SODA User manual

Bhavesh Patel (bpatel@calmi2.org)
Parya Aghasafari (paghasafari@ucdavis.edu)
Karl Helmer (helmer@nmr.mgh.harvard.edu)
Harshit Srivastava (hsrivastava@calmi2.org)

October 2019

Contents

Page

1 What is SODA? 2

2 Download and Open the Application 3

3 Main Interface 6

4 Prepare Metadata 7

5 Prepare Dataset 7
5.1 Specify dataset files . . . 8
5.1.1 Organize my dataset . . . 8
5.1.2 Dataset already organized . . . 15
5.1.3 Smart Organize . . . 16
5.2 Handle file format conversion . . . 16
5.3 Specify metadata files to be attached . . . 16
5.4 Validate dataset before generating . . . 18
5.5 Generate dataset . . . 18
5.5.1 Modify current dataset . . . 18
5.5.2 Create new dataset locally . . . 19
5.5.3 Create new dataset directly on Blackfynn . . . 20
5.5.4 Progress window . . . 20

6 Manage and Submit Datasets 21
6.1 Login to your Blackfynn acount . . . 22
6.2 Create new dataset . . . 24
6.3 Add metadata to Blackfynn dataset . . . 25
6.4 Submit new dataset . . . 25
6.5 Manage dataset permission . . . 26
6.6 Annotate dataset . . . 29

7 Help 29

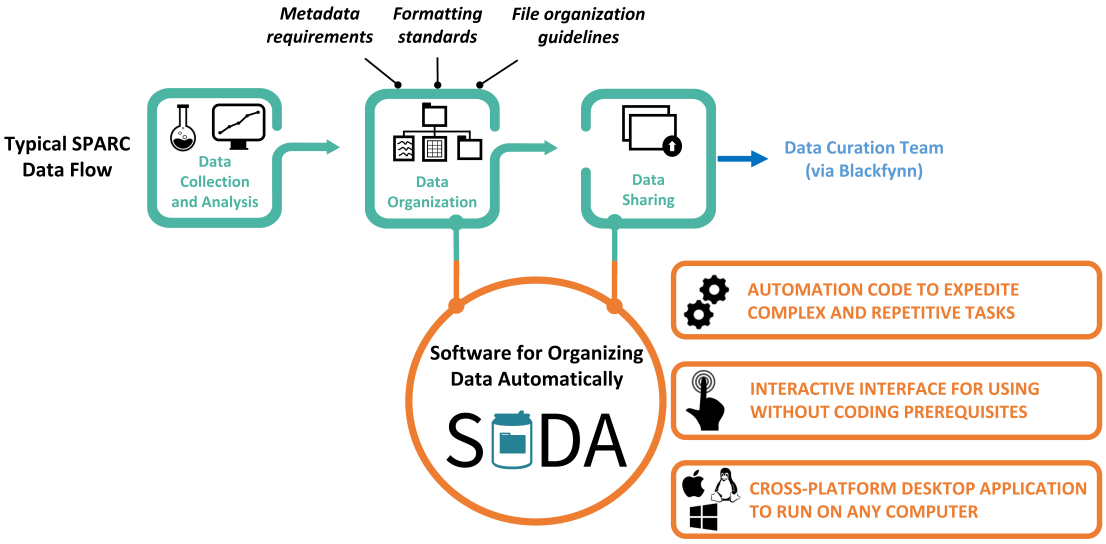
1 What is SODA?

Once a milestone of a project is reached, SPARC investigators need to organize and prepare the corresponding datasets according to standards established by the Data Standard Committee, before sharing them on the Blackfynn platform for review by the Curation Team. To achieve that, investigators must identify the applicable guidelines for their specific datasets and generate metadata. This data organization process could become time-consuming and overwhelming as more and more data is generated, and eventually steer focus away from data acquisition and analysis, which should ideally remain the primary focus of the SPARC investigators.

Software for Organizing Data Automatically (SODA) is a computer program intended to facilitate the data organization and curation processes through interactive tools and automation. If at any point you think: "Ah, if only this step could be simplified", actually it can be with SODA! Especially, SODA would allow users to:

- Conveniently organize datasets following the SPARC folder structure
- Convert file format to SPARC defined standards
- Generate metadata files with pre-populated fields, some automatically
- Validate dataset with the validator developed by the Curation Team
- Upload organized dataset directly on Blackfynn to avoid duplicating files locally

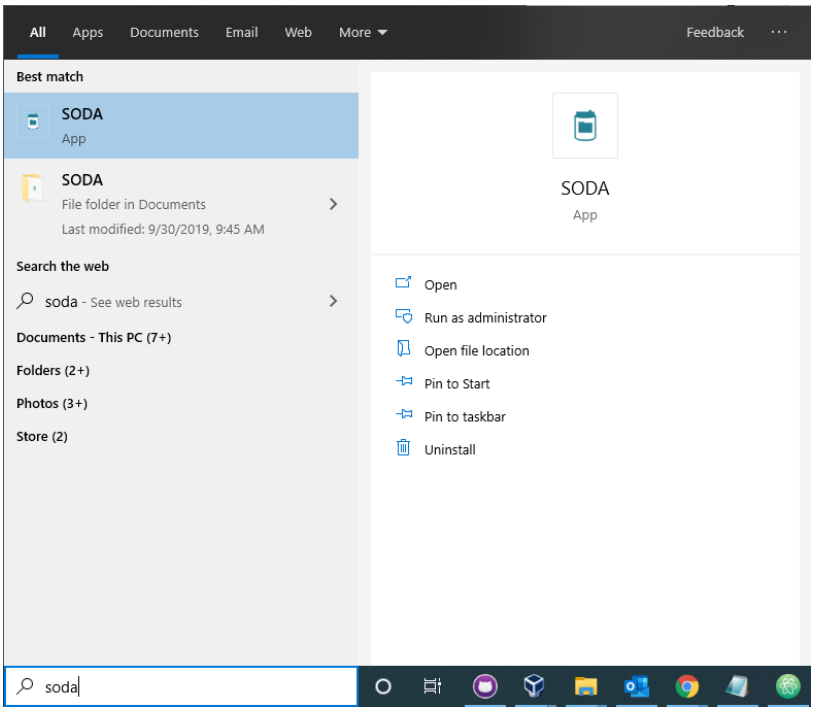
SODA is distributed as an easy to install application for Windows, Mac, and Linux platforms. The front-end (Graphical User Interface or GUI) of SODA is built with Electron, an open-source framework developed and maintained by GitHub that conveniently combines HTML, CSS, and Javascript, while the back-end is developed in Python (v3.6). All source codes and files are shared with an open source license (MIT) to permit user modification without restrictions.



2 Download and Open the Application

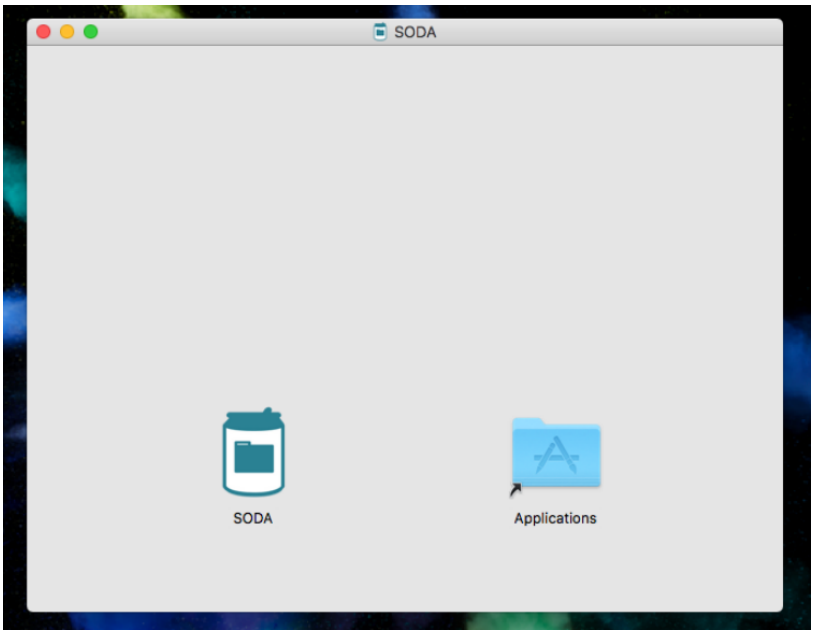
To use SODA, follow the instructions below for your specific platform:

- Windows (developed on Windows 10, tested on Windows 10 and 7)
    1. Download Windows installer from the SODA Github page: https://github.com/bvhpatel/SODA
    2. Double click on the downloaded installer, it will start the installation process (if there is a permission issue try right click and select "Run as administrator").
    3. Select the folder where you want to install
    4. You can chose the option of having a shortcut on the Desktop or Quick Launch
    5. Once the application is installed, you can run the application by either running the desktop shortcut, searching for "SODA" on the start menu, or from the installation folder.
    6. To stop the application, simply click on the cross in the top right corner.
    7. The downloaded installer file could be deleted since it is not necessary to run SODA. The SODA app can be uninstalled from the Program manager (under Control Panel) as any other app on Windows.



