 3 Project Action Buttons may appear when moving the mouse pointer over the project summary. These buttons, in order, are Download Dataset, Edit Project, and Delete Project (icons are circled in green below).

3.4.1 Download Dataset

The first Project Action Button is used to download the dataset that is associated with the project. The browser’s built-in download manager will be used to retrieve the file, so
depending on browser and operating settings used, the user may be prompted for a location to save the file, or whether to open the dataset.

3.4.2 Edit Project

This Project Action Button will only appear if a user is the owner/author of the project. A dialog like the one used to add a new project and dataset will appear when this button is selected. Please see the section on Contributing to the Data Commons for more detail on the fields in this dialog.

3.4.3 Delete Project

This Project Action Button is used to remove the project and dataset from the Data Commons. A dialog will appear asking the user to confirm deletion before the project is removed from the Data Commons. Note that if the project is removed accidentally it may be possible to recreate using backups of the Data Commons, but this is not guaranteed so please be sure that a project should be deleted prior to removing the project and dataset.
4 Contributing to the Data Commons

Contributing to the Data Commons is done by using the Add a New Dataset link at the top of the main Portal page. Note that any user that has a valid Ohio State name.# login can contribute a Project Dataset. However, the administration team does reserve the right to ban users if it is determined that a user is misusing or abusing the Data Commons.

To begin the dialog to contribute to the Data Commons, select the “Add New Dataset” button at the top of the Data Commons project list.

Add New Dataset

4.1 Publish New Dataset dialog

Once the Add New Dataset link has been selected, the first screen of the “Publish New Dataset” dialog will appear:

4.1.1 Author(s)

The Author’s email address. Currently only valid OSU email addresses are accepted. Enter the email address and click the Verify button. If the email entered is a valid OSU email address it will be confirmed and appear under the email address dialog box. Multiple Authors may be added to the project.
4.1.2 Title
A title for the project. Up to 50 characters are allowed.

4.1.3 Keywords
Add any keywords that can be used when searching for this project. New keywords can be added by clicking the + button next to the keyword dropdown. At least one keyword needs to be selected, and multiple keywords can be selected.

4.1.4 Summary
A short description for the project. A maximum of 200 characters are allowed. This is the description that will appear in the list of projects under the project title.

4.1.5 Long Description
A longer description for the project. This field accepts formatted text and can be used for a description of the data, links to tools that may be useful for working with the data, etc. If this description is for a non-Public project, it’s recommended that this information be placed in a document and added to an archive with the dataset. The archive can be uploaded to the Data Commons so any consumers of the dataset will also have background information on the dataset.

4.1.6 Access Level
Access levels are Public, Private or OSU. As the entire TDAI Data Commons is behind OSU name / password authentication, the Public and OSU access levels are functionally identical. At this time, it is recommended that projects are set to Public unless a project requires a closed set of collaborators.

4.1.6.1 Private Access Level
If a project dataset is set to Private access level, the email addresses of named OSU collaborators need to be added to the Private Emails list. At least one email address (of the contributor) should be added to the list. The list is verified in the same manner as the author list, where valid OSU email addresses are allowed.

4.1.7 Categories
This is similar in function to the Keywords tags. At least one category should be associated with the project. The Category assigned to the project will appear in the Project Summary page, to the left of the Title and Summary of the project.

4.1.8 Featured Image (deprecated)
This image appears behind the Category banner on the Project Summary page. It does not work consistently and is not required.

4.1.9 Upload Files
This opens a dialog so a file can be selected for upload. It is recommended that the dataset to added to an archive (.zip, .tar, etc.) so that multiple files may be included in the
upload, as only one file may be associated with each project. Within the archive, it is recommended that the dataset be included along with any metadata, documentation, tools, etc. If tools or other materials are available elsewhere on the Internet, it may be appropriate to include a README file with the location of other materials.

4.1.10 Features (deprecated)
As there are already fields for Keywords and Category, the use of the Features field for further tagging is not needed at this time.

