

SODA User manual

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

July 2019



Contents

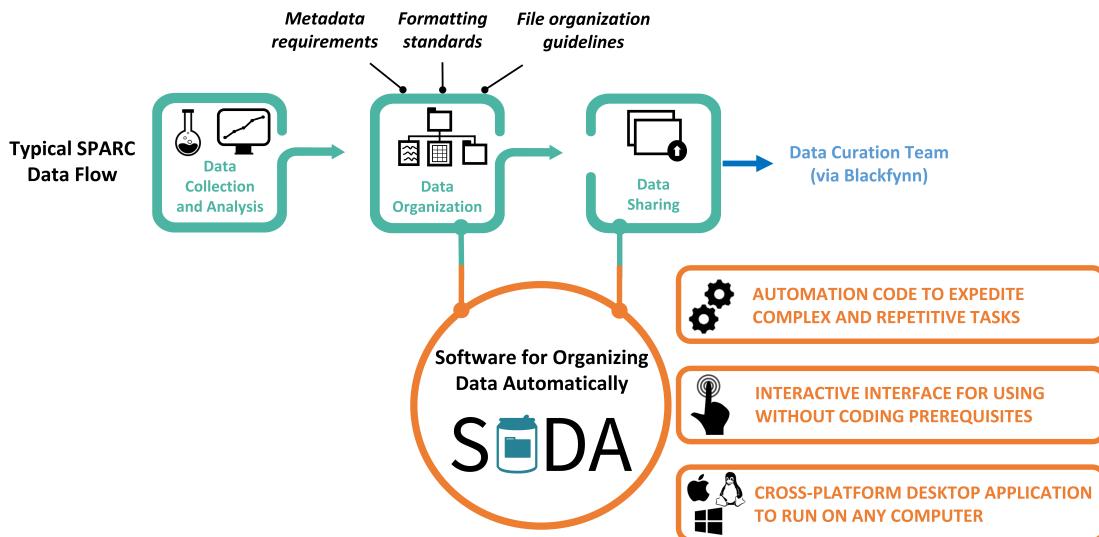
	Page
1 What is SODA?	2
2 Download and Open the Application	3
3 Main Interface	5
4 Prepare Dataset	6
4.1 Organize	7
4.1.1 Dataset already organized	7
4.1.2 Organize my dataset	8
4.2 Convert file format	15
4.3 Generate metadata	15
4.4 Validate	17
4.5 Prepare dataset	17
4.5.1 Modify current dataset	17
4.5.2 Create new dataset locally	18
4.5.3 Create new dataset directly on Blackfynn	19
4.5.4 Progress window	19
5 Manage and Submit Datasets	20
5.1 Login to your Blackfynn account	20
5.2 Create new dataset folder	21
5.3 Submit new dataset	22
5.4 Share a dataset with the curation team	24
5.5 Annotate dataset	24
6 Help	24

1 What is SODA?

Once a milestone of a project is reached, SPARC investigators need to organize and curate 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, which vary for each type of dataset. 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 process through interactive tools and automation. If at any point during your data organization process you think: "Ah, if only this step could be simplified", it actually could be with SODA!

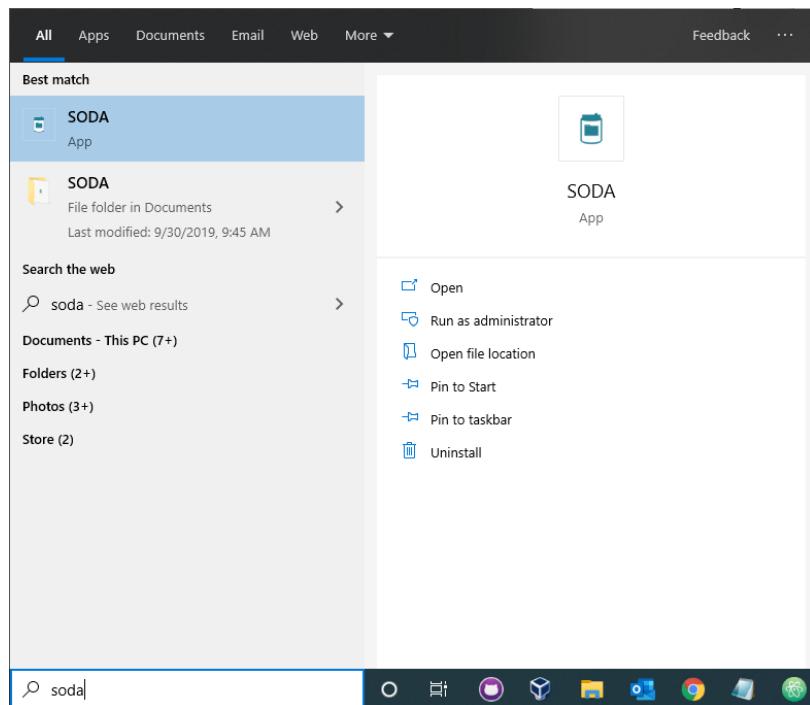
SODA is distributed as an easy to install application for Windows, Mac, Linux platforms. The front-end (Graphical User Interface or GUI) of SODA is built with Electron, an open-source framework developed and maintained by GitHub that conveniently combines HTML, CSS, and Javascript, while the back-end is developed in Python (v3.6). SODA is distributed as an easy to install application for Windows, Mac, Linux platforms. 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

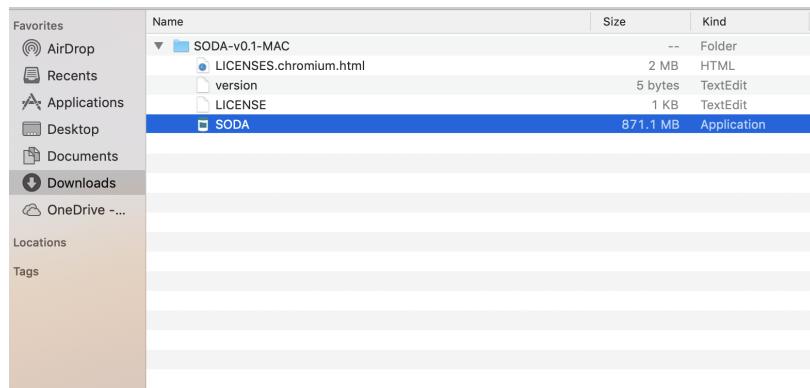
To use SODA, follow the instructions below for your specific platform:

- Windows (developed on Windows 10)
 1. Click here and download the "SODA-v0.2-Windows.exe" file
 2. Double click on the executable, it will start the installation process
 3. Select the folder where you want to extract the files
 4. You can chose the option of having a shortcut at the Desktop or Quick Launch
 5. Once the application is installed, you can run the application by either searching for "SODA" on the start menu, running the desktop shortcut or from the folder where all files got extracted
 6. To stop the application, simply click on the cross in the top right corner.