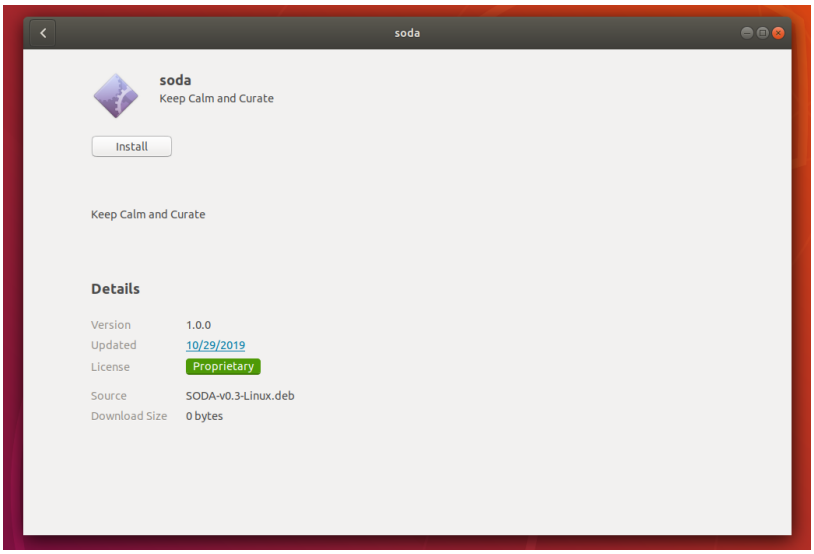
- MAC-OS (developed and tested on Mojave 10.14)

    1. Download MAC installer from the SODA Github page: https://github.com/bvhpatel/SODA
    2. Double-click on the downloaded installer and, in the installation window, drag the SODA icon into the "Applications" folder. Security preferences may need to be changed (System Preferences>Security & Privacy>General Tab>"Open anyway").
    3. To stop the application, click on the red circle/button in the top left corner.
    4. The downloaded installer could be deleted since it is not necessary to run SODA. To uninstall the SODA app, simply delete it from your computer's "Applications" folder.



- Linux (developed and tested on Ubuntu 18.04.2 LTS)

  1. Download Linux installer from the SODA Github page: https://github.com/bvhpatel/SODA
  2. Run the downloaded installer by double-clicking it.
  3. Click on 'Install' in the 'Ubuntu Software' window.
  4. Enter your password (if prompted).
  5. Installed application will be visible in 'Activities' section on Ubuntu.
  6. To stop the application click on the red circle/button in the top left corner.
  7. The downloaded installer could be deleted since it is not necessary to run SODA. To remove the SODA app, run 'Ubuntu Software', search for 'SODA' and click on 'Remove'.

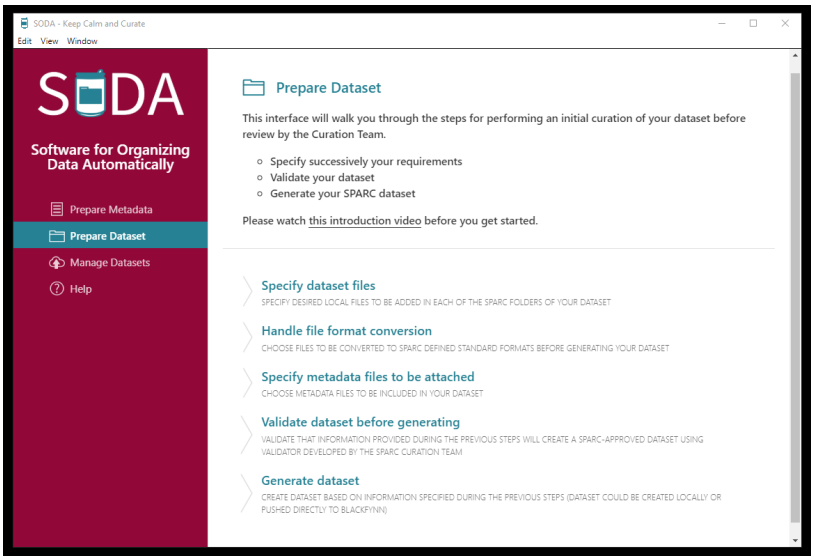


Note:

- Screenshots provided subsequently in this User Manual are from a Windows platform, but they are essentially the same for MAC and Linux baring some OS-specific native design differences.

3 Main Interface

A picture of the main interface is shown below. The side menu allows to navigate between the three main features of SODA, "Prepare Metadata", "Prepare Dataset" and "Manage and Submit Datasets", as well as access the "Help" section. The design of the GUI is based on the Electron demo code, which is freely available.



Note:

- The window size of the application frame can be easily adjusted to fit your screen by dragging the borders.
- Closing the interface will stop any on-going process (dataset organization, uploading, etc.).
- If any error occurs, close SODA and re-open it.
- Tooltips ***(i)*** are used throughout the interface to provide the user with additional information.

4 Prepare Metadata

Coming soon... tentative date: December 2019

5 Prepare Dataset

This UI will walk you step by step through the data preparation process. The following options that are described:

1. Specify dataset files
2. Convert file format
3. Specify metadata files

4. Validate
5. Generate dataset

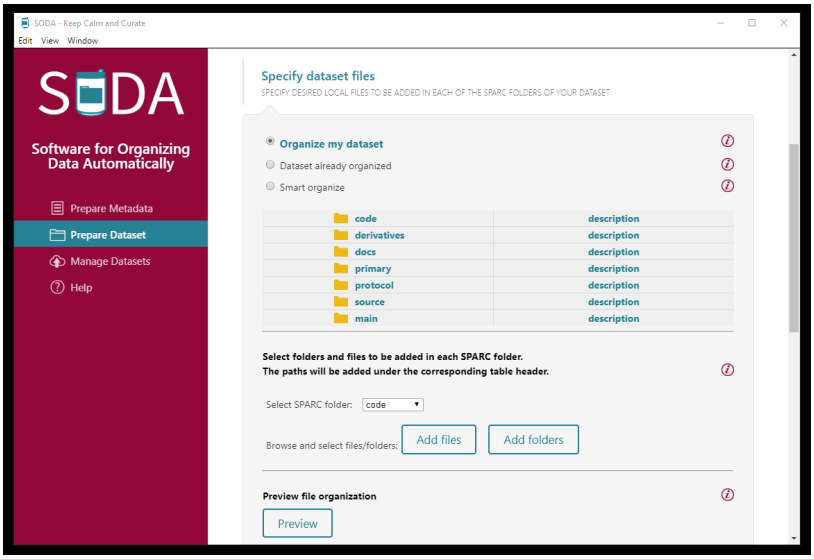
The first four options permit the selection of desired actions on the dataset which can then be applied under "Generate dataset" (nothing is actually generated before that step).

5.1 Specify dataset files

During this step, you can specify the files/folders that you want to include in each SPARC folder of your dataset. The paths of the files/folders you decide to include will be shown in a table (where each header corresponds to one SPARC folder, and the "main" header corresponds to the main dataset folder containing all SPARC folders). The program will then use this table to generate your dataset when you reach the "Generate dataset" step.

5.1.1 Organize my dataset

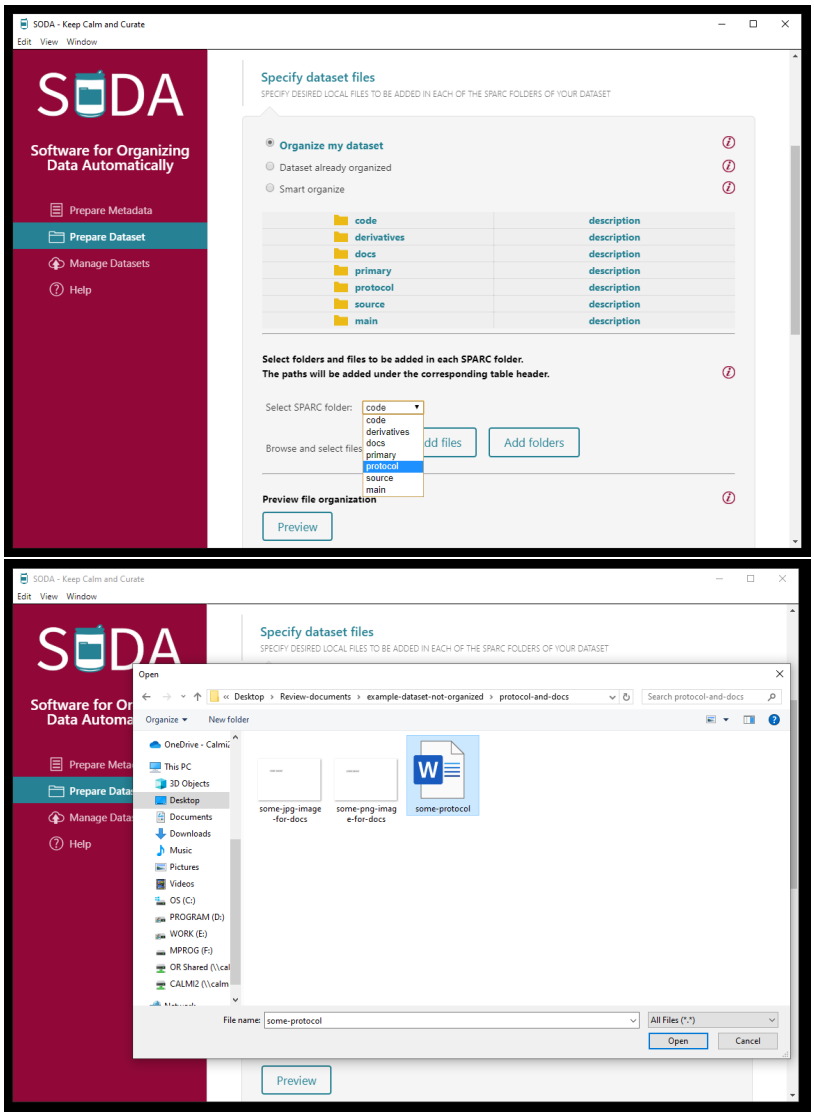
If your dataset is not organized according to the SPARC defined folder structure, this is where you can specify the files/folders that you want to include in your dataset within each of the SPARC folders.

