

SODA User manual

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

July 2019



Contents

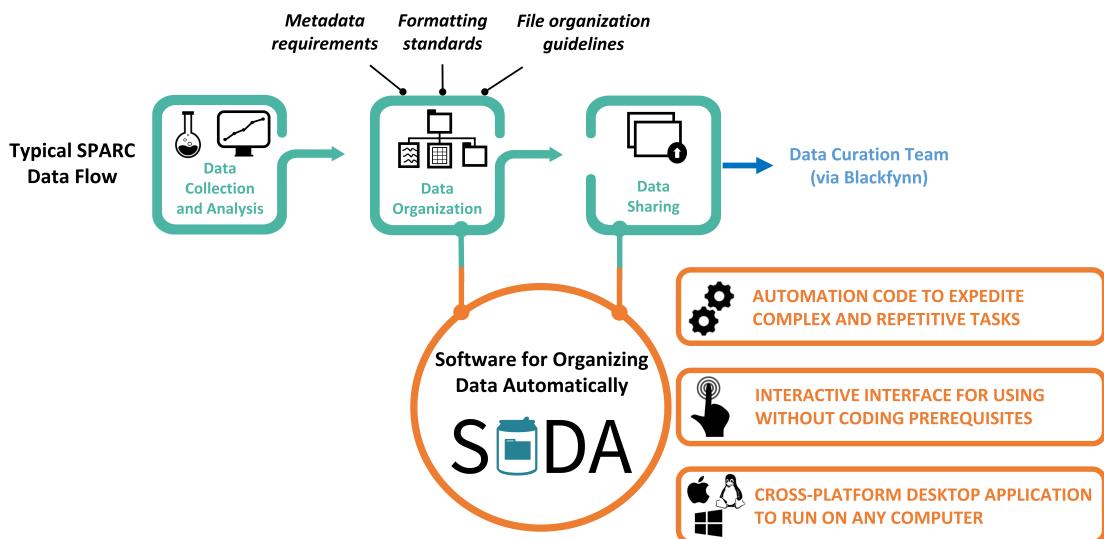
	Page
1 What is SODA?	2
2 Download and Open the Application	3
3 Main Interface	5
4 Curate Dataset	5
4.1 Organize dataset	6
4.2 Convert file format	7
4.3 Generate metadata	8
4.4 Validate dataset	9
4.5 Curate dataset	9
4.5.1 Modify current dataset	9
4.5.2 Create new dataset locally	10
4.5.3 Create new dataset directly on Blackfynn	11
4.5.4 Progress window	11
5 Manage and Submit Datasets	12
5.1 Login to your Blackfynn account	12
5.2 Visualize existing datasets	13
5.3 Create new dataset folder	14
5.4 Submit new dataset	15
5.5 Share a dataset with the curation team	17
5.6 Annotate dataset	17
6 Help	17

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!

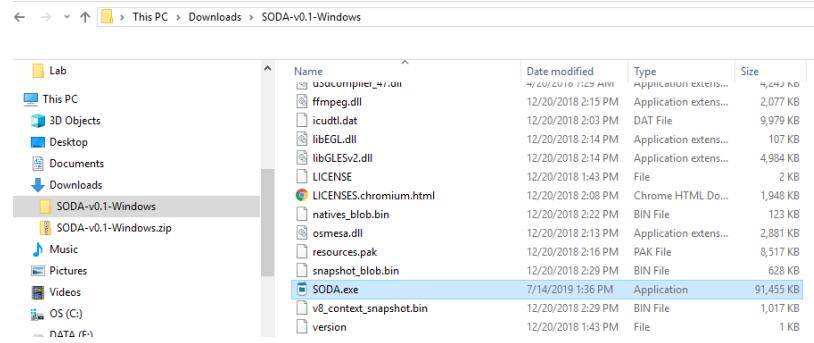
The front-end (Graphical User Interface or GUI) of SODA is built with Electron, an open-source framework developed and maintained by GitHub that conveniently combines HTML, CSS, and Javascript, while the back-end is developed in Python (v3.6). SODA is distributed as an easy to install application for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



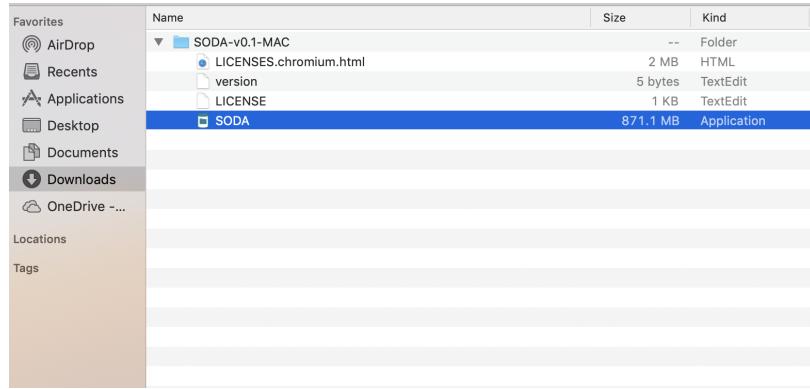
2 Download and Open the Application

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

- Windows (developed on Windows 10)
 1. Click here and download the "SODA-v0.1-Windows.zip" file
 2. Unzip the file if necessary (e.g., right click and select "Extract all")
 3. Open the unzipped folder "SODA-v0.1-Windows"
 4. Click on "SODA.exe" to launch the application (right click and select "Run as administrator" if opening is prevented)
 5. To stop the application, simply click on the cross in the top right corner.