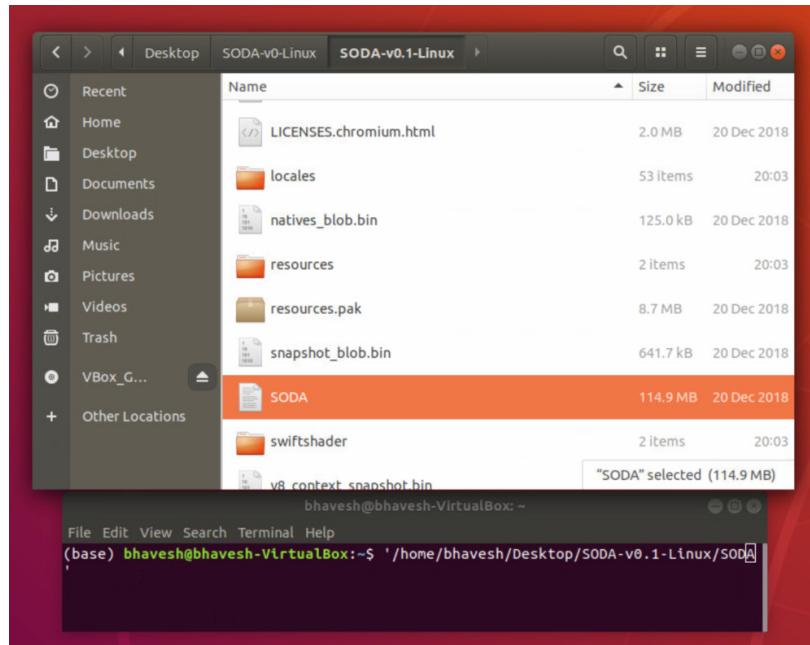
- MAC-OS (developed on MAC 10.14)
 1. Click here and download the "SODA-v0.2-MAC.zip" file
 2. Unzip the file if necessary (e.g., double click to extract with with "Archive Utility" tool)
 3. Open the unzipped folder "SODA-v0.2-MAC"

4. Click on "SODA.app" to launch the application (right click and select "Open" if opening is prevented). Security preferences may need to be changed (System Preferences>Security Privacy>General Tab>"Open anyway").
5. To stop the application, click on the red circle/button in the top left corner. You may also need to right click on the SODA icon in the Dock and select "Quit".
6. The file could be uninstalled conveniently by removing



Favorites	Name	Size	Kind
AirDrop	SODA-v0.1-MAC	--	Folder
Recents	LICENSES.chromium.html	2 MB	HTML
Applications	version	5 bytes	TextEdit
Desktop	LICENSE	1 KB	TextEdit
Documents	SODA	871.1 MB	Application
Downloads			
OneDrive - ...			
Locations			
Tags			

- Linux (developed on Ubuntu 18.04.2 LTS)
 1. Click here and download the "SODA-v0.2-Linux.zip" file
 2. Unzip the file if necessary
 3. Run the "SODA" file located in the unzipped folder "SODA-v0.2-Linux" from the Terminal
 4. To stop the application close the main interface

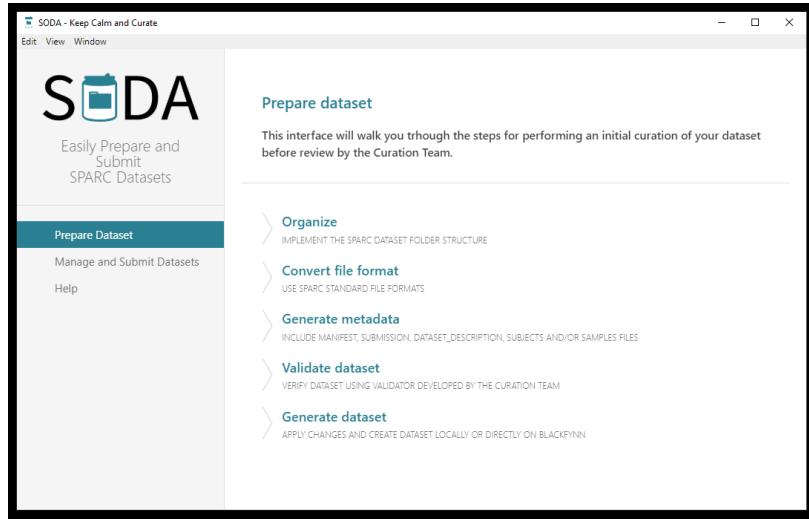


Note:

- The ".zip" folder can subsequently be deleted since it is not required to use SODA.
- Screenshots provided 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 two main features of SODA, "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 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 the app and re-open it.
- There are helpful tooltips (?) spread across the UI which provide some information about the features.

4 Prepare Dataset

This interface will walk you step by step through the data preparation process and includes the following options that are described in the next sections:

1. Organize
2. Convert file format
3. Generate metadata
4. Validate
5. Prepare dataset

The first four options permit to select desired actions on the dataset which could then be applied under "Prepare dataset".

4.1 Organize

4.1.1 Dataset already organized

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

- Click on "Select Dataset Folder" and select the already organized dataset
- Path of files / folder from that dataset will appear in the table located below

The image shows two side-by-side screenshots of the SODA software interface. Both screenshots feature a left sidebar with the SODA logo and the text "Easily Curate and Submit SPARC Datasets".

The top screenshot is titled "Organize dataset" and contains the following text:
IMPLEMENT THE SPARC DATASET FOLDER STRUCTURE
Dataset already organized
Organize my dataset

A "Select Dataset Folder" button is highlighted with a red box. Below it is a table with two columns: "code" and "description". The "code" column lists directory paths, and the "description" column lists file names. A "Clear table" button is at the bottom.

code	description
derivatives	description
docs	description
protocol	description
samples	description
sourcedata	description
subjects	description
main	description

The bottom screenshot is titled "Prepare dataset" and contains the following text:
Easily Prepare and Submit SPARC Datasets
Prepare dataset
Manage and Submit Datasets
Help

A "Select Dataset Folder" button is highlighted with a red box. Below it is a table with three columns: "code", "description", and "Edit". The "code" column lists directory paths, and the "description" column lists file names. Each row has an "Edit" button next to the "code" entry.

code	description	Edit
C:\Users\HSrivastava\Desktop\example-dataset-organized\code\some-abaqus-input.ip		Edit
C:\Users\HSrivastava\Desktop\example-dataset-organized\code\some-abaqus-model.cae		Edit
C:\Users\HSrivastava\Desktop\example-dataset-organized\code\some-python-code.py		Edit
derivatives	description	
C:\Users\HSrivastava\Desktop\example-dataset-organized\derivatives\random folder		Edit
C:\Users\HSrivastava\Desktop\example-dataset-organized\derivatives\some-derivative-data.csv		Edit
docs	description	
C:\Users\HSrivastava\Desktop\example-dataset-organized\docs\Readme		Edit

You can click on the "Edit" button to add a description for the corresponding file/folder (click on "Save" when you are done). These descriptions will be included in the manifest files if you request them under "Generate metadata".