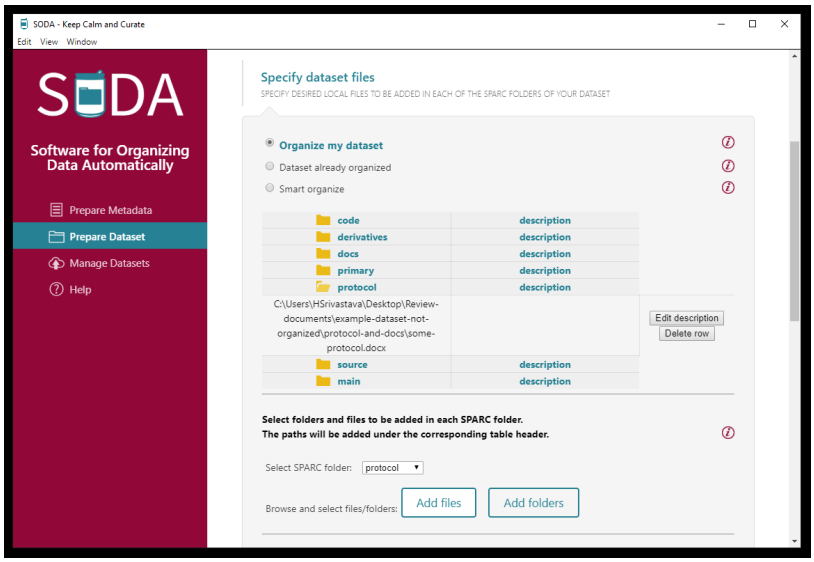


There are several ways to do that:

1. Browse and add:
    - Select SPARC folder from the drop-down list
    - Click on "Add files" or "Add folders" to browse and select

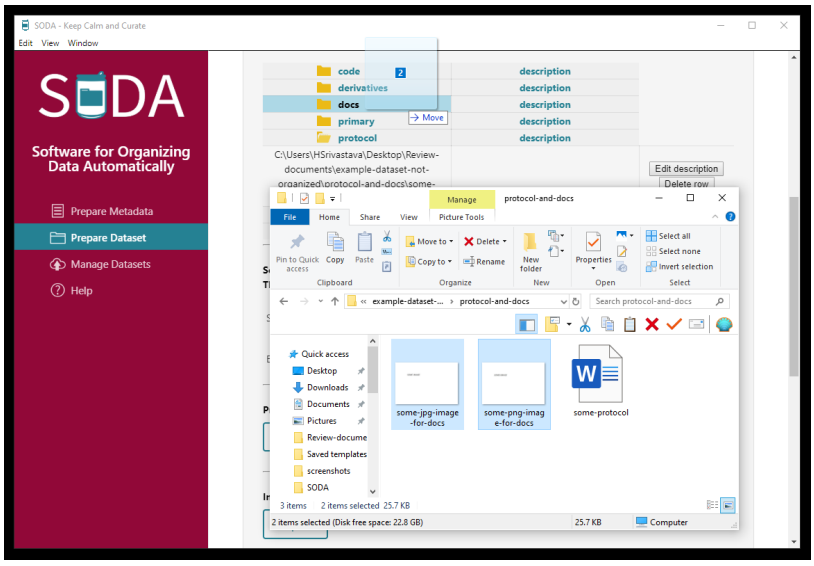
- Path of selected files/folder will appear in the table, below the header of the selected SPARC folder

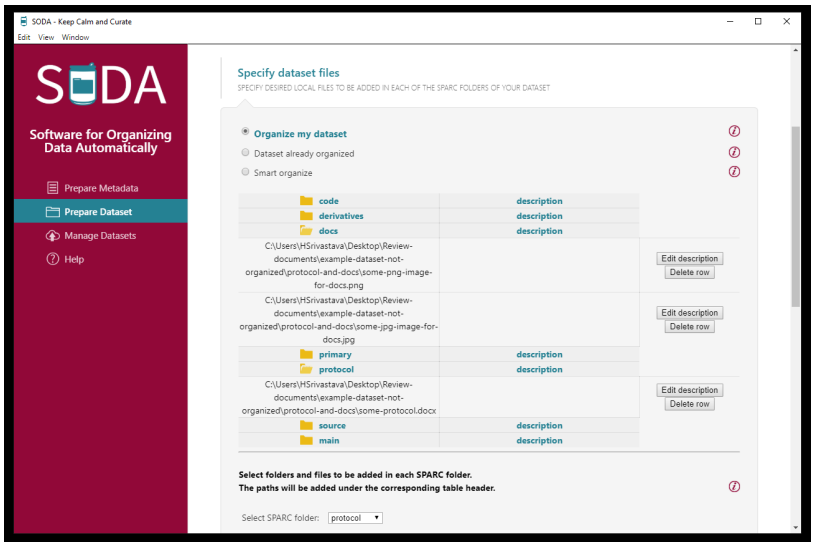




2. Drag and drop:

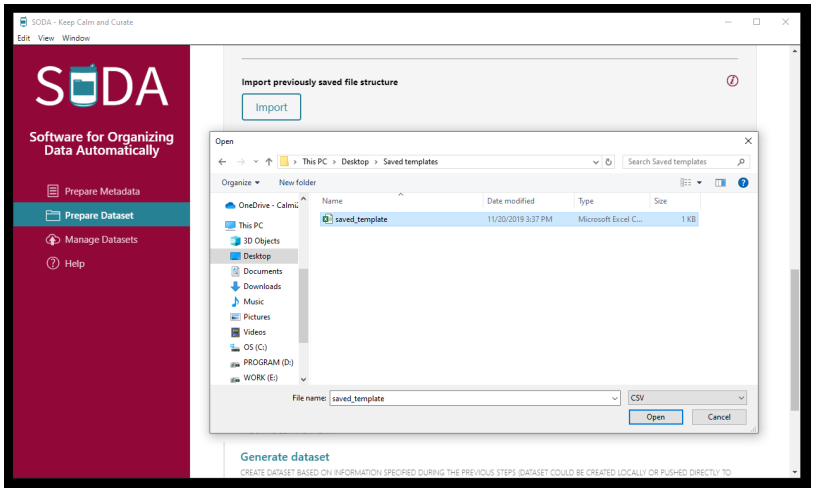
   - Drag any local files/folders to any of the table headers (e.g. "code", "derivative", etc.). Mouse pointer should be inside the box of the desired folder (font color of the header will change to black when pointer is properly positioned).
   - Path of dragged files/folder will appear in the table, below the corresponding header

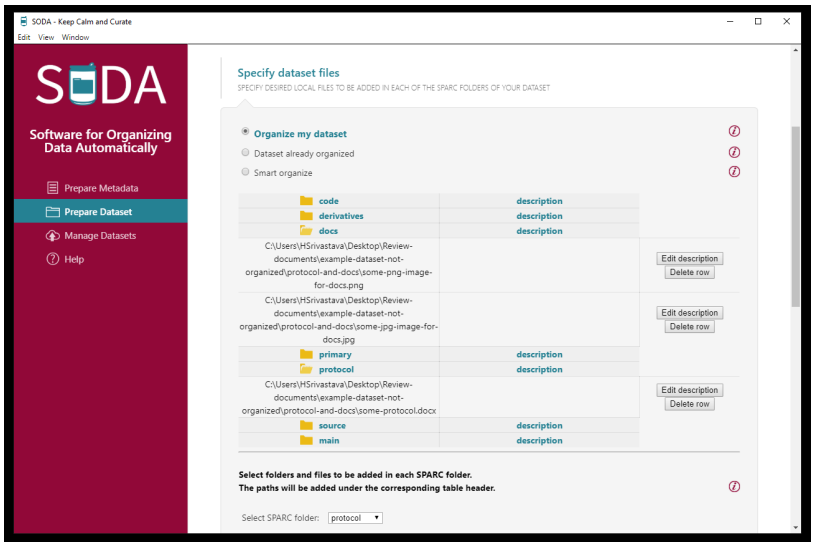




3. Import previously saved organization:

   - Click on "Import"
   - Select previously saved progress (see explanation of "Save" button below)
   - Paths will be added to the table along with the description (see below for explanation of "description" field). The paths/descriptions already in the table will not be removed.

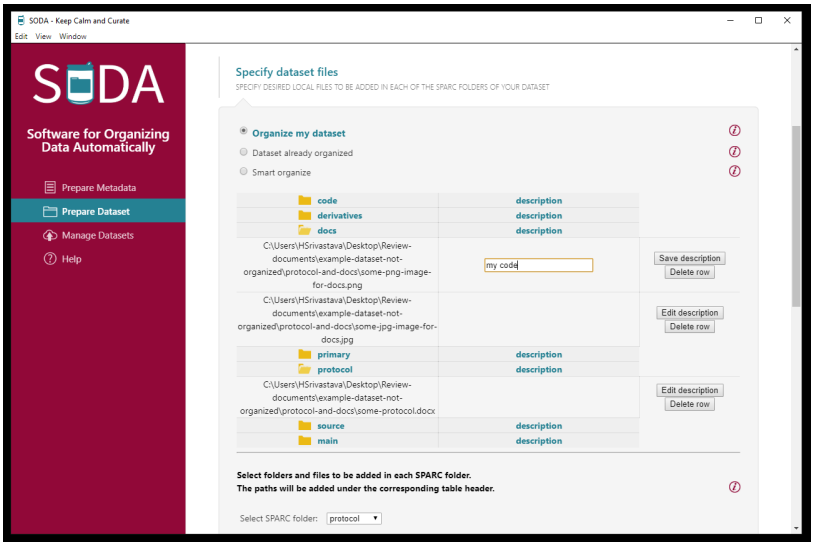


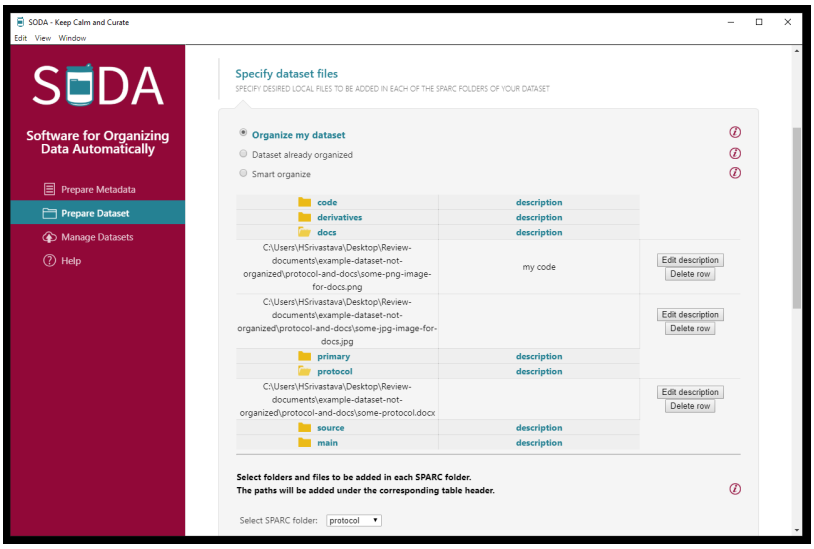


Click on "Delete row" to remove any item from the table.