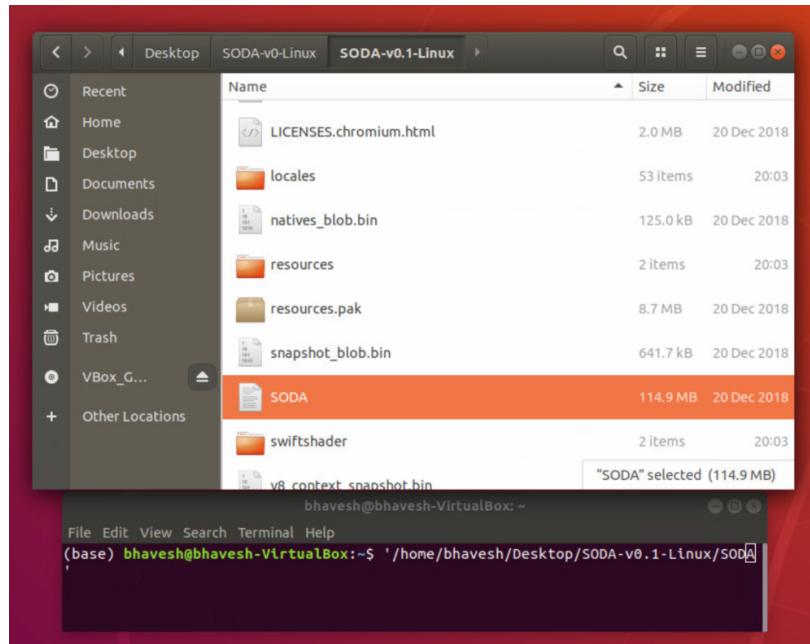


- MAC-OS (developed on MAC 10.14)
 1. Click here and download the "SODA-v0.1-MAC.zip" file
 2. Unzip the file if necessary (e.g., double click to extract with "Archive Utility" tool)
 3. Open the unzipped folder "SODA-v0.1-MAC"
 4. Click on "SODA.app" to launch the application (right click and select "Open" if opening is prevented)
 5. To stop the application, click on the cross in the top left corner. You may also need to right click on the SODA icon in the Dock and select "Quit".



- Linux (developed on Ubuntu 18.04.2 LTS)

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



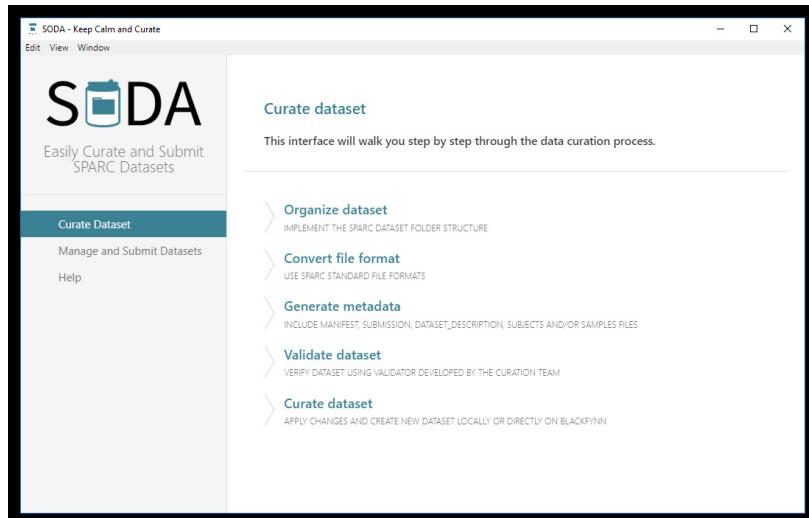
Note:

- 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, "Curate 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.
- When you reopen the application, the interface opened last time before closing is conveniently displayed again and the last frame size is remembered as well.
- Closing the interface will stop any on-going process (dataset organization, uploading, etc.).
- If any error occurs, close the app and re-open it.

4 Curate Dataset

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

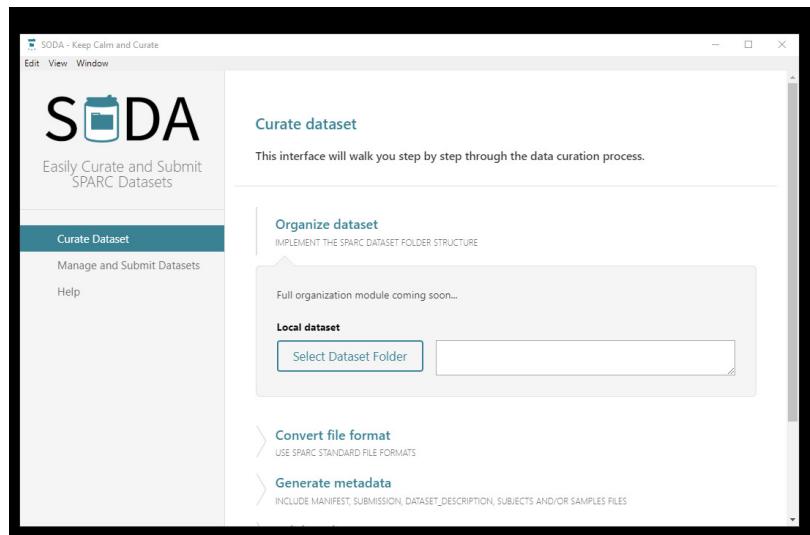
1. Organize dataset
2. Convert file format
3. Generate metadata
4. Validate dataset
5. Curate dataset