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

Note:

- The table automatically clears out 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.

4.1.2 Organize my dataset

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

The screenshot shows a user interface for organizing a dataset. At the top, there are two radio button options: "Dataset already organized" and "Organize my dataset". The "Organize my dataset" option is selected. To its right are two help links: "(?)" and "(?)". Below this is a section titled "Upload previously saved file structure" with a "Upload" button. A horizontal line separates this from the main configuration area. The main area starts with a title "Select folders and files to be added in each SPARC folder" and a help link "(?)". It includes a dropdown menu "Select SPARC folder:" set to "code" with a dropdown arrow, and two buttons: "Add files" and "Add folders". Below this is a table for defining file organization:

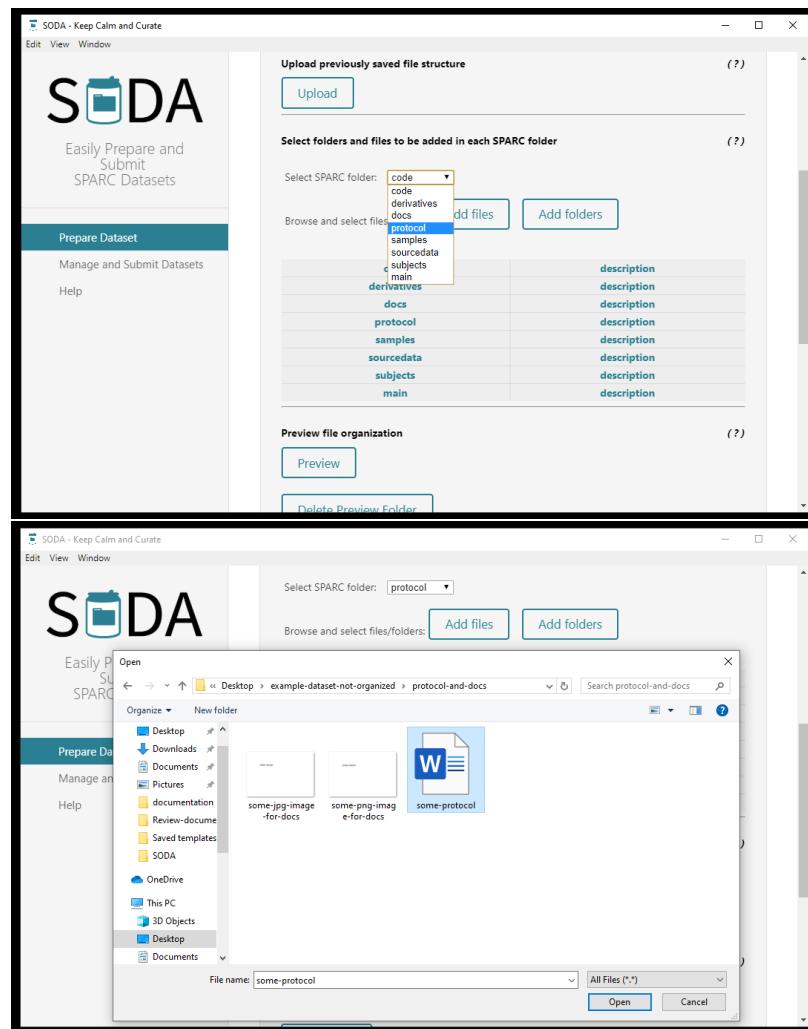
code	description
derivatives	description
docs	description
protocol	description
samples	description
sourcedata	description
subjects	description
main	description

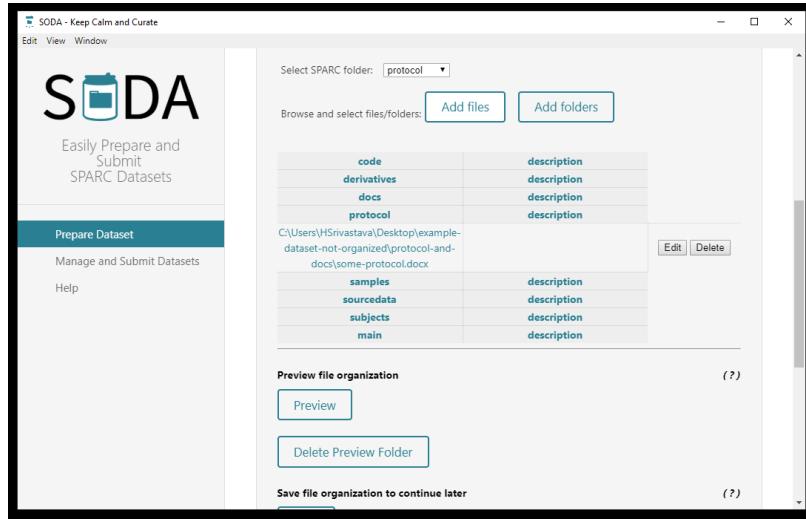
Below the table is a section titled "Preview file organization" with a help link "(?)". It contains a "Preview" button and a "Delete Preview Folder" button. Another horizontal line separates this from the bottom section. The bottom section is titled "Save file organization to continue later" with a help link "(?)". It contains a "Save" button and a "Clear table" button.