You can add a description for each of the files in the table which will then be added to the manifest file.

- Click on "Edit description" next to the path to add a description
- Enter the description in the textbox
- Click on "Save" to record

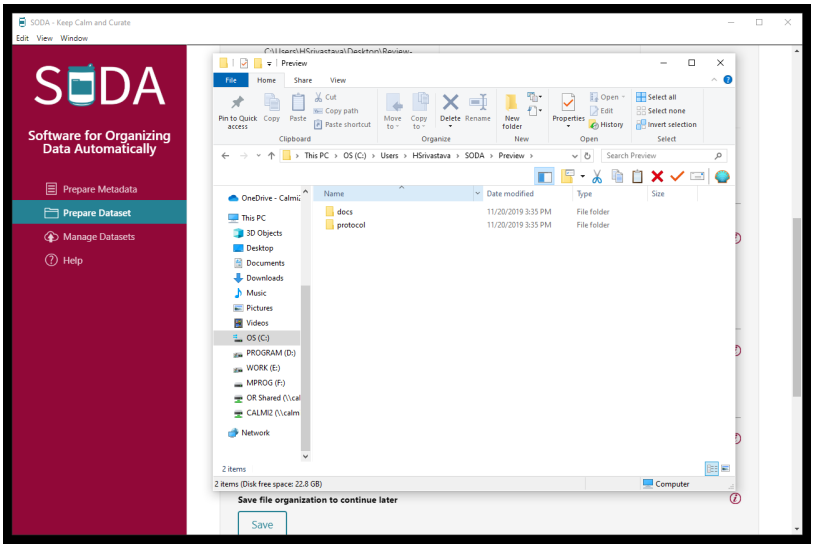




Click on a header of the table to collapse the corresponding rows with file/folder paths if you want a cleaner looking work space.

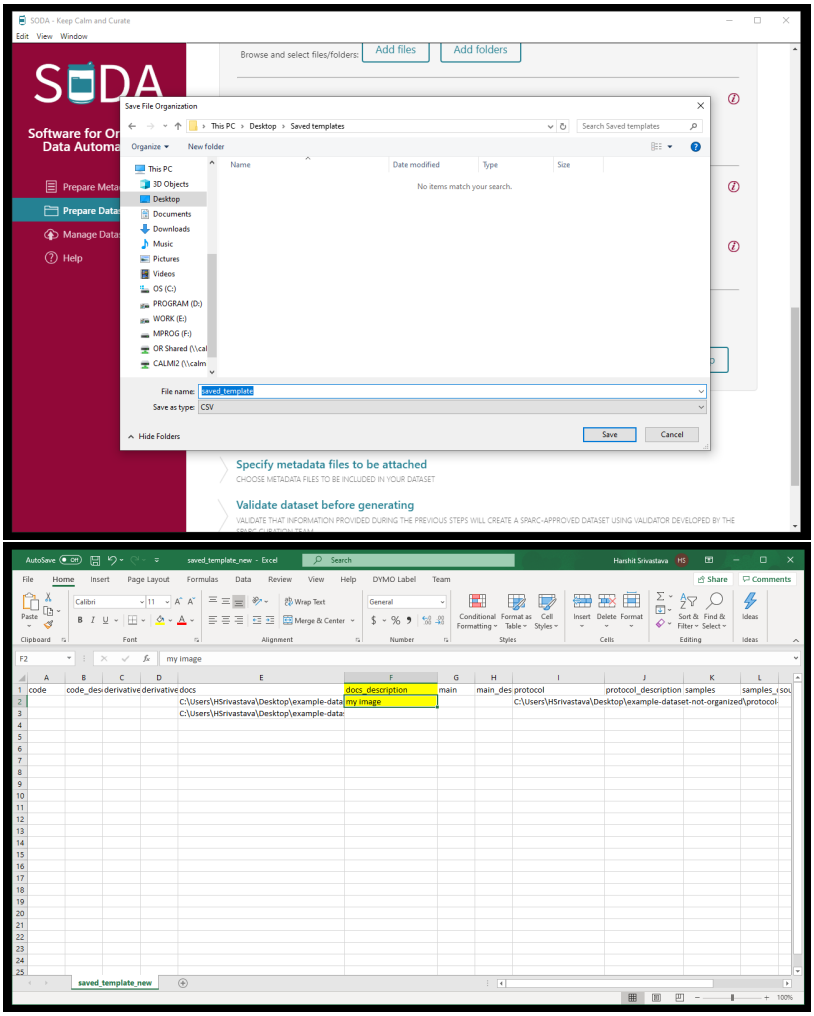
You can conveniently preview file organization to ensure it's organized as desired before the actual dataset is created during the "Generate dataset" step.

- Click on "Preview" to view the file structure of the dataset you organized
- All the files/folders displayed are mock files which resemble the original files but don't take up any space on the user's computer
- Click on "Delete Preview Folder" to get rid of the files created for preview (only one preview folder is allowed to be created at a time to avoid clustering user's computer)



You can save your dataset organization progress by saving the table as a CSV file.

- Click on "Save" and browse to the desired location to save all the paths from the table into a CSV file
- Saved file will contain all the paths as well as descriptions
- Saved file could be edited through a CSV editor (e.g. MS Excel) to manually include paths and descriptions

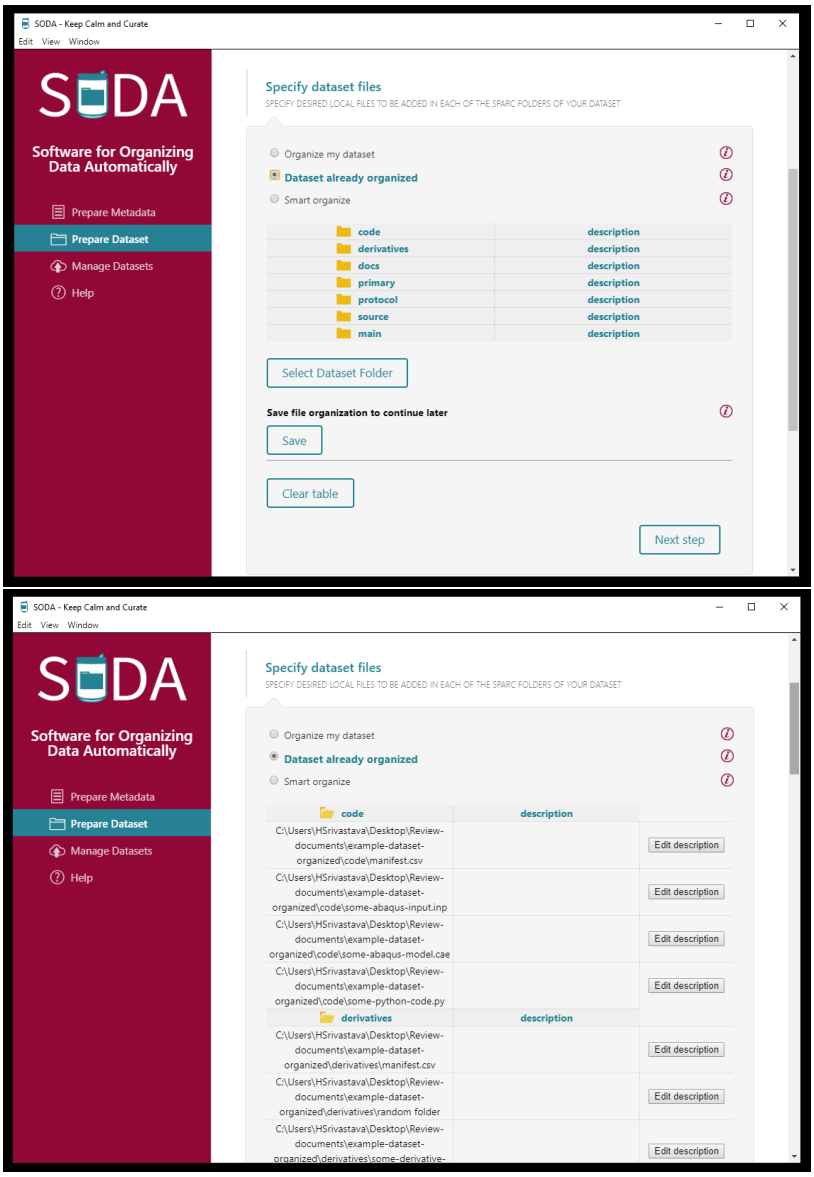


You can clear ALL information contained in the table by clicking on "Clear table". Use this option when you want to remove all paths and descriptions from the table and start fresh.

5.1.2 Dataset already organized

If your dataset is already organized according to the the SPARC defined folder structure, this is where you can import it for further processing.

- Click on "Select Dataset Folder", browse files, and select the desired organized dataset
- Path of files/folders from that dataset will appear in the adjacent table.



You can click on the "Edit description" button to add a description for the corresponding file/folder (click on "Save description" when you are done). These descriptions will be in-

cluded in the manifest files if you request them under "Generate metadata".

The table can be cleared with the "Clear table" button.

Note:

- The table automatically clears if you select another valid dataset.
- An error message will be displayed if the folders within the selected dataset do not match with the SPARC folder structure.

