

SODA User manual

Bhavesh Patel (bpatel@calmi2.org)
Parya Aghasafari (pagh@safari@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 Dataset	7
4.1 Organize	7
4.1.1 Dataset already organized	7
4.1.2 Organize my dataset	8
4.2 Convert file format	16
4.3 Generate metadata	16
4.4 Validate	18
4.5 Generate dataset	18
4.5.1 Modify current dataset	18
4.5.2 Create new dataset locally	19
4.5.3 Create new dataset directly on Blackfynn	20
4.5.4 Progress window	20
5 Manage and Submit Datasets	21
5.1 Login to your Blackfynn account	21
5.2 Create new dataset folder	22
5.3 Submit new dataset	23
5.4 Share a dataset with the curation team	25
5.5 Annotate dataset	25
6 Help	25

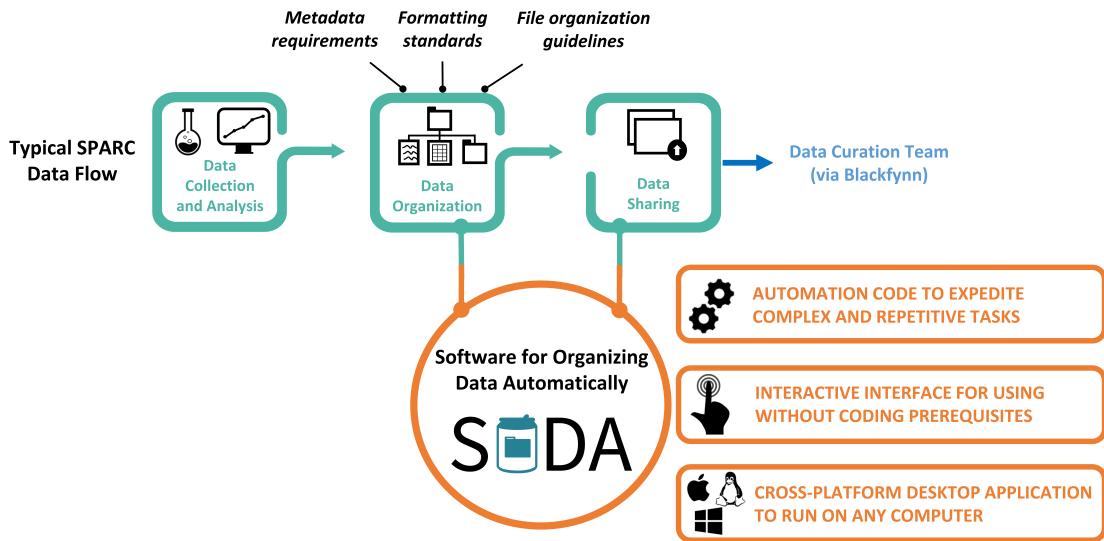
1 What is SODA?

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! Especially, SODA would allow user 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 same validator used by the Curation Team
- Upload organized dataset directly on Blackfynn to avoid duplicating files locally

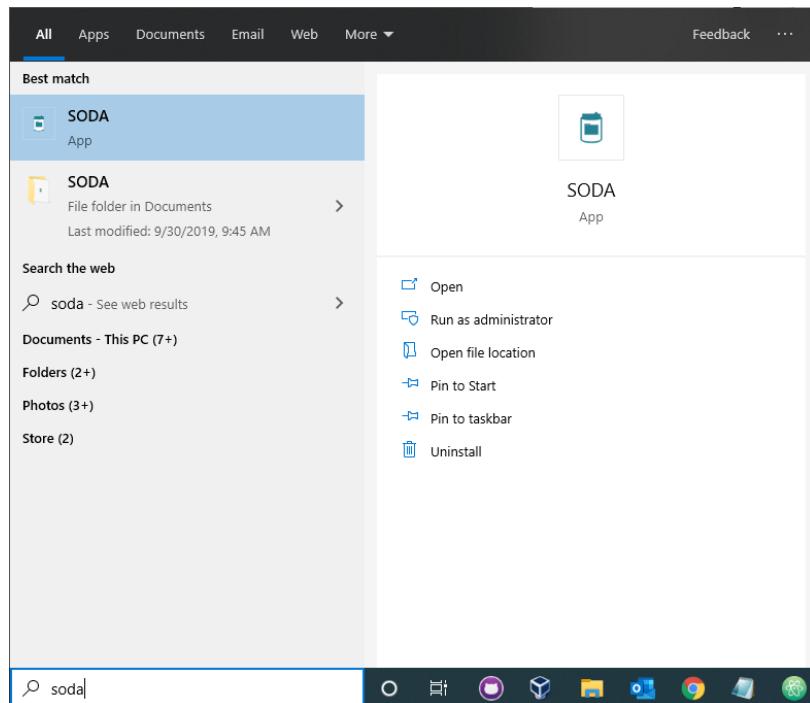
SODA is distributed as an easy to install application for Windows, Mac, Linux platforms. The front-end (Graphical User Interface or GUI) of SODA is built with Electron, an open-source framework developed and maintained by GitHub that conveniently combines HTML, CSS, and Javascript, while the back-end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



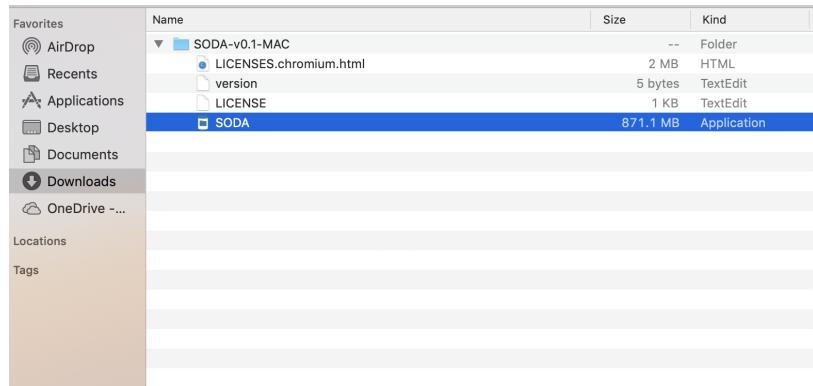
2 Download and Open the Application

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

- Windows (developed on Windows 10)
 1. Download the "SODA-v0.2-Windows.exe" file: One Drive / Google Drive
 2. Double click on the executable, it will start the installation process (if there is a permission issue do right click and select "Run as administrator").
 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.
 7. The downloaded exe 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.