There are several ways to do that:

1. Browse and add:

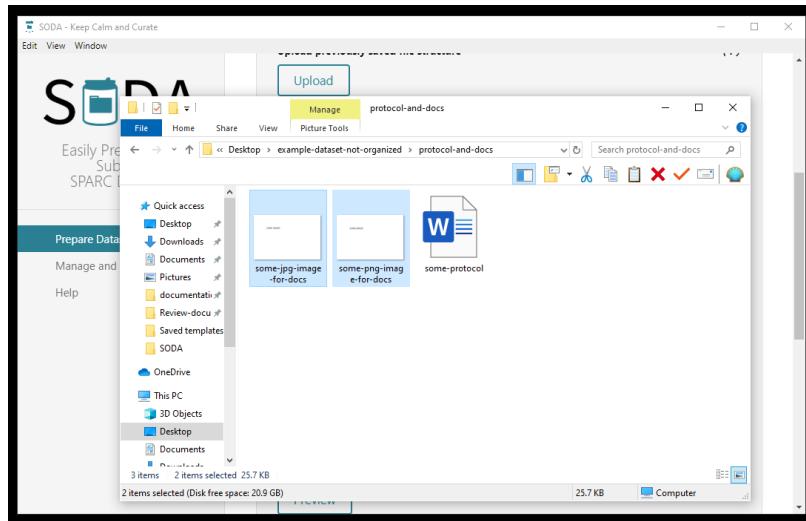
- Select SPARC folder from the drop-down list
- Click on "Add files" or "Add folders" to browse and select
- Path of selected files / folder will appear in the table below the header of the selected SPARC folder





2. Drag and drop:

- Drag any local files/folders to any of the table header ("code", "derivative", etc.)
- Mouse pointer should be inside the box for each folder to correctly capture the path
- Path of dragged files / folder will appear in the table below the corresponding header



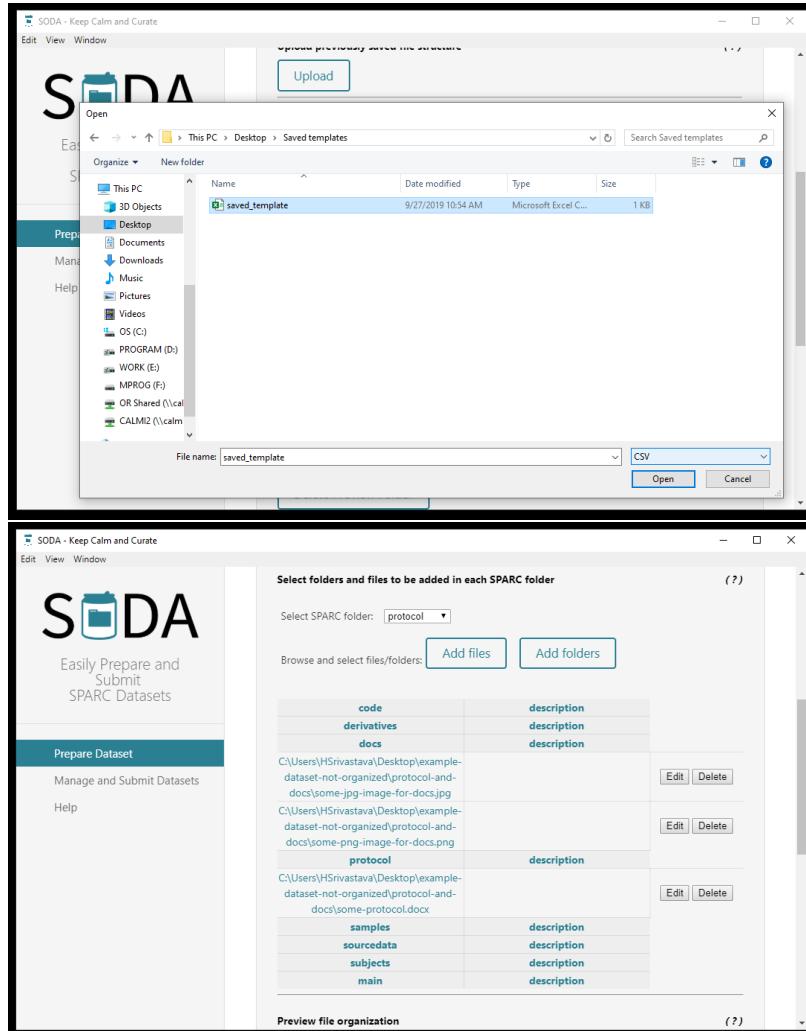
The image consists of two side-by-side screenshots of the SODA software interface. Both screenshots show a left sidebar with the SODA logo and the text "Easily Prepare and Submit SPARC Datasets". Below this is a "Prepare Dataset" section with "Manage and Submit Datasets" and "Help" options.

The top screenshot shows a window titled "Upload" with the sub-instruction "Select folders and files to be added in each SPARC folder". It displays a table with two columns: "code" and "description". The "code" column contains "protocol", "derivatives", and "docs". The "description" column contains "description", "description", and "description" respectively. There are "Edit" and "Delete" buttons for each row. A tooltip "Move" is visible over the "protocol" row. Below the table is a preview section with a "Preview" button.

The bottom screenshot shows a similar window with the same table structure. The "code" column contains "protocol", "samples", "sourcedata", "subjects", and "main". The "description" column contains "description", "description", "description", "description", and "description" respectively. There are "Edit" and "Delete" buttons for each row. Below the table is a preview section with a "Preview" button.

3. Upload previously saved organization:

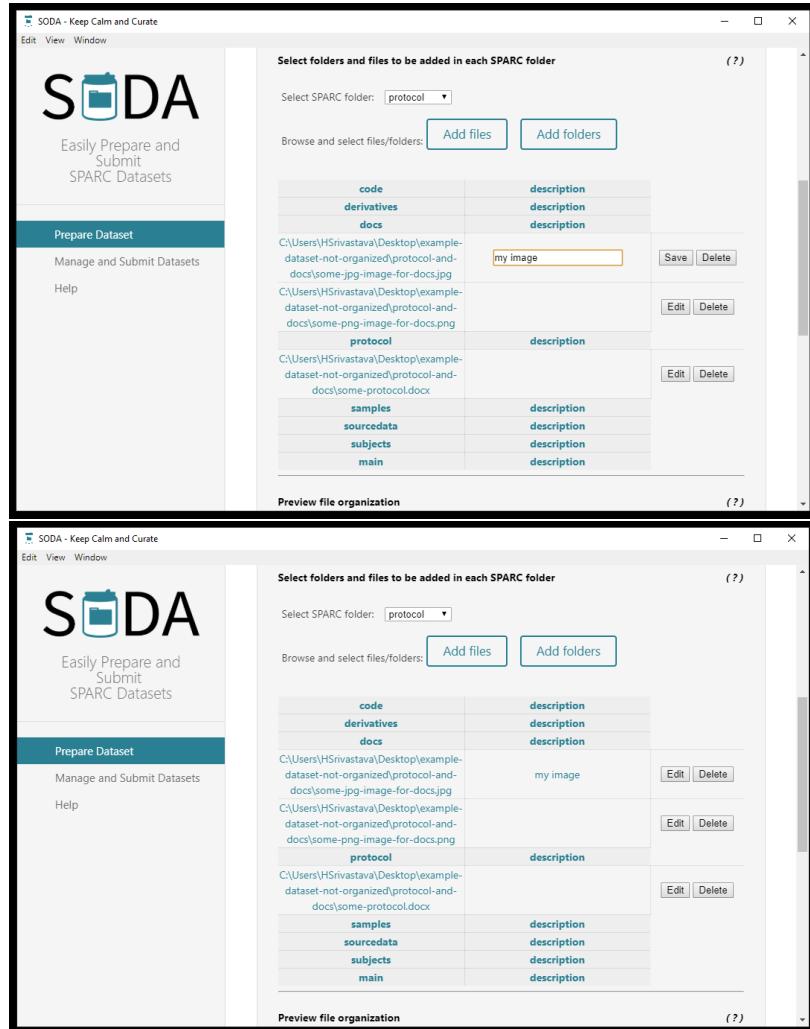
- Click on "Upload"
- Select previously saved table template (see explanation below)
- Paths for each header will be added to the table along with the description (see below)