5.1.3 Smart Organize

Coming soon... Tentative date: December 2019

5.2 Handle file format conversion

Coming soon... Tentative date: May 2020

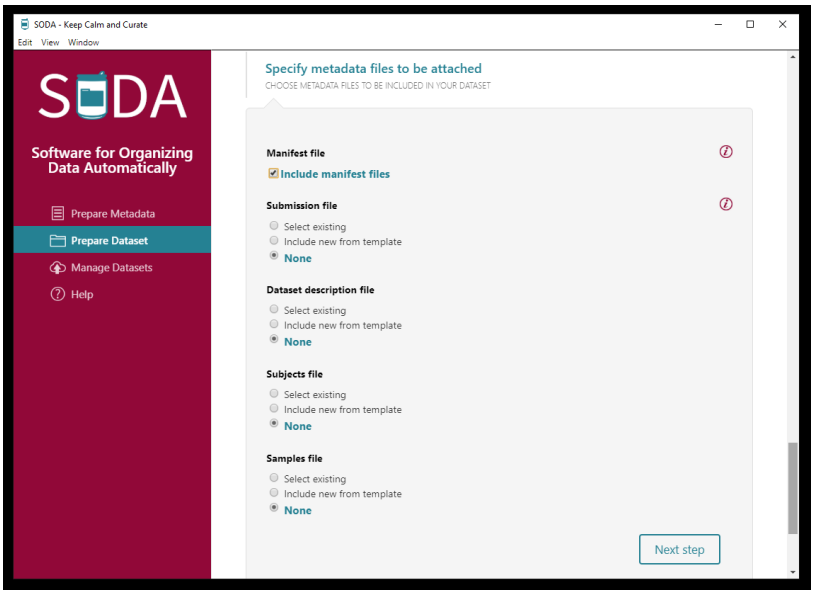
5.3 Specify metadata files to be attached

During this step, you can specify the metadata files you wish to include in your dataset:

- manifest
- submission
- dataset_description
- subjects
- samples

Requested metadata files will then be included in the dataset during the "Generate dataset" step.

To include manifest files, simply check the "Include manifest files" box. A "manifest.xlsx" file will be added to each SPARC folder with the "filename", "timestamp", and "file type" fields automatically populated while the "description" field will be filled with the descriptions provided (if any) in the table during the "Specify dataset files" step.



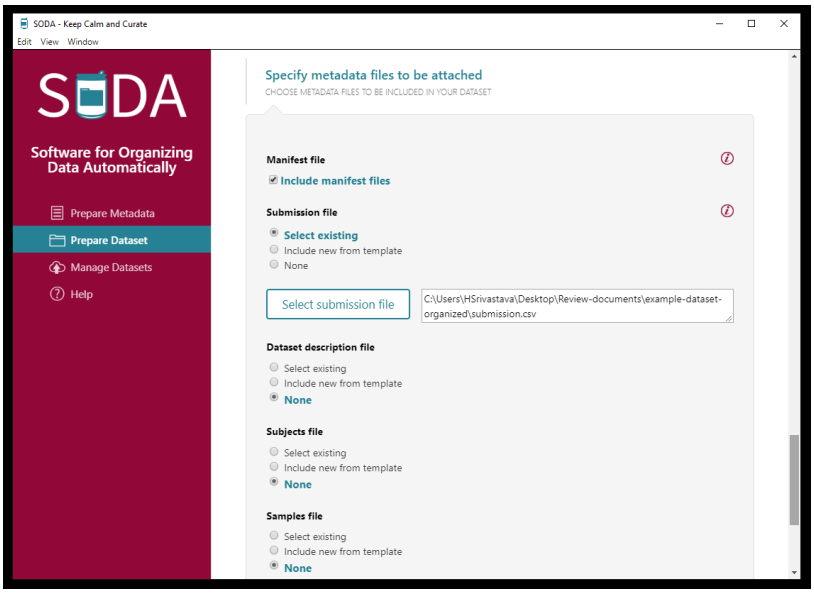
For all other files, there are three options:

1. "Select existing": this option allows to select an existing file by clicking on the associated button and navigating to the file or entering the file path directly in the textbox.

   NOTE: Only a file with the expected name should be selected (for eg - only "submission.xlsx" or "submission.csv" will be accepted for submission file), and all other file names will result in an error later during dataset generation step.

2. "Include new from template": this option allows to include an empty template provided by the Curation Team, which could then be filled out later on.

3. "None": select this option to not include the corresponding file (e.g., if it is already included in the selected dataset folder or if you wish to add it later).



5.4 Validate dataset before generating

Coming soon... tentative date: December 2019

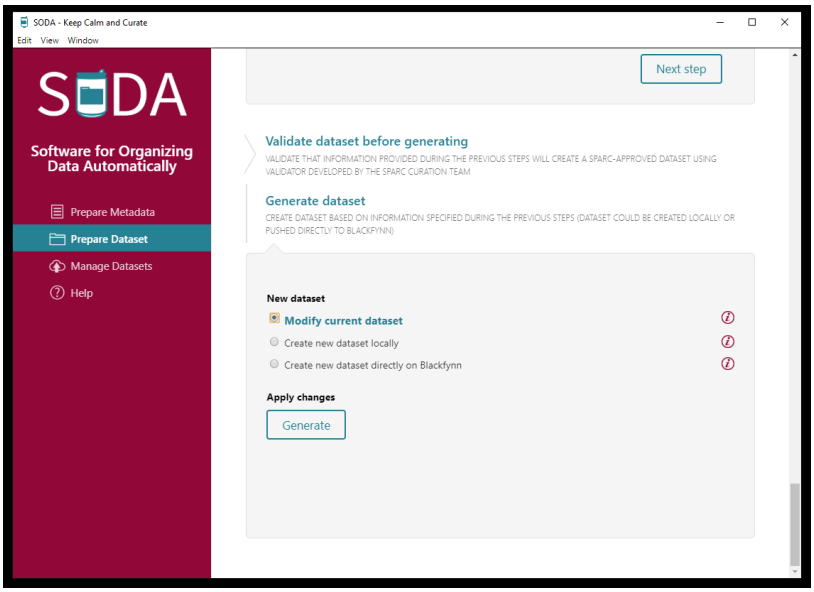
5.5 Generate dataset

Once you have selected all desired actions from the previous stepss, you can apply them here through one of the three options explained below.

5.5.1 Modify current dataset

If "Modify current dataset" is selected, the requested actions will be applied to the current dataset folder (selected under the "Dataset already organized" option of the "Specify dataset files" step). Simply click on "Generate" to do so.

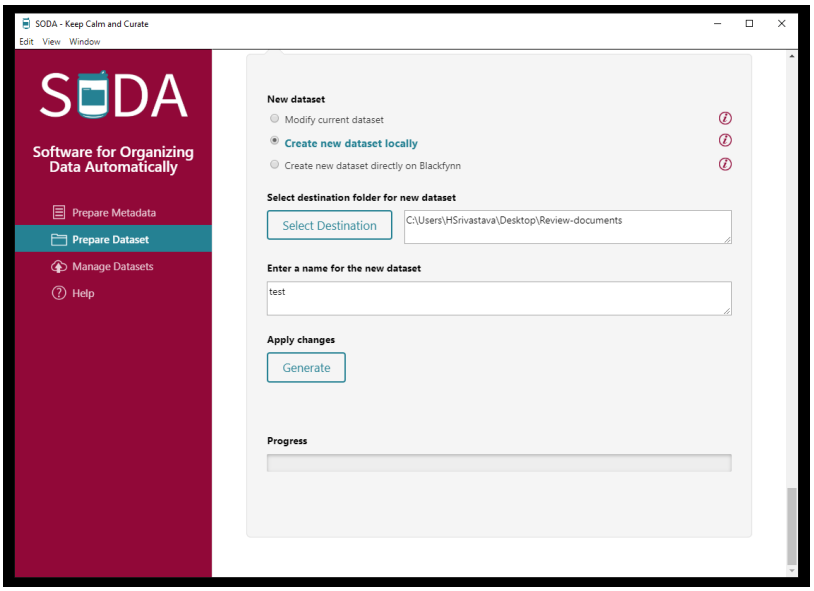
NOTE: This option is only available when you select the "Dataset already organized" option of the "Specify dataset files" step.



5.5.2 Create new dataset locally

If "Create new dataset locally" is selected, existing dataset will be copied and the requested actions will be applied to a new folder created at a desired location. No modifications will be made to the original dataset.

- Click on "Select destination" to navigate to the desired location of the new dataset or enter the path directly in the textbox. Only a folder can be selected.
- Enter the name of the new dataset (a folder with this name will be created at the selected destination)
- Click "Generate" to apply requested actions



S

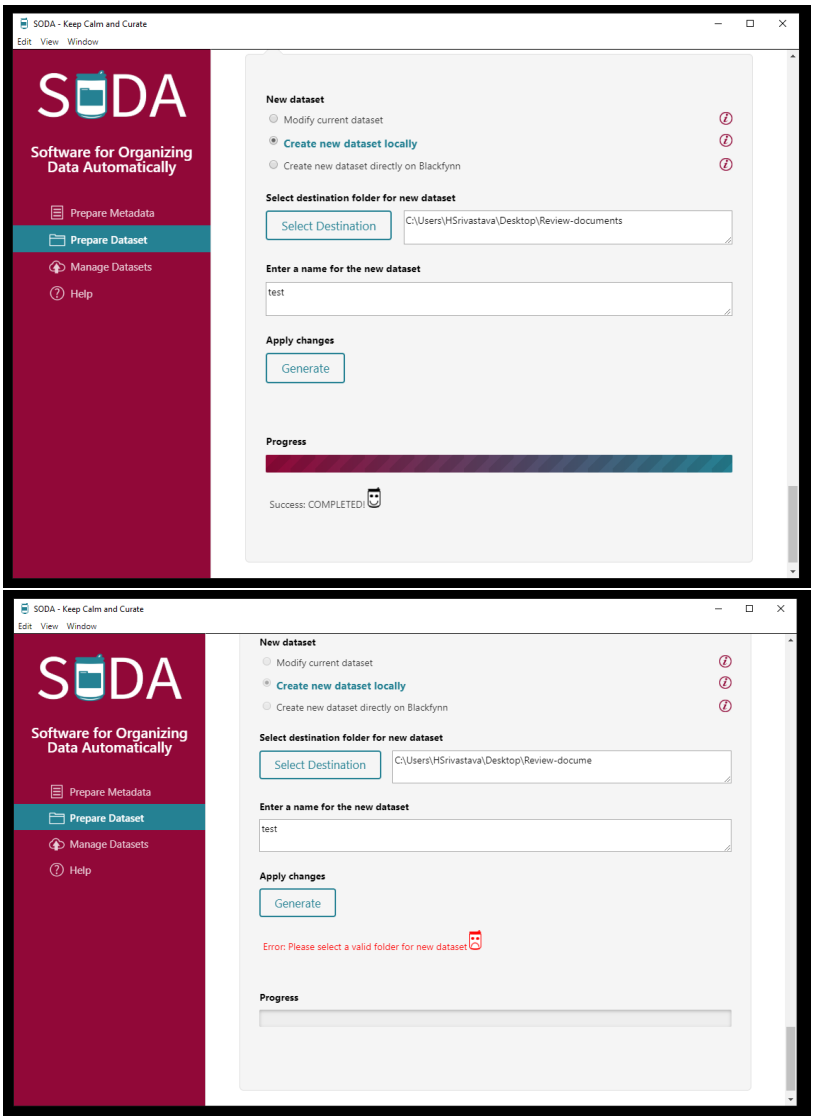
5.5.3 Create new dataset directly on Blackfynn