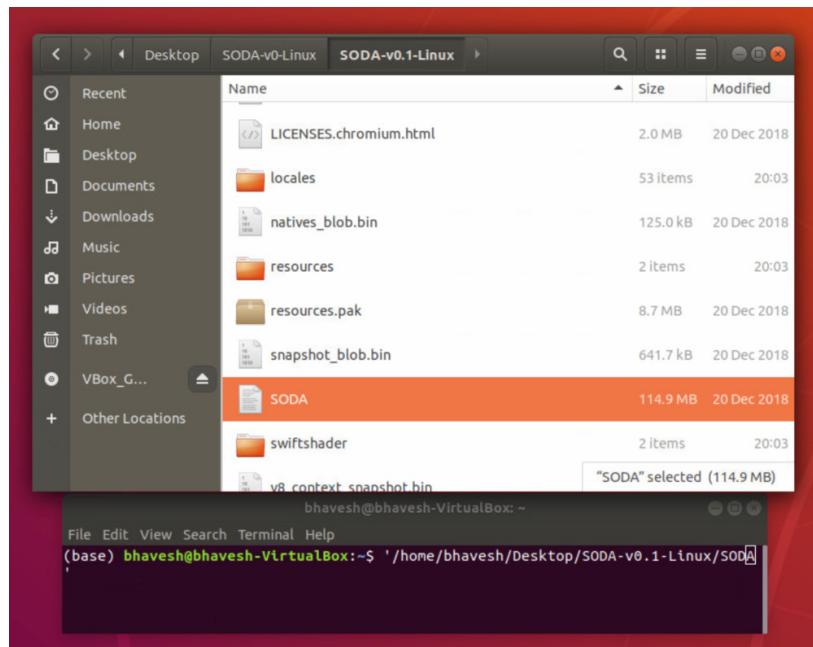


- MAC-OS (developed on MAC 10.14)
 1. Download the "SODA-v0.2-MAC.zip" file: One Drive / Google Drive
 2. Unzip the file if necessary (e.g., double click to extract with "Archive Utility" tool)
 3. Open the unzipped folder "SODA-v0.2-MAC"
 4. Click on "SODA.app" to launch the application (right click 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 downloaded zip file could be deleted since it is not necessary to run SODA. To remove the SODA app, simply delete the unzipped folder.



- Linux (developed on Ubuntu 18.04.2 LTS)

1. Download the "SODA-v0.2-Linux.zip" file: One Drive / Google Drive
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
5. The downloaded zip file could be deleted since it is not necessary to run SODA. To remove the SODA app, simply delete the unzipped folder.

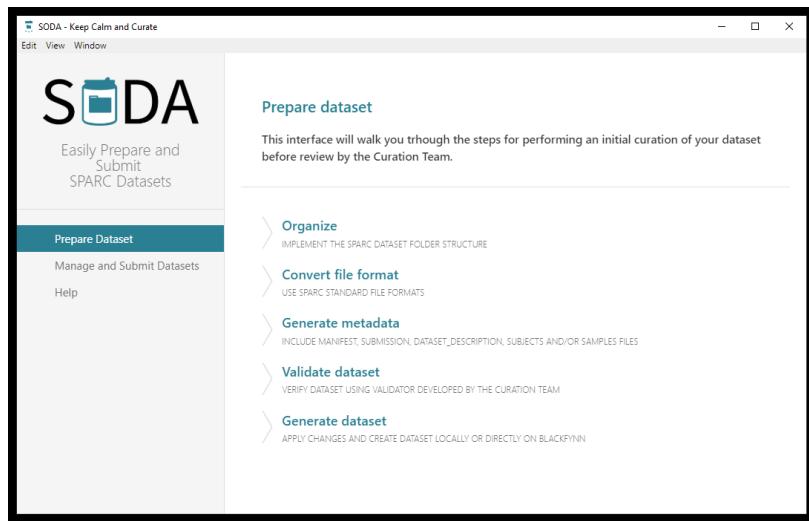


Note:

- The ".zip" folder can subsequently be deleted since it is not required to use SODA.
- 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 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. Generate dataset

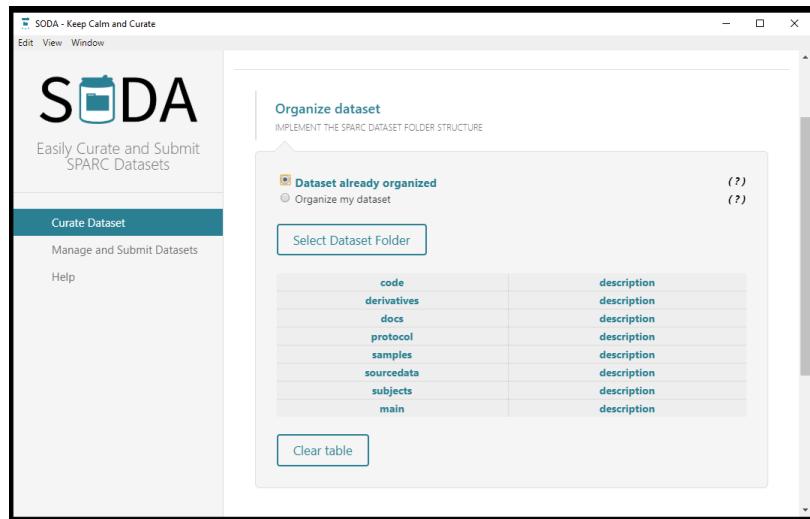
The first four options permit to select desired actions on the dataset which could then be applied under Generate dataset".