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

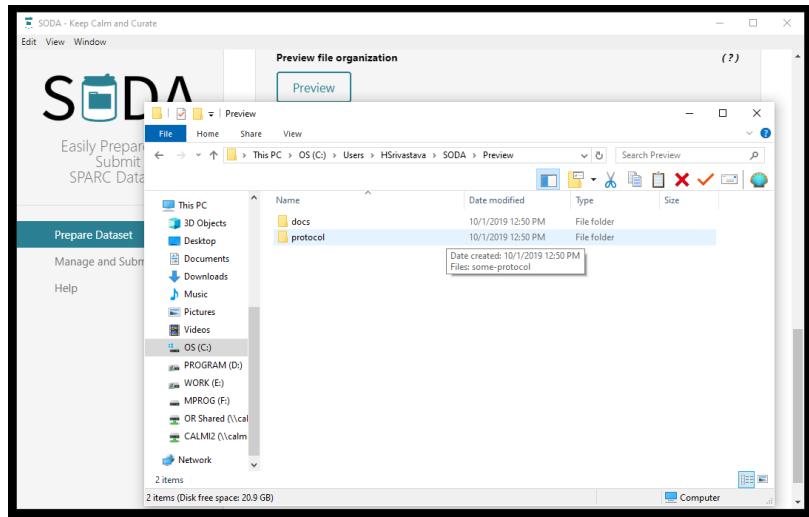
You can add a description for each of the files from the interface which would then be added to the manifest file.

- Click on 'edit' next to the path to add a description
- Enter the description in the textbox
- Click on 'save' to record



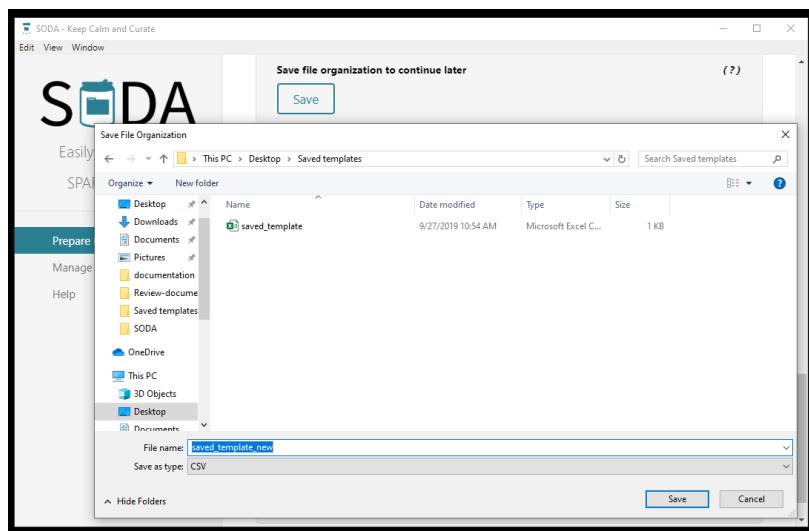
You can conveniently preview file organization before continuing to ensure it's organized as desired before the actual dataset is created.

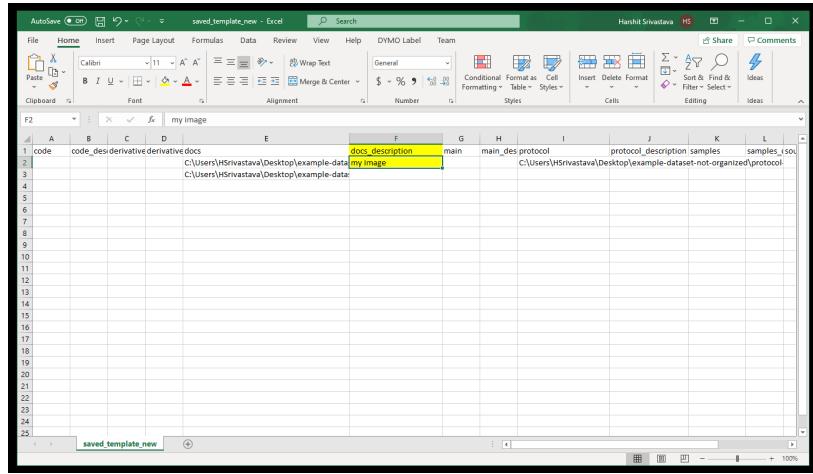
- Click on "Preview" to view the file structure of the dataset you organized
- All the files / folders displayed are mock files which resemble the original files but don't take up any space in the disk
- Click on "Delete Preview Folder" to get rid of the files created for preview



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

- 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 (for eg- MS Excel) to manually include path and description





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.

4.2 Convert file format

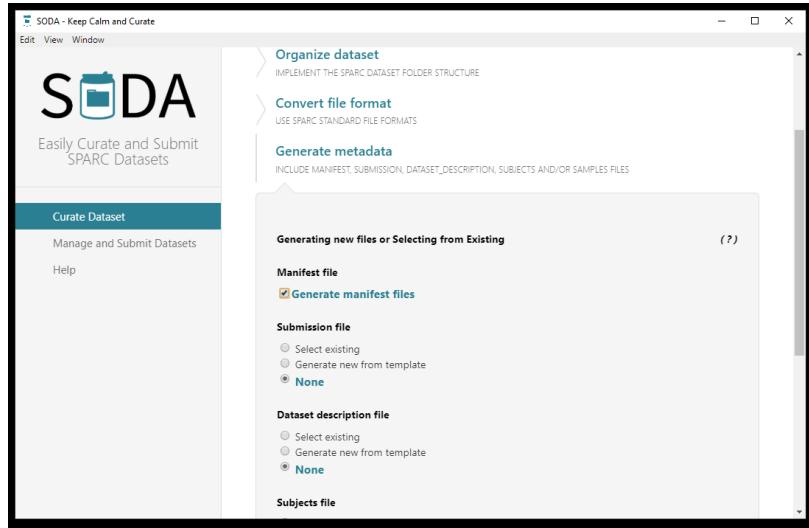
Coming soon

4.3 Generate metadata

You can generate set of required descriptive files (metadata) for your dataset:

- manifest.xlsx within each first level folders of your dataset (it is not required in sub-folders)
 - submission.xlsx
 - dataset_description.xlsx
 - subjects.xlsx
 - samples.xlsx

To generate manifest files, simple check the "Generate manifest files" box. A "manifest.xlsx" file will be added to each folder of the selected dataset with the "filename", "timestamp", and "file type" fields automatically populated.