If "Create new dataset directly on Blackfynn" is selected, selected files and folders from local system will be uploaded to the corresponding Blackfynn dataset along with the requested metadata files. No modifications will be made to the original dataset.

- Click on "Load Existing Account(s)" to load the Blackfynn accounts stored on the system.
- Select one of the accounts from the "Select Account" drop-down list.
- Click one of the datasets from the "Select Dataset" drop-down list.
- Click on 'Generate' to start uploading the files and folders on to your Blackfynn dataset.

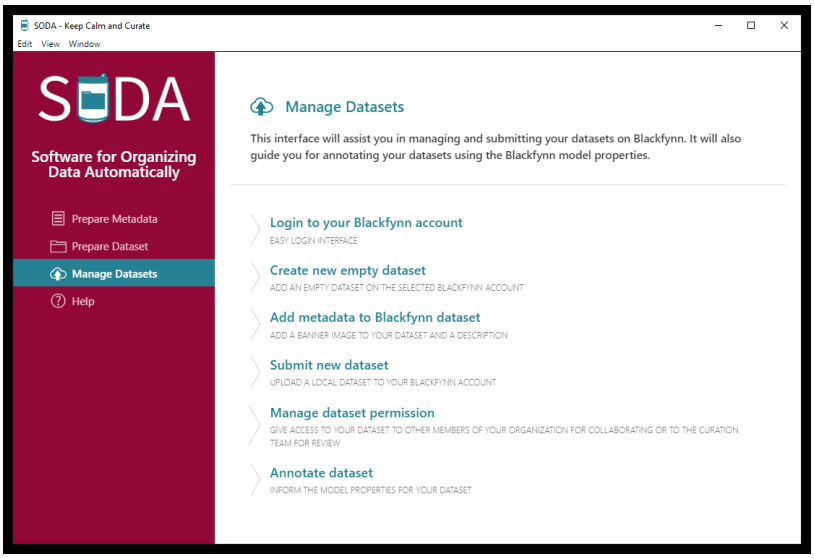
5.5.4 Progress window

For all three options, progress of the preparation process is displayed in the "Progress" box located below the "Generate" button. It ends with a success message when finished without error or displays an error message in red informing the user about the nature of the error.



6 Manage and Submit Datasets

This interface will assist you in managing and submitting your datasets on Blackfynn. It will also guide you for annotating your datasets using Blackfynn's model properties.



Uploading files on Blackfynn directly from a computer without using their webpage (thus with SODA) requires installing the Blackfynn Agent. Please follow the instructions located at https://developer.blackfynn.io/agent/index.html before continuing.

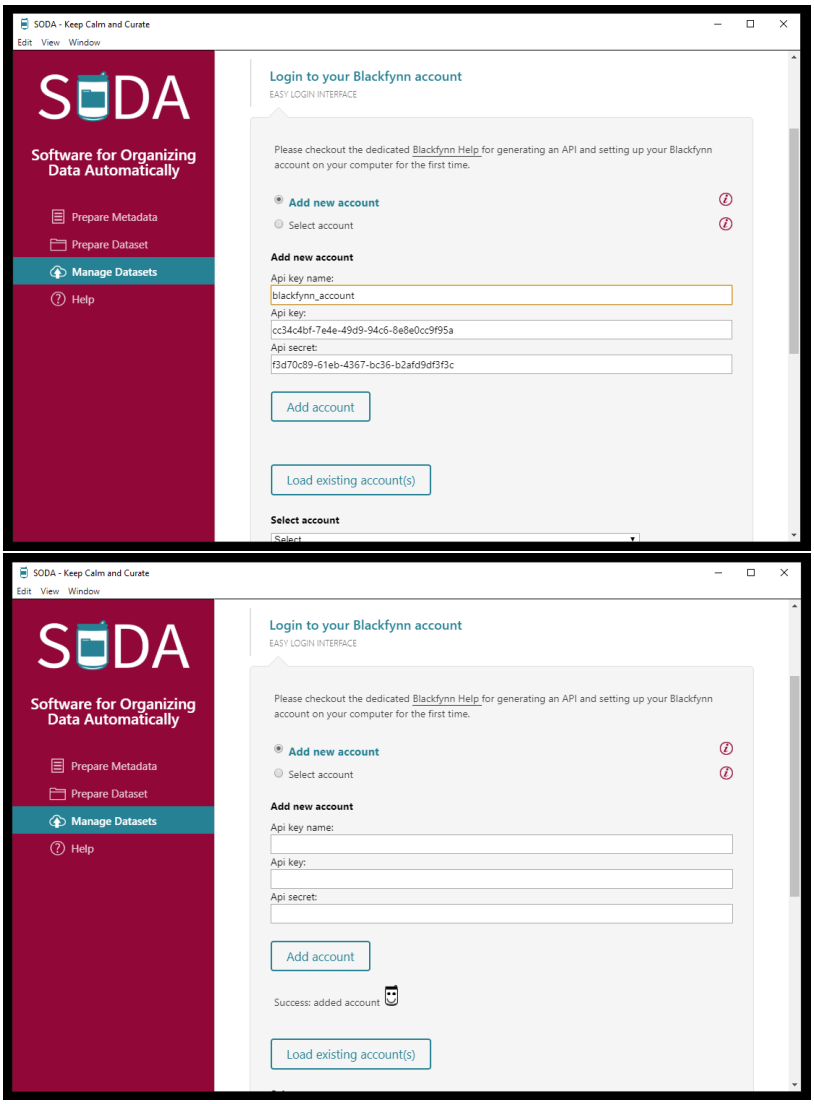
6.1 Login to your Blackfynn acount

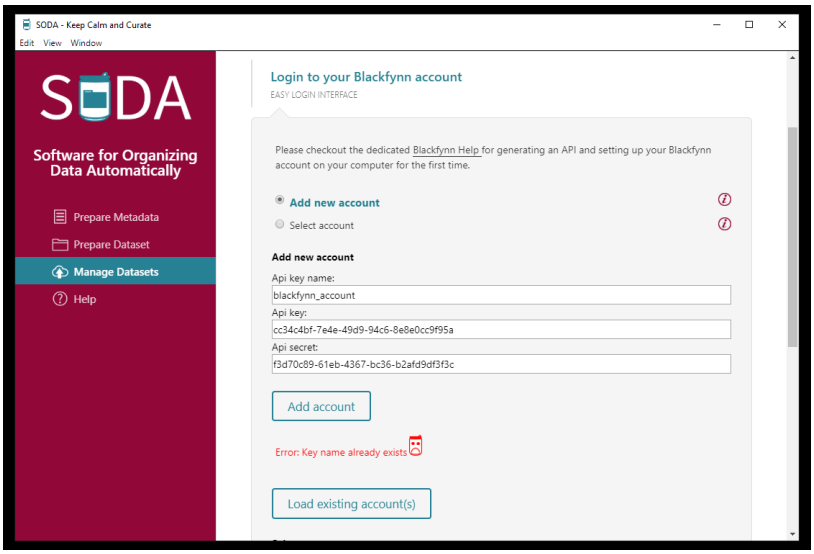
You can easily login to your Blackfynn account through the SODA interface:

- Follow the instruction on the Blackfynn help page to get an key name, API key, and secret.
- Enter them in the corresponding fields in the interface.
- Click "Add account". The adjacent textbox will display a success message or prompt an error message with the nature of the error.
- If successful, the account key name will appear in the dropdown list under "Select account".

Note:

- Selecting an account under "Select account" is necessary to perform subsequent actions.