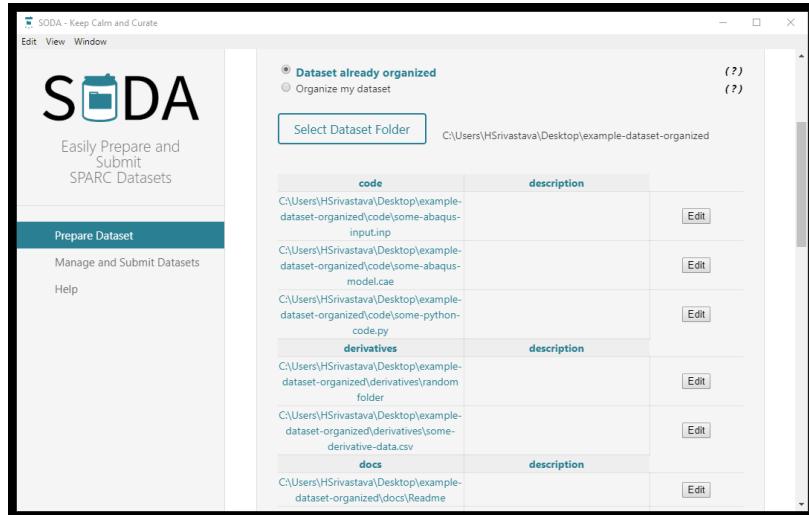
4.1 Organize

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", browse files, and select the already organized dataset
- Path of files/folders from that dataset will appear in the table located below





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 include in your dataset within each of the SPARC folders.

Dataset already organized (?)
Organize my dataset (?)

Upload previously saved file structure (?)

Upload

Select folders and files to be added in each SPARC folder (?)

Select SPARC folder: **code** ▾

Browse and select files/folders: **Add files** **Add folders**

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

Preview file organization (?)

Preview

Delete Preview Folder

Save file organization to continue later (?)

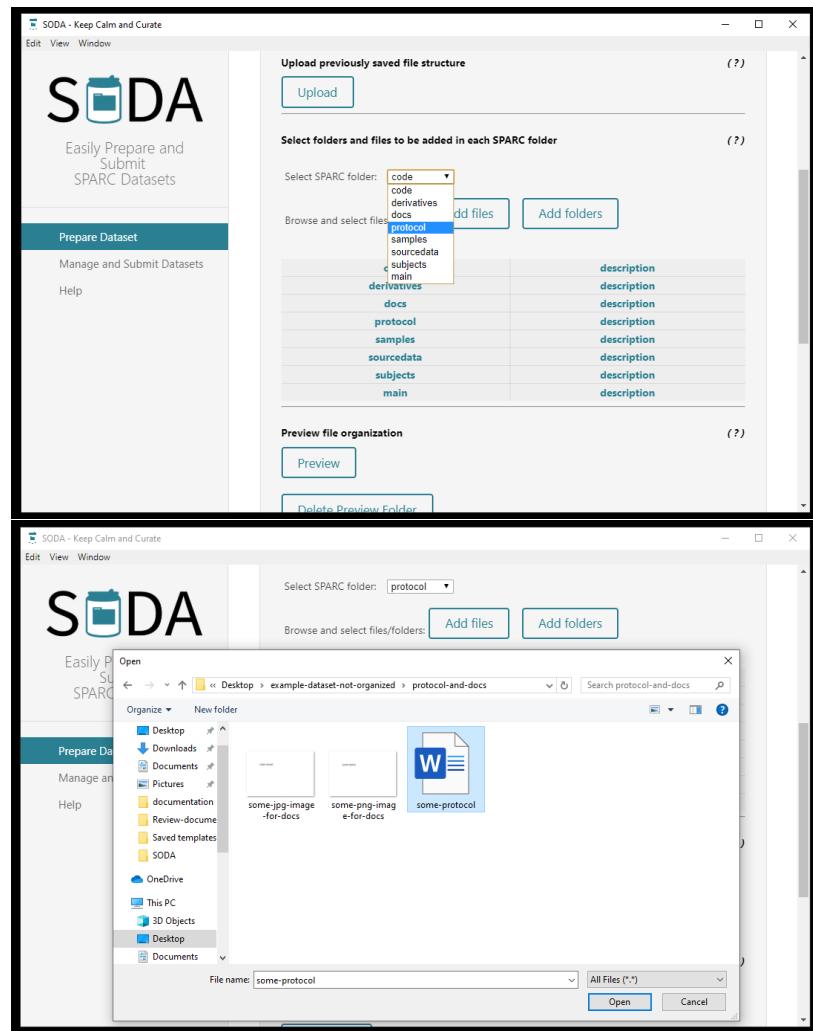
Save

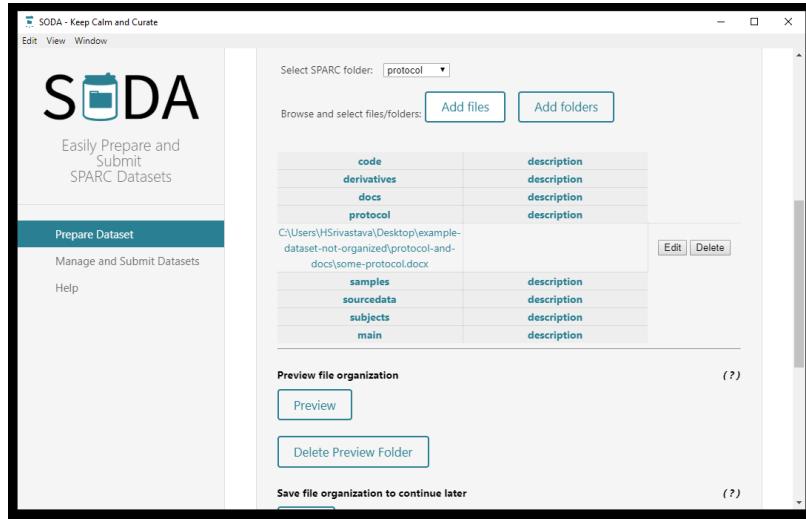
Clear table

There are several ways to do that:

1. Browse and add:

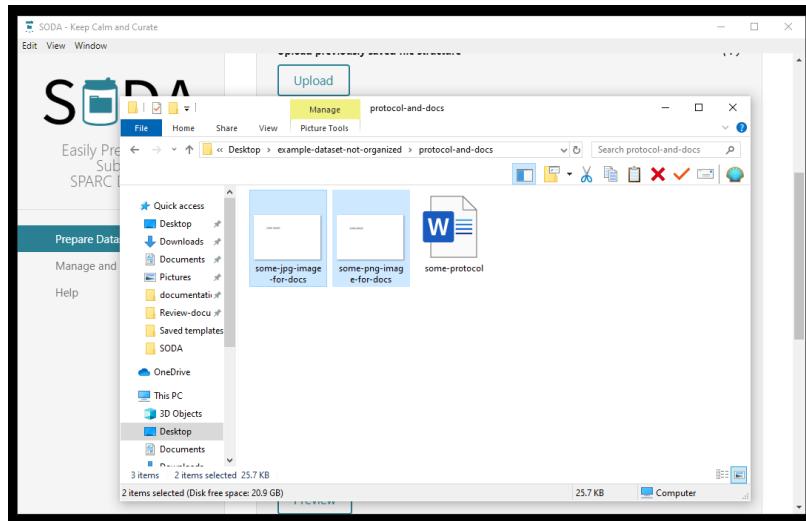
- Select SPARC folder from the drop-down list
- Click on "Add files" or "Add folders" to browse and select
- Path of selected files/folder will appear in the table, below the header of the selected SPARC folder





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

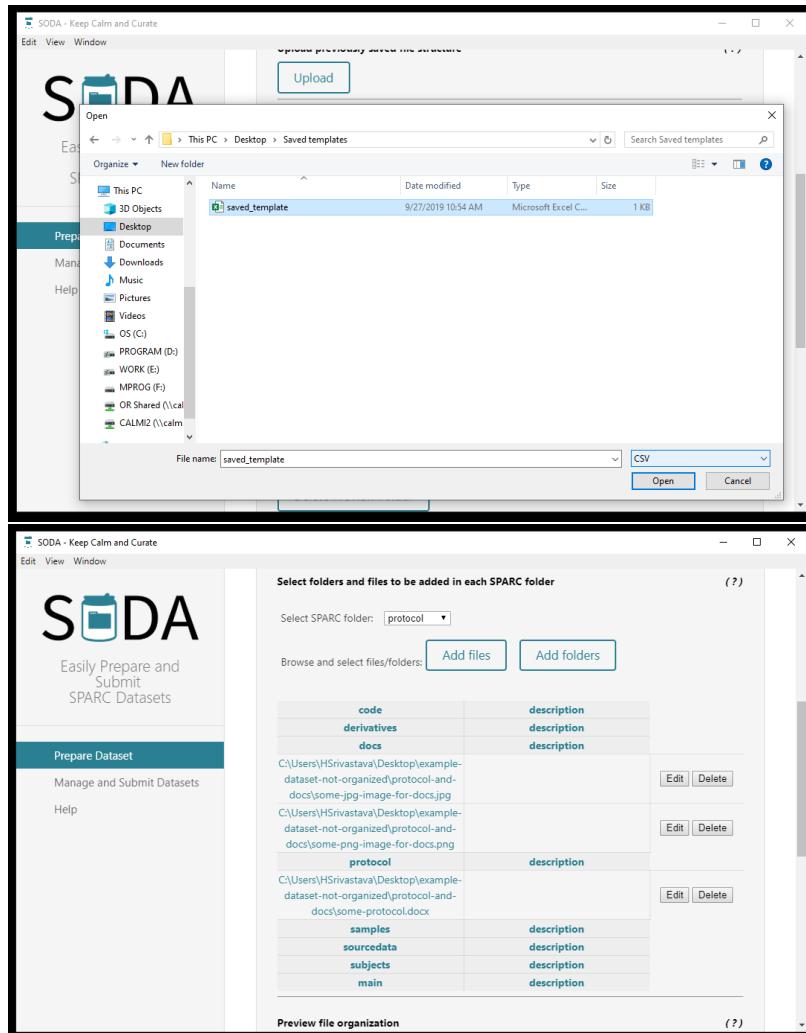


The screenshots show the SODA software interface for managing dataset organization. The top window displays a table for the 'protocol' folder, with columns 'code' and 'description'. The bottom window shows a more detailed table for the same folder, including additional rows for 'samples', 'sourcedata', 'subjects', and 'main'. Both windows include 'Edit' and 'Delete' buttons for each row.

code	description
derivatives	description
docs	description
C:\Users\Hsrivastava\Desktop\example-dataset-not-organized\protocol-and-docs\some-jpg-image-for-docs.jpg	
C:\Users\Hsrivastava\Desktop\example-dataset-not-organized\protocol-and-docs\some-png-image-for-docs.png	
protocol	description
C:\Users\Hsrivastava\Desktop\example-dataset-not-organized\protocol-and-docs\some-protocol.docx	
samples	description
sourcedata	description
subjects	description
main	description