4.2 Dataset Publishing completion
Once the Submit Dataset button is pressed an email will be sent to the contributor to accept the contributor role for this project. The Project will not be added to the Project List on the Data Commons until the contributor has confirmed that they are the owner for this project.
5  Data Commons Policies

About the TDAI Data Commons

The TDAI Data Commons is a collection of data sets, metadata, tools and methods that is meant to be a place to collaborate using data. It is maintained by the Translational Data Analytics Institute at the Ohio State University as a place where data can be used for collaboration and research.

About the Ohio State University and the Translational Data Analytics Institute

https://www.osu.edu
https://tdai.osu.edu/about/

Accessing the TDAI Data Commons

Permission to access to view and obtain data from the Data Commons depends on the controls assigned to a data set. Each individual data set must be set to one of the following levels of access:

- **Private** – only users that are explicitly listed by the data contributor can collaborate on this data set. The contributor may add or remove members after the data set has been added to the Data Commons.
- **OSU** – users with a valid osu.edu login may use and collaborate on this data set
- **Public** – currently this is effectively the same as the OSU access level, as the portal is only accessible to users with valid osu.edu credentials. Eventually the portal will be available to the general public whereupon this level of access would grant access to anybody (even without a login)

*In the future, access to any Public data will not require a login.* Those wishing to contribute data to the Data Commons must have an account to do so; data may not be contributed anonymously.

*Future iterations of the Data Commons will contain finer access controls such that other portal pages within the Data Commons become accessible depending on privilege level. Those sub-portals will contain data sets, tools, etc. specific to project. Currently all projects are self-contained within a single Data Set Entry.*

About the Data

TDAI and the Ohio State University make no claim on the accuracy, completeness, level of de-identification or quality of the data found in the TDAI Data Commons. The views and opinions of authors that may be expressed within data sets in the TDAI Data Commons do not necessarily state or reflect those of TDAI or the Ohio State University and may not be used for advertising or product endorsement purposes.

In the TDAI Data Commons, data sets of any kind are allowed. At this time, only approved contributors with valid .osu.edu credentials will be permitted to contribute data to any of the TDAI Data Commons. It is recommended, but not required, that contributions are
released with an open license, such as with the Open Knowledge framework (https://okfn.org)

Using the Data

Data contributed to the TDAI Data Commons should be anonymized prior to its contribution. Additionally, it is a violation of our data use policy for consumers of the data to use the data to reveal any individual identities.

Consumers of the data should acknowledge TDAI and the dataset(s) utilized in any research, presentation, publication, etc.

Contributing Data

At this time, only approved contributors with valid .osu.edu credentials will be permitted to contribute data to the TDAI Data Commons. Contributions may be reviewed by TDAI prior to becoming available for public consumption.

File Format: Any file format will be accepted; however if there are multiple files associated with the data set, or if there are metadata files, tools, or documentation that go along with the data, all of these files should be archived (.zip, .tar.gz, etc.) prior to contribution. If tools or source code is available at other repositories (github, sourceforge, etc.) you may include a README in the archive with links to the external repository.

Size Limitation: Currently there is no size limitation for data set contribution. However, if you are contributing data exceeding 50GB please contact us first using the Support link on the Data Commons portal.

Access Restrictions and Changes: Contributions that are set to a Private level of access require the contributor to manage the list of accounts that are allowed access. Additionally, the contributor may set the data set to a different level (either more or less restrictive). If ownership of a data set changes or if the original contributor would like to delegate contributor privileges to a different individual, please contact us using the Support link on the portal page.

Data Privacy: Please ensure that data are anonymized such that any personally identifying information is removed or obscured.

Curation: Data contributors are expected to play a significant role in the curation of the data contributed. Curation assistance is available from TDAI with the data curation processes (please contact us through the Support link on the Data Commons portal), but the contributor should always be the primary curator as the contributor is most familiar with the data.