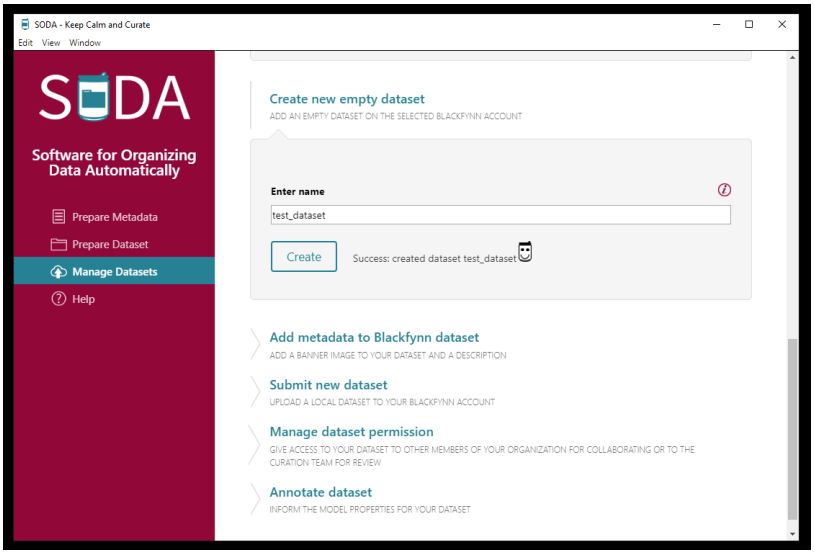

SODA - Keep Calm and Curate
Edit View Window
SODA
Software for Organizing Data Automatically
Prepare Metadata
Prepare Dataset
Manage Datasets
Help
Login to your Blackfynn account
EASY LOGIN INTERFACE
Please checkout the dedicated Blackfynn Help for generating an API and setting up your Blackfynn account on your computer for the first time.
Add new account
Select account
Add new account
Api key name:
blackfynn_account
Api key:
cc34c4bf-7e4e-49d9-94c6-8e8e0cc9f95a
Api secret:
f3d70c89-61eb-4367-bc36-b2afd9df3f3c
Add account
Load existing account(s)
Select account
Select
SODA - Keep Calm and Curate
Edit View Window
SODA
Software for Organizing Data Automatically
Prepare Metadata
Prepare Dataset
Manage Datasets
Help
Login to your Blackfynn account
EASY LOGIN INTERFACE
Please checkout the dedicated Blackfynn Help for generating an API and setting up your Blackfynn account on your computer for the first time.
Add new account
Select account
Add new account
Api key name:
Api key:
Api secret:
Add account
Success: added account
Load existing account(s)



6.2 Create new dataset

To create a new dataset on Blackfynn:

- Enter the desired name for the dataset
- Click "Create"
- A success message will be displayed in the adjacent textbox if the dataset is created. Otherwise, an error message will be displayed with the nature of the error.



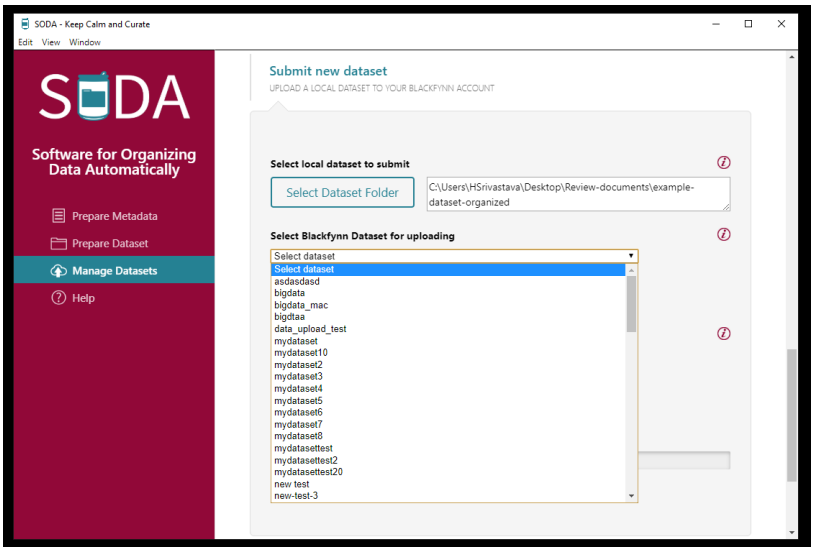
6.3 Add metadata to Blackfynn dataset

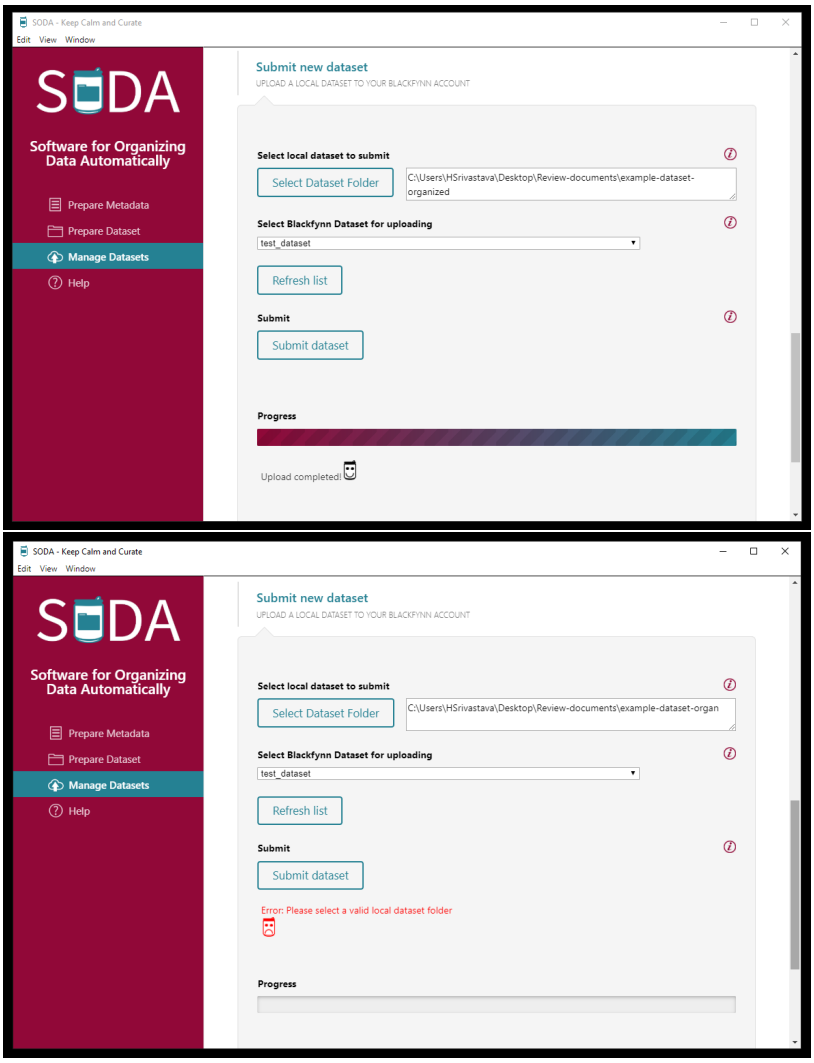
Coming soon

6.4 Submit new dataset

To submit a dataset to Blackfynn:

- Click on "Select Dataset Folder" and navigate to the folder you wish to submit or type the path directly in the adjacent texbox
- "Select Blackfynn Dataset for uploading" lets you select the dataset on Blackfynn, to which, your newly organized dataset will be uploaded
- Click on "Submit dataset". The content of your local dataset will be uploaded to the dataset on Blackfynn selected under "Select Blackfynn Dataset for uploading"
- The progress bar will inform about the progress of the files being uploaded until a success message indicates completion or prompt an error message stating the nature of the error