3. Upload previously saved organization:

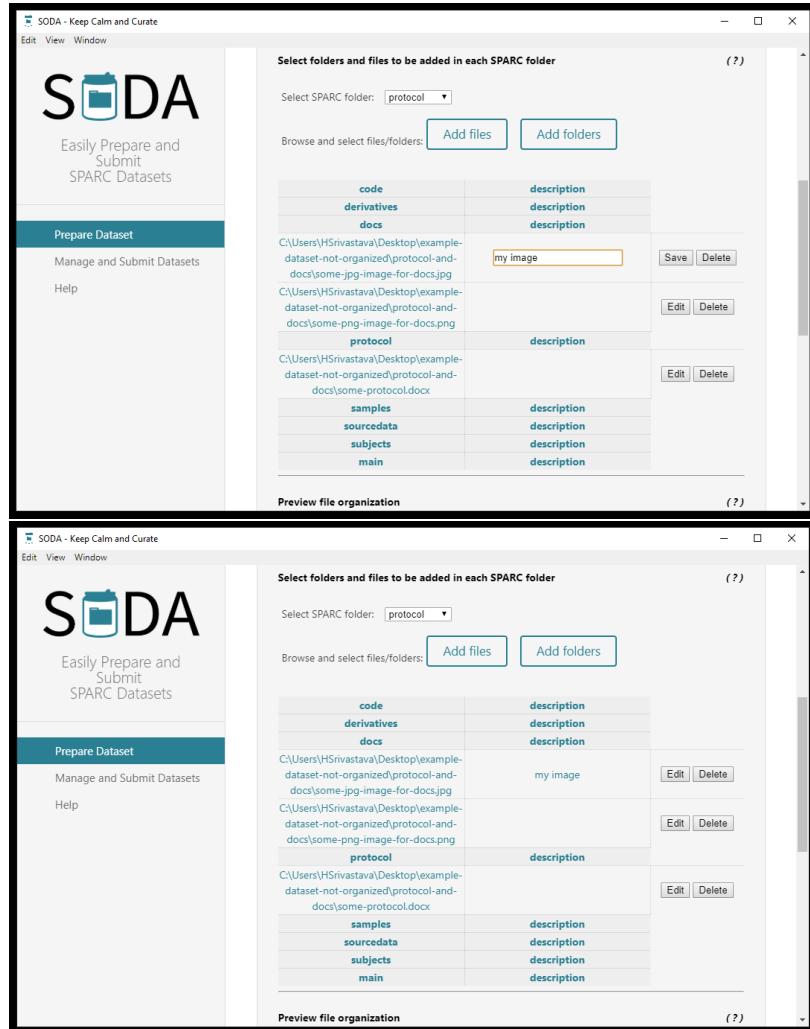
- Click on "Upload"
- Select previously saved table template (see explanation below)
- Paths for each header will be added to the table along with the description (see below). The paths/descriptions already in the table will not be removed.



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

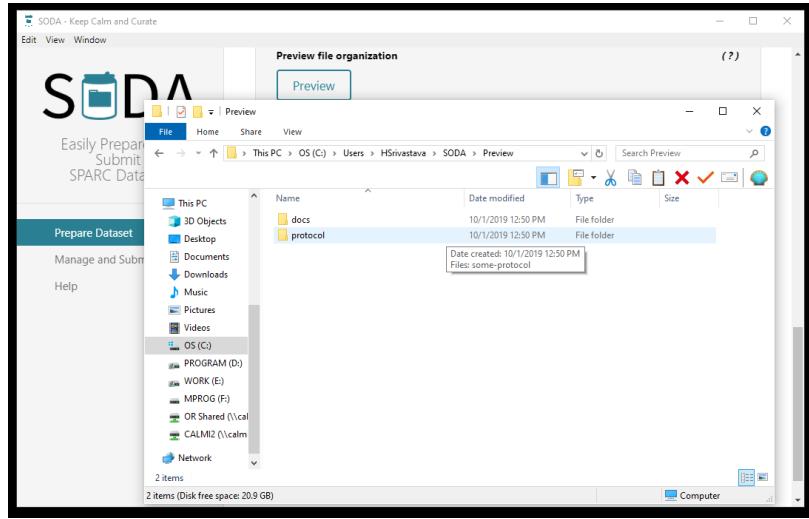
You can add a description for each of the files from the interface which would then be added to the manifest file.

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



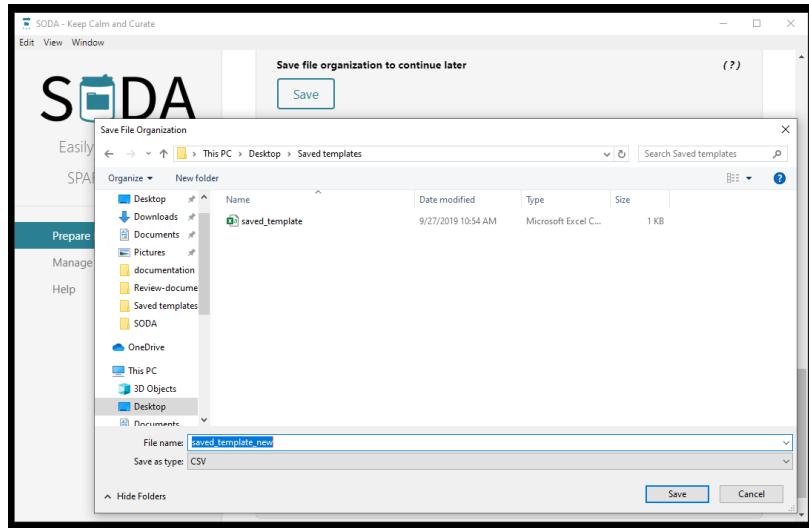
You can conveniently preview file organization to ensure it's organized as desired before the actual dataset is created.

- 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 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 (for e.g. MS Excel) to manually include paths and descriptions



code	docs_descriptive_derivative_docs	does_description	main	main_des_protocol	protocol_description	samples	samples_size
1	code_descriptive_derivative_docs	C:\Users\Hsrivastava\Desktop\example-data\my image	main	main_des_protocol	C:\Users\Hsrivastava\Desktop\example-dataset-not-organized\protocol	samples	samples_size
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							
24							
25							

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

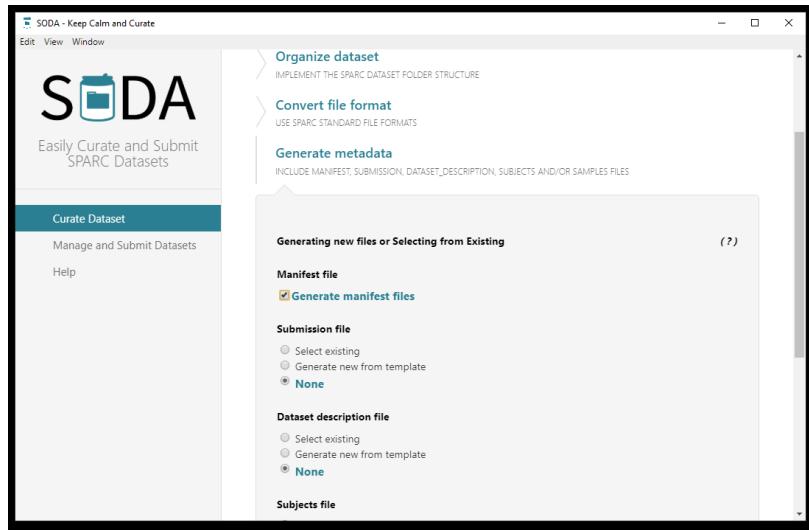
Coming soon

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.