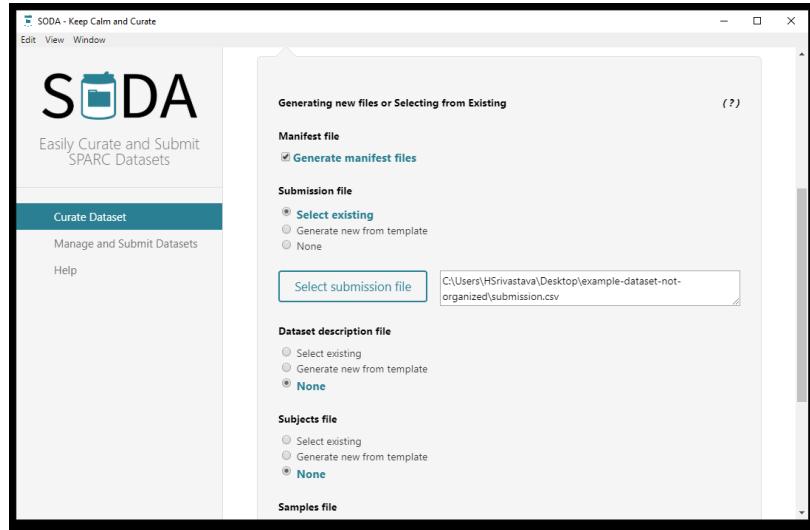


For all other files, there are three options:

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: The file should be saved with the same name as metadata (for eg - only "submission.xlsx" will be accepted for submission file and all other file names will result in an error (to prevent incorrect submission)).

2. "Generate new from template": this option allows to create a file from the 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).



Note:

- By default, the manifest check box is unchecked and all metadata file options are set to "None".
- Only .xlsx, .xls, and .csv file are authorized for all metadata files.

4.4 Validate

Coming soon

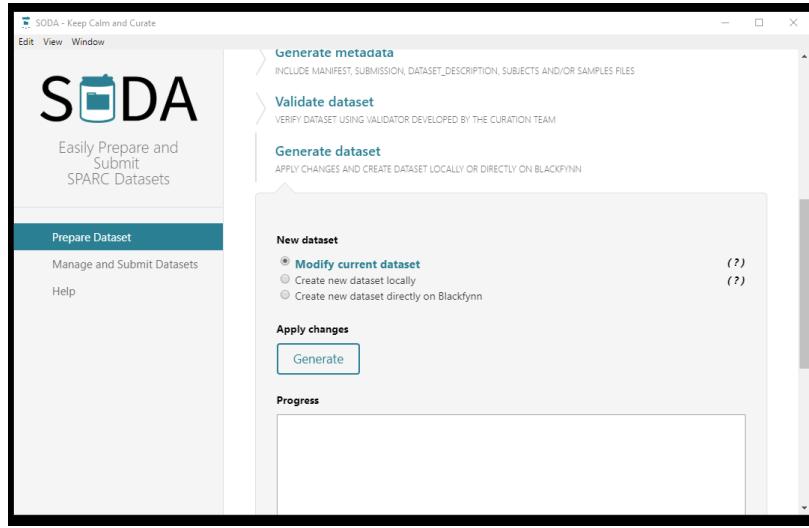
4.5 Prepare dataset

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

4.5.1 Modify current dataset

If "Modify current dataset" is selected, the requested actions will be applied to the current dataset folder (selected under "Organize dataset"). Simply click on "Prepare" to do so.

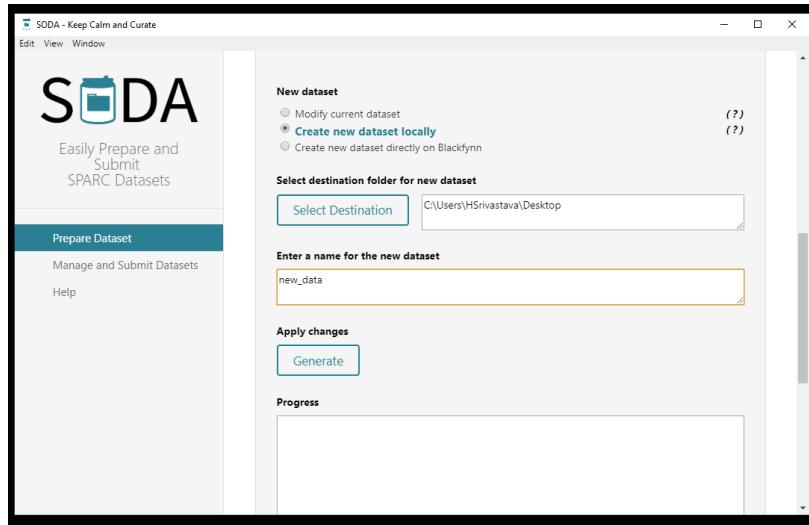
NOTE: This option is only available when you select "Dataset already organized" in the "Organize Dataset" section.