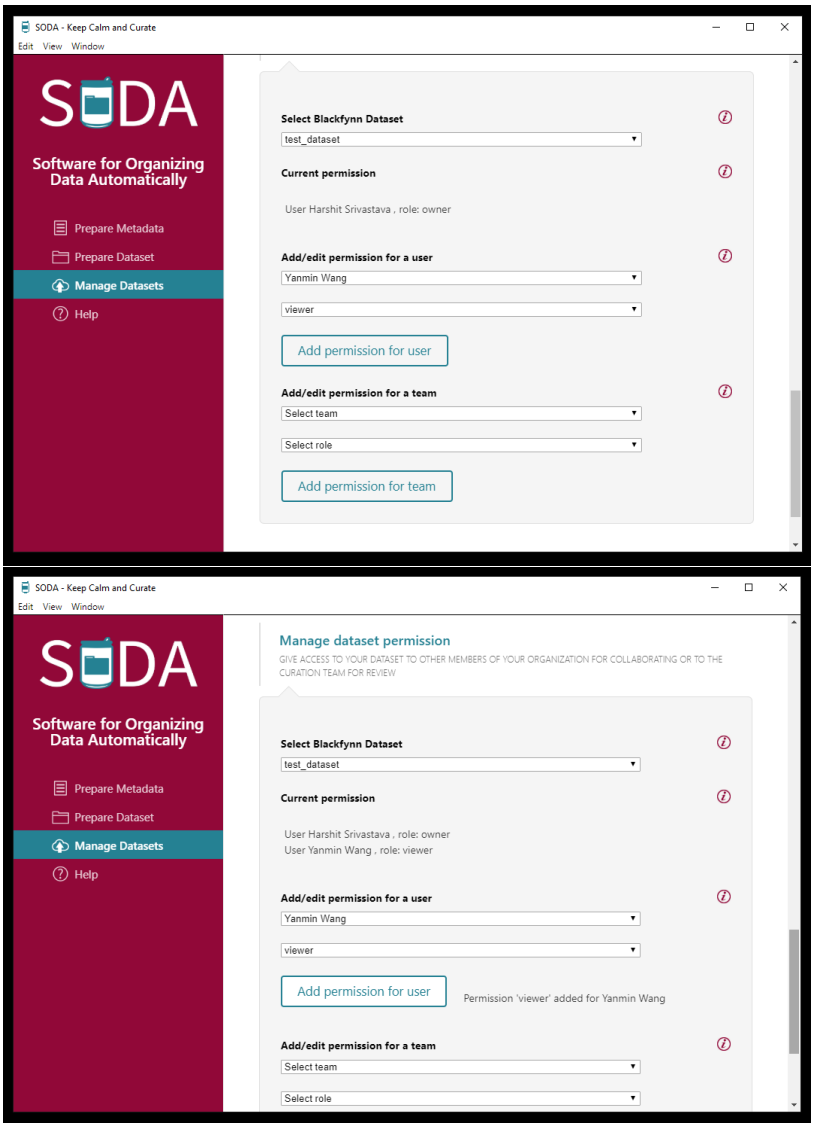




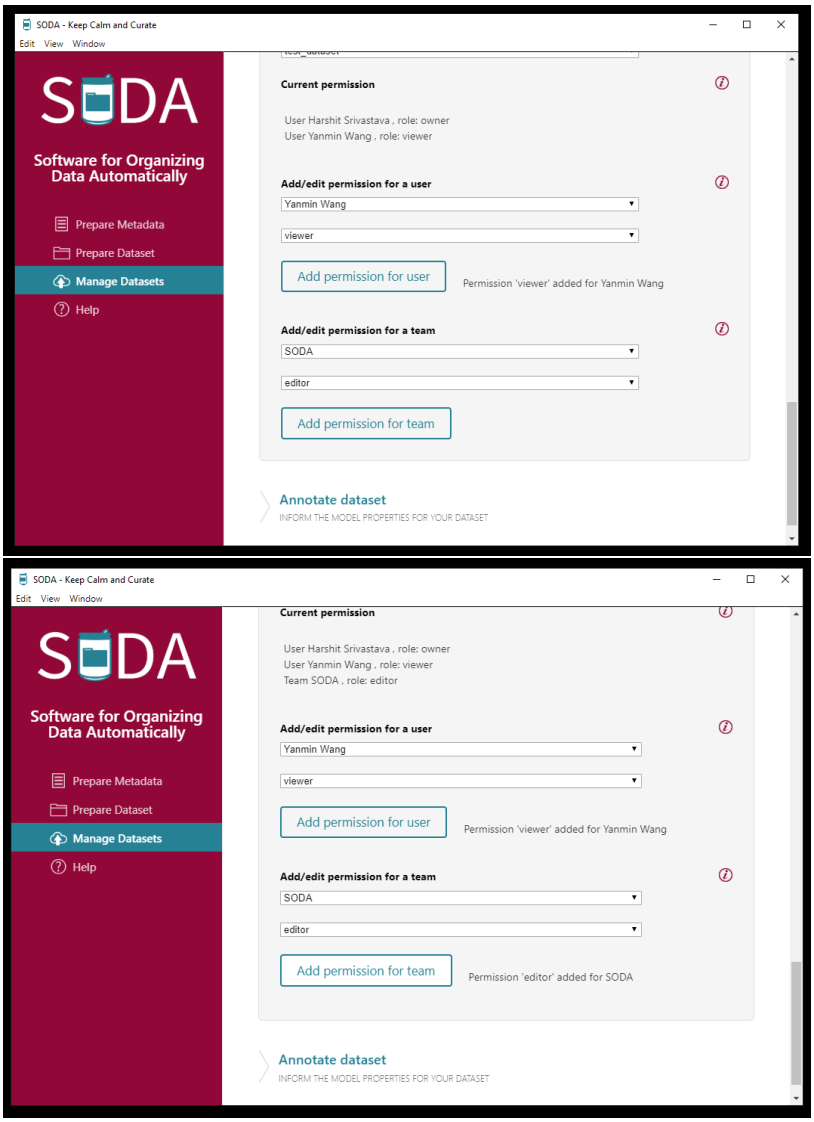
6.5 Manage dataset permission

The permission of a dataset can be changed through this feature (e.g. make PI of a SPARC award "owner" of a dataset as required per SPARC rules or add "SPARC Curation Team" as "manager" to share dataset with the Curation Team for review).

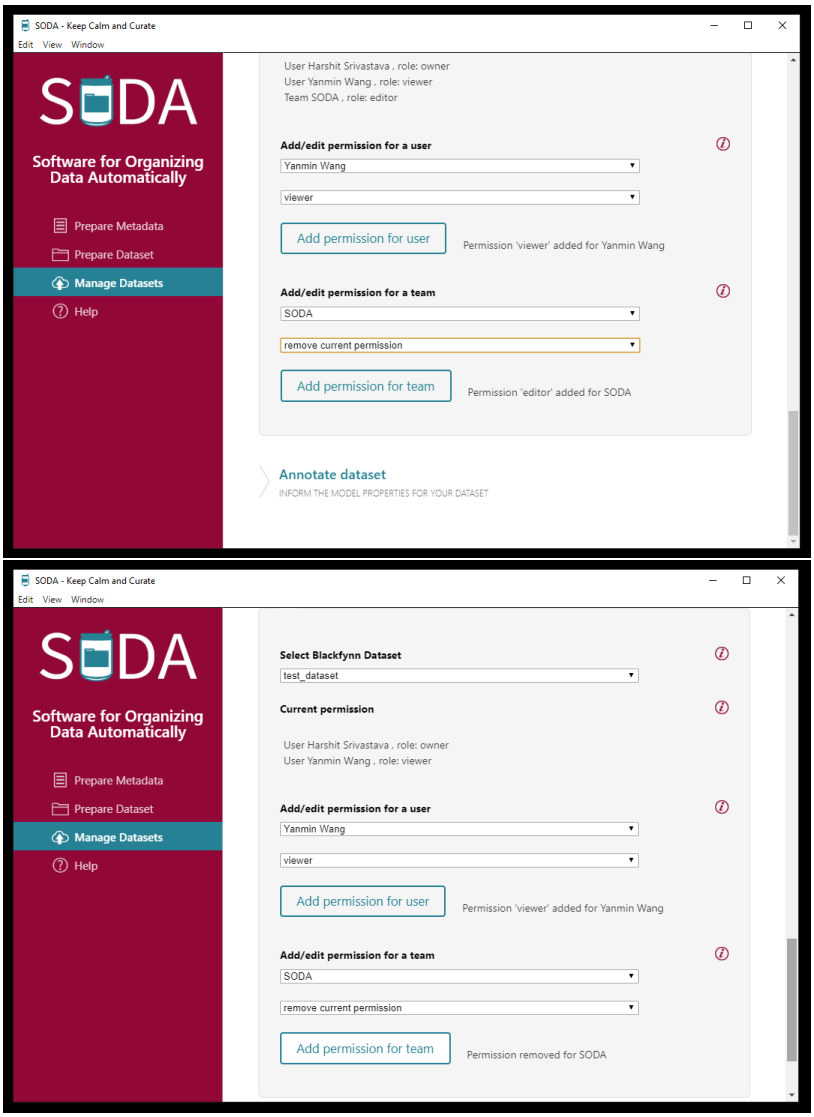
- Select a dataset on your Blackfynn account from the drop-down list
- The current permission on the dataset is shown under "Current permission"
- Follow subsequent indications in the interface to add/remove permission for a user (member of your organization) or a team


SODA - Keep Calm and Curate
Edit View Window
SODA
Software for Organizing Data Automatically
Prepare Metadata
Prepare Dataset
Manage Datasets
Help
Select Blackfynn Dataset
test_dataset
Current permission
User Harshit Srivastava , role: owner
Add/edit permission for a user
Yanmin Wang
viewer
Add permission for user
Add/edit permission for a team
Select team
Select role
Add permission for team

SODA - Keep Calm and Curate
Edit View Window
SODA
Software for Organizing Data Automatically
Prepare Metadata
Prepare Dataset
Manage Datasets
Help
Manage dataset permission
GIVE ACCESS TO YOUR DATASET TO OTHER MEMBERS OF YOUR ORGANIZATION FOR COLLABORATING OR TO THE CURATION TEAM FOR REVIEW
Select Blackfynn Dataset
test_dataset
Current permission
User Harshit Srivastava , role: owner
User Yanmin Wang , role: viewer
Add/edit permission for a user
Yanmin Wang
viewer
Add permission for user
Permission 'viewer' added for Yanmin Wang
Add/edit permission for a team
Select team
Select role


SODA - Keep Calm and Curate
Edit View Window
SODA
Software for Organizing Data Automatically
Prepare Metadata
Prepare Dataset
Manage Datasets
Help
Current permission
User Harshit Srivastava , role: owner
User Yanmin Wang , role: viewer
Add/edit permission for a user
Yanmin Wang
viewer
Add permission for user
Permission 'viewer' added for Yanmin Wang
Add/edit permission for a team
SODA
editor
Add permission for team
Annotate dataset
INFORM THE MODEL PROPERTIES FOR YOUR DATASET

SODA - Keep Calm and Curate
Edit View Window
SODA
Software for Organizing Data Automatically
Prepare Metadata
Prepare Dataset
Manage Datasets
Help
Current permission
User Harshit Srivastava , role: owner
User Yanmin Wang , role: viewer
Team SODA , role: editor
Add/edit permission for a user
Yanmin Wang
viewer
Add permission for user
Permission 'viewer' added for Yanmin Wang
Add/edit permission for a team
SODA
editor
Add permission for team
Permission 'editor' added for SODA
Annotate dataset
INFORM THE MODEL PROPERTIES FOR YOUR DATASET

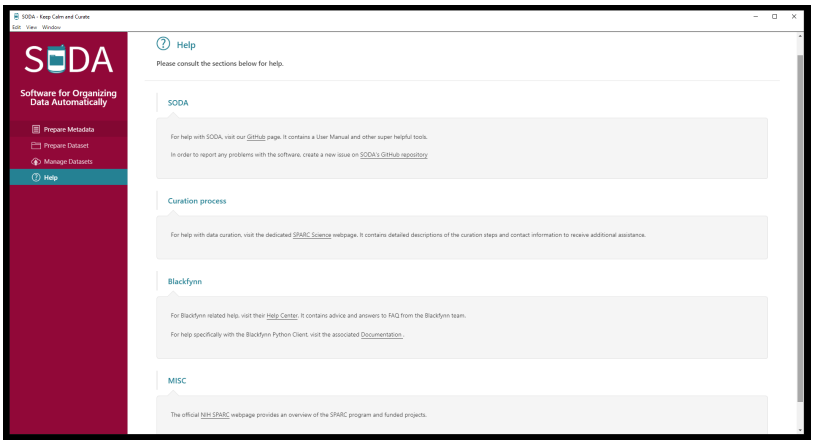


6.6 Annotate dataset

Coming soon... tentative date: April 2020

7 Help

The help section contains links to various resources for assistance with SODA, the dataset preparation process, Blackfynn, and other.



For reporting any problems/bugs with the software, please open a new issue on the https://github.com/bvhpate
Provide adequate information (operating system, steps leading to error, screenshots) so we can help you efficiently.