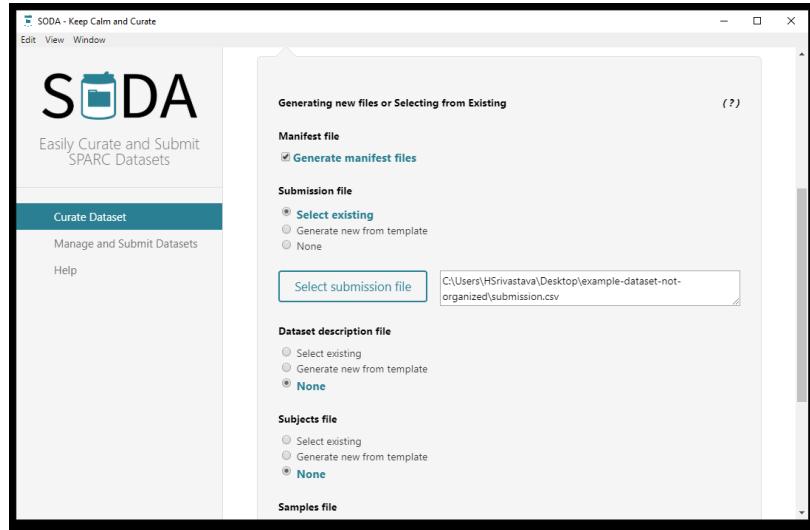


For all other files, there are three options:

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

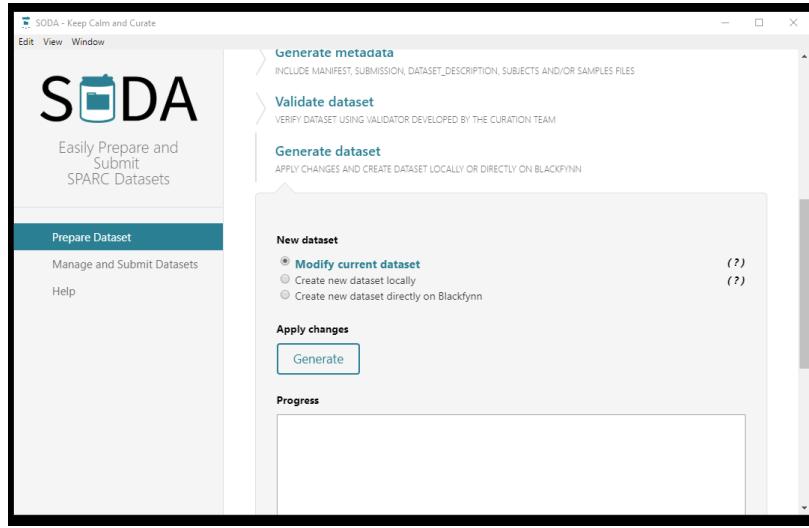
4.5 Generate dataset

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 "Generate" to do so.

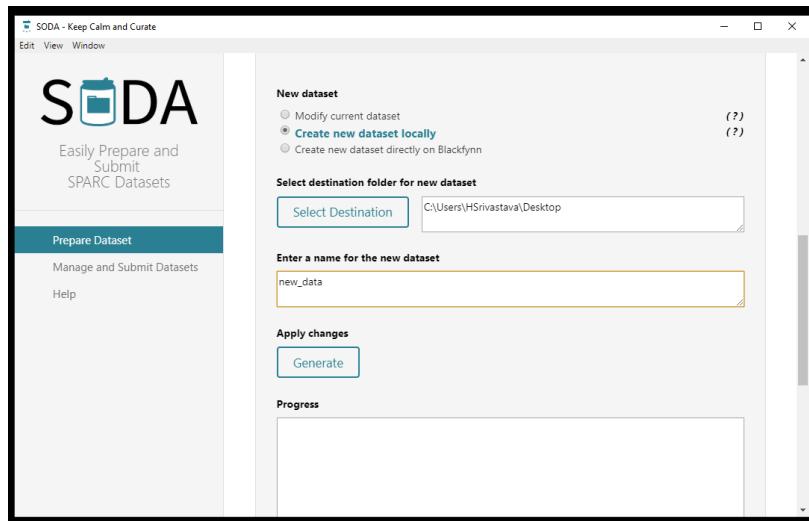
NOTE: This option is only available when you select "Dataset already organized" in the "Organize" section.



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 at a desired location with the desired name. 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