4.5.2 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 with the same name as the current dataset folder 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 folder in which new dataset will be exported
- Click "Prepare" to apply requested actions

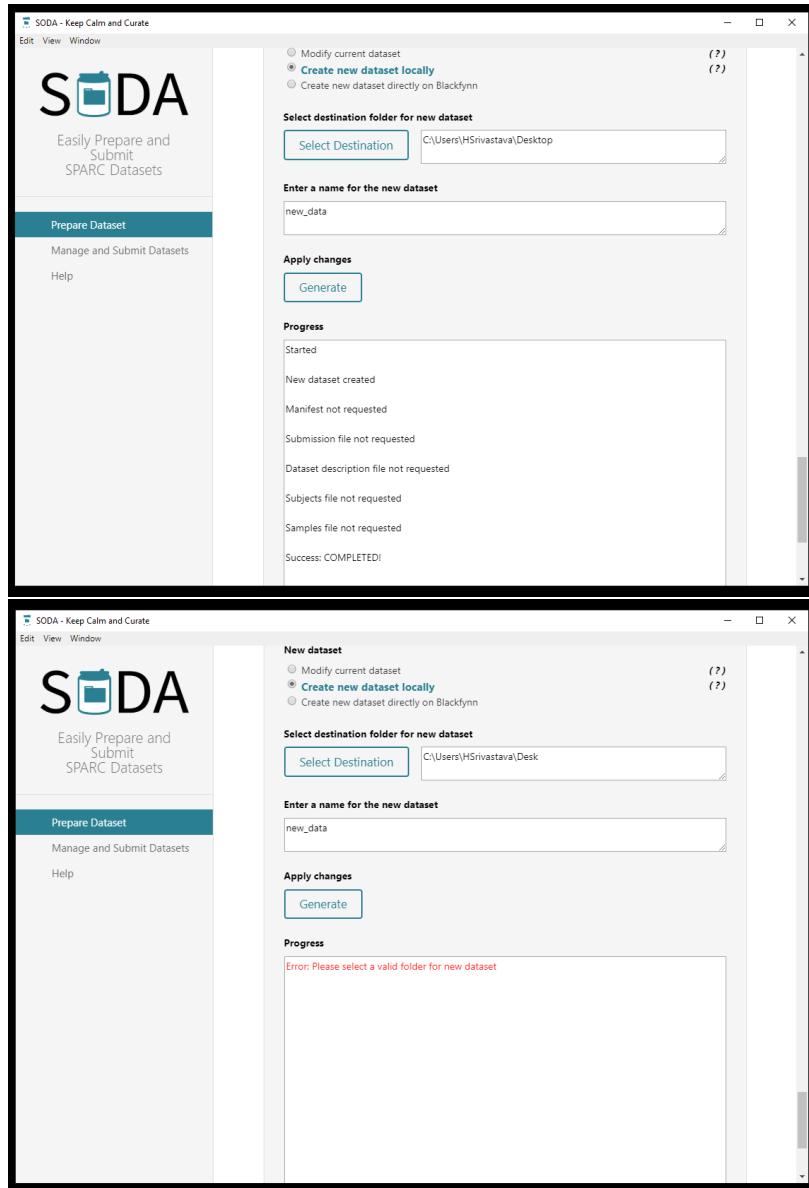


4.5.3 Create new dataset directly on Blackfynn

Coming soon

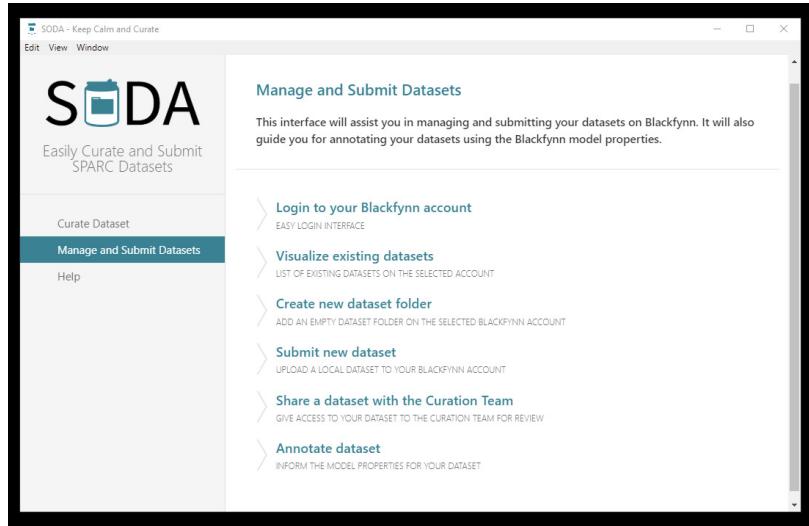
4.5.4 Progress window

For all three options, progress of the preparation process is displayed in the "Progress" box located below the "Prepare" 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.



5 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.

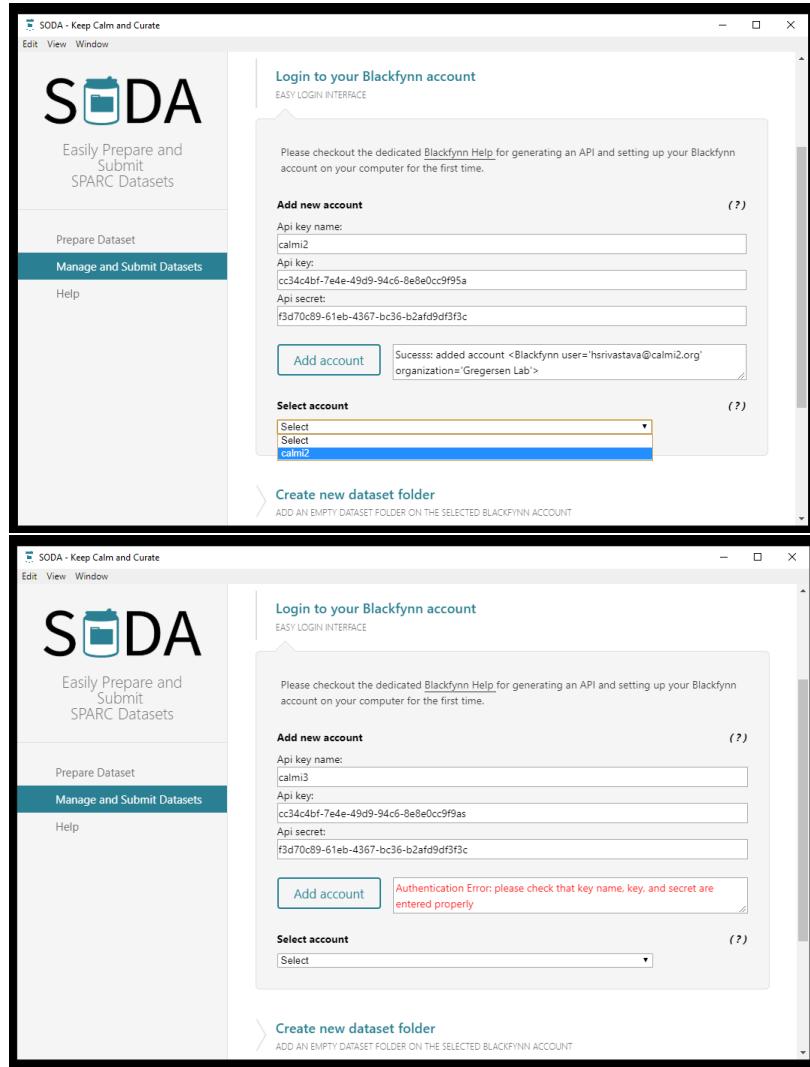
5.1 Login to your Blackfynn account

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

- Follow the instruction on the Blackfynn help page to get a key name, 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:

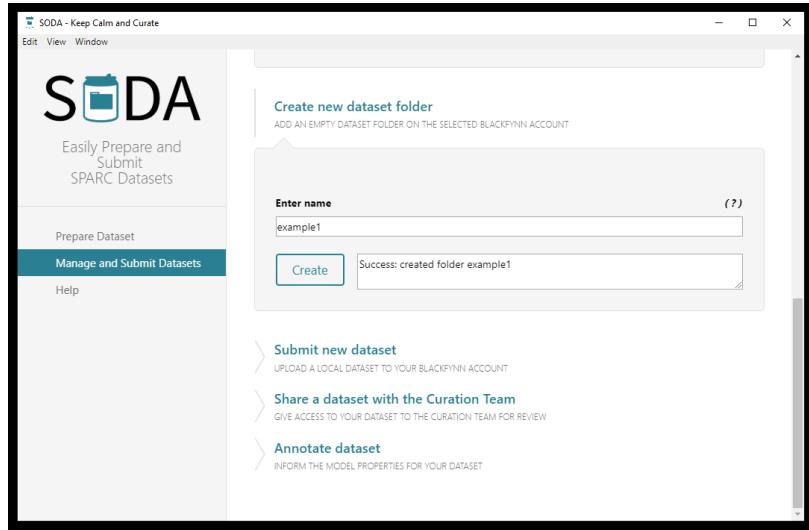
- Selecting an account under "Select account" is necessary to perform subsequent actions.



5.2 Create new dataset folder

To create a new dataset folder:

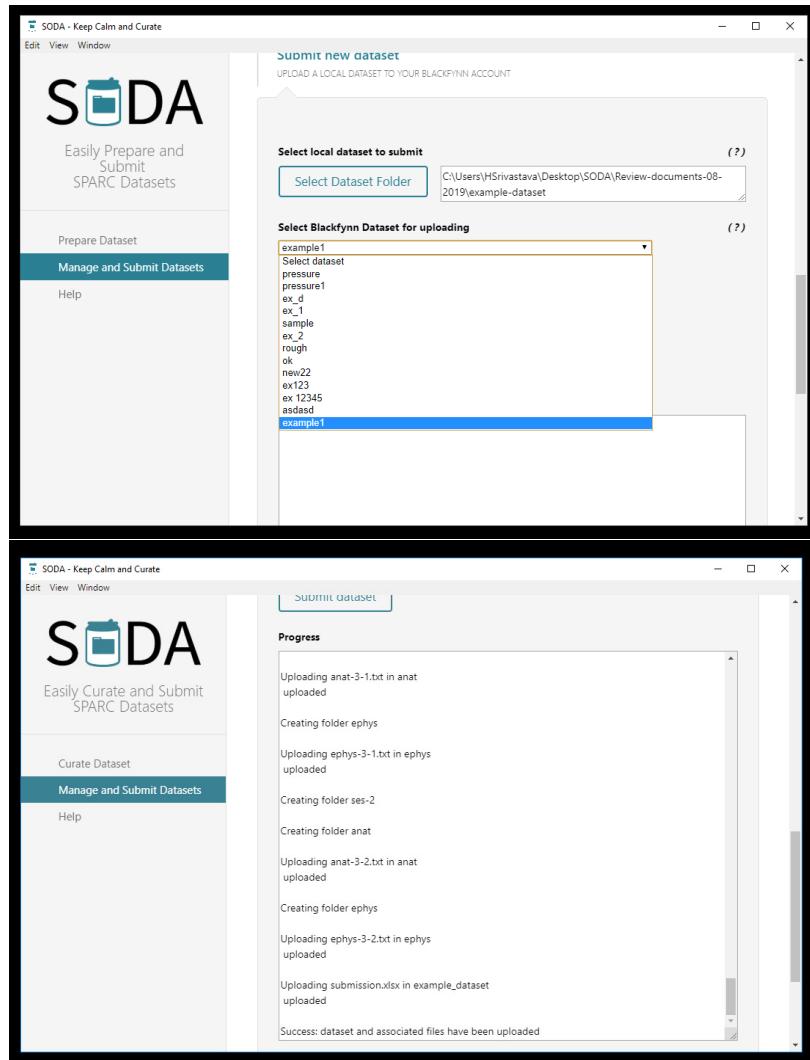
- Enter the desired name for the folder
- Click "Create"
- A success message will be displayed in the adjacent textbox if the dataset is created (it will also be added automatically to the dataset list in section "Submit new dataset" under "Select Blackfynn Dataset for uploading"). Otherwise, an error message will be displayed with the nature of the error.

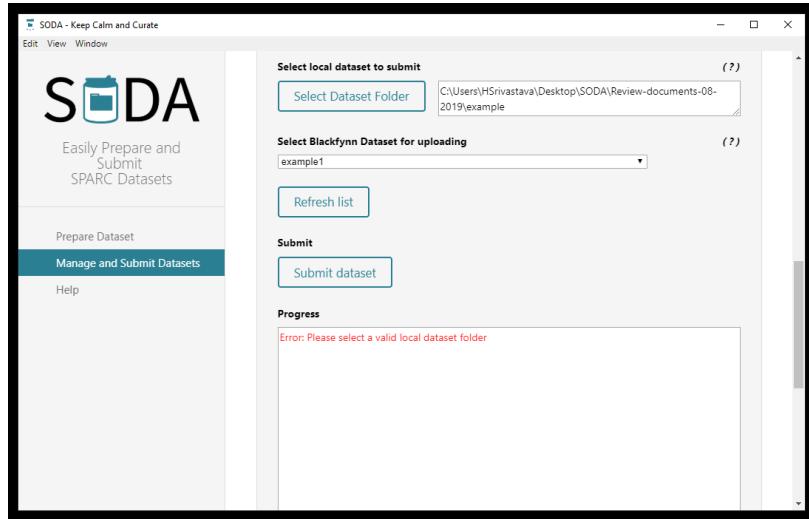


5.3 Submit new dataset

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 textbox
- "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 folder on Blackfynn selected under "Visualize existing datasets"
- The progress box will display the progress of the files being uploaded until a success message indicates completion or prompt an error message stating the nature of the error





5.4 Share a dataset with the curation team

Coming soon

5.5 Annotate dataset

Coming soon

6 Help

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