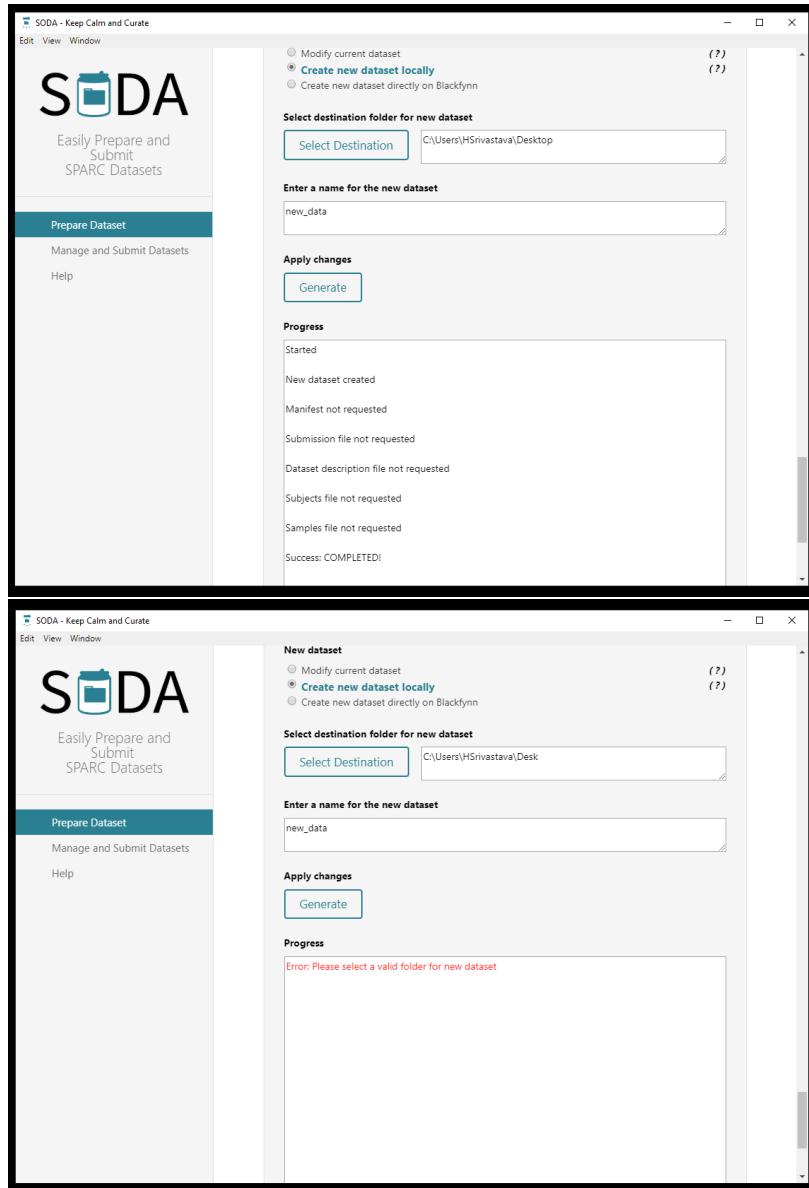


4.5.3 Create new dataset directly on Blackfynn

Coming soon

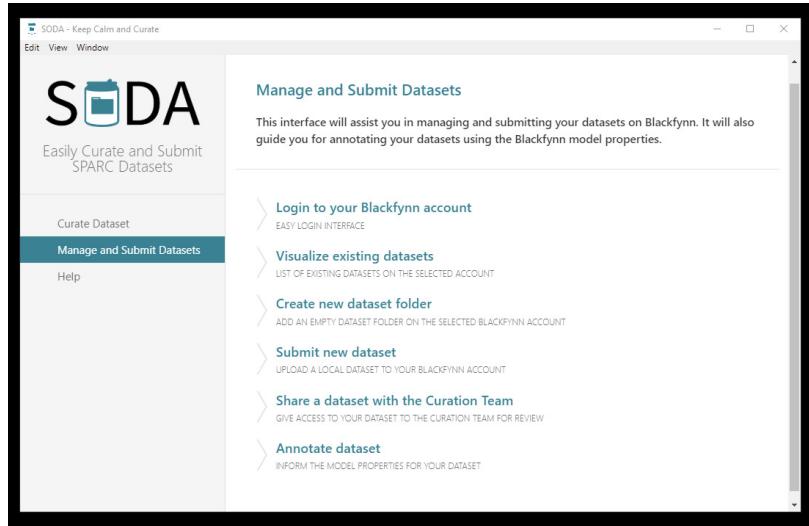
4.5.4 Progress window

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



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.

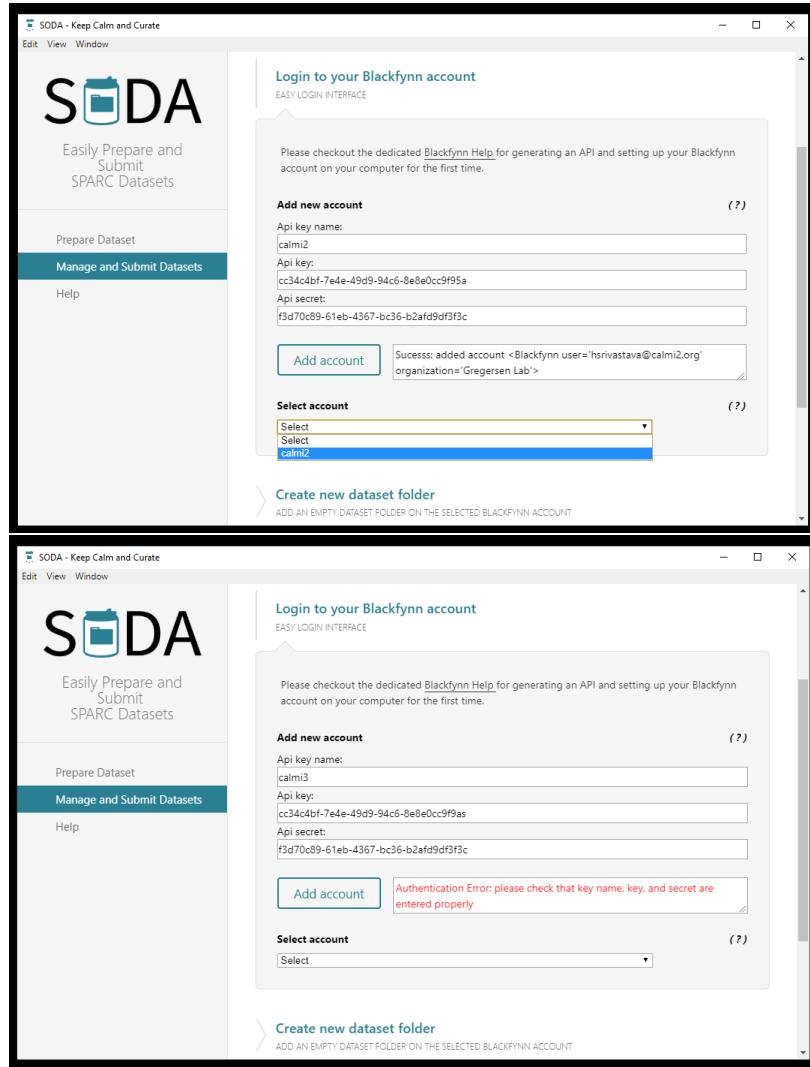
5.1 Login to your Blackfynn account

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:

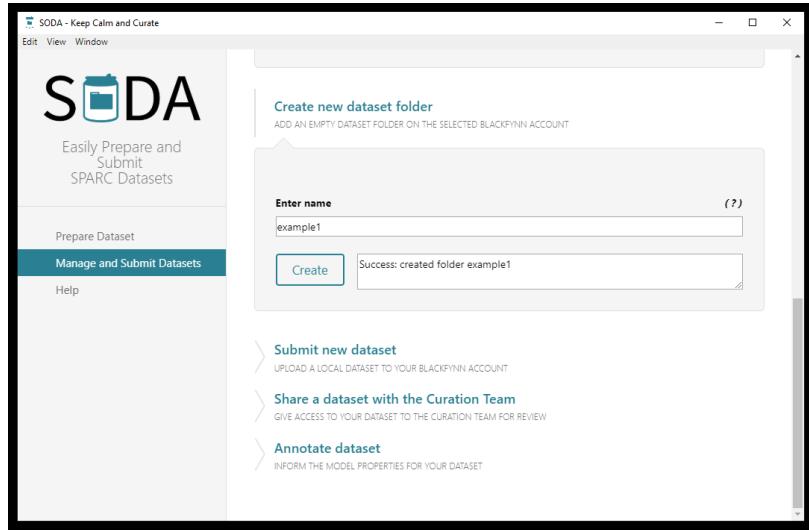
- Selecting an account under "Select account" is necessary to perform subsequent actions.



5.2 Create new dataset folder

To create a new dataset folder:

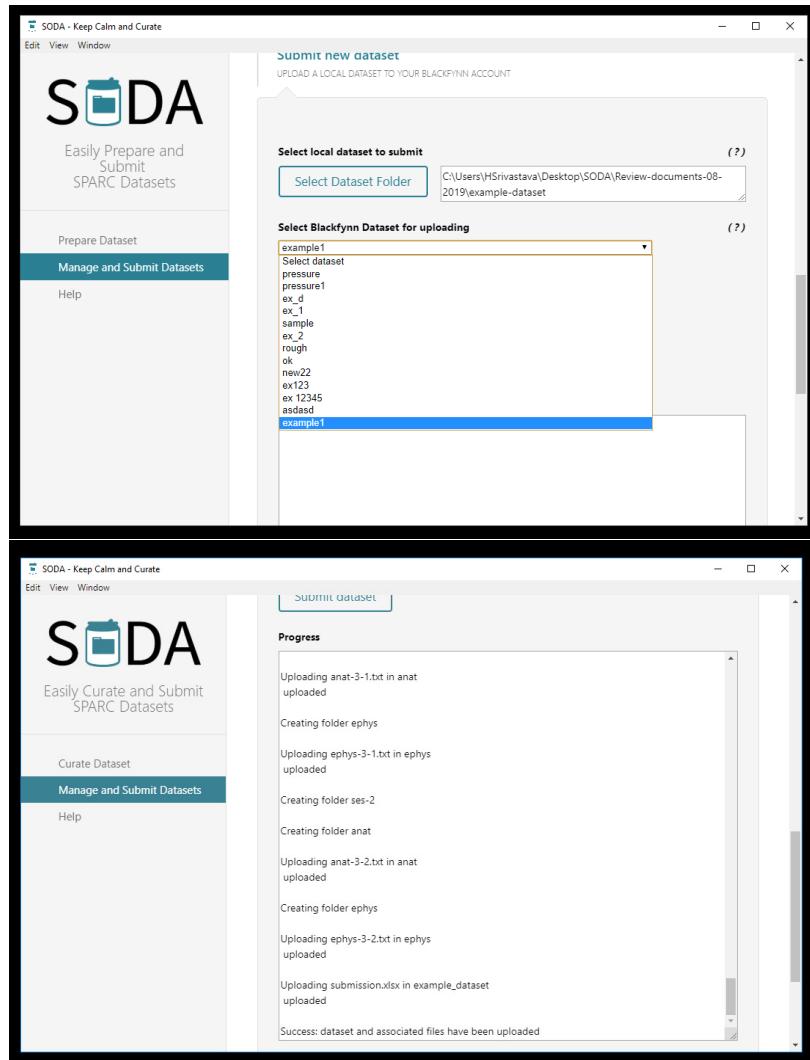
- Enter the desired name for the folder
- Click "Create"
- A success message will be displayed in the adjacent textbox if the dataset is created (it will also be added automatically to the dataset list in section "Submit new dataset" under "Select Blackfynn Dataset for uploading"). Otherwise, an error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

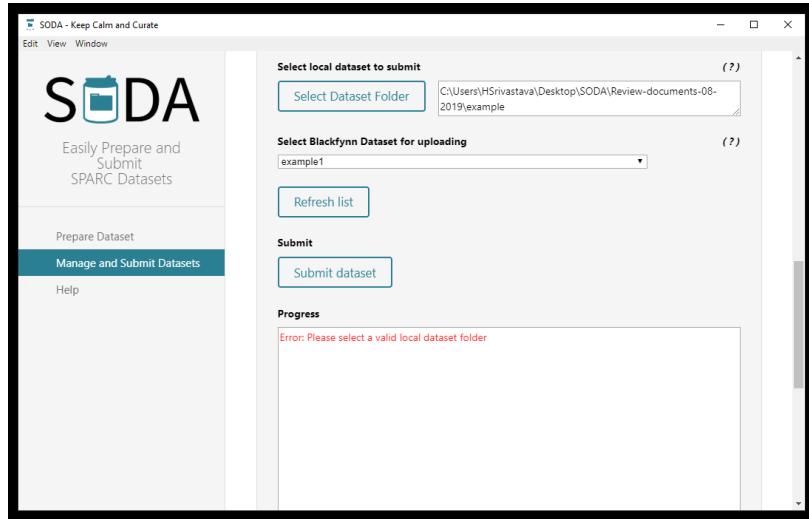


5.3 Submit new dataset

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 "Select Blackfynn Dataset for uploading"
- 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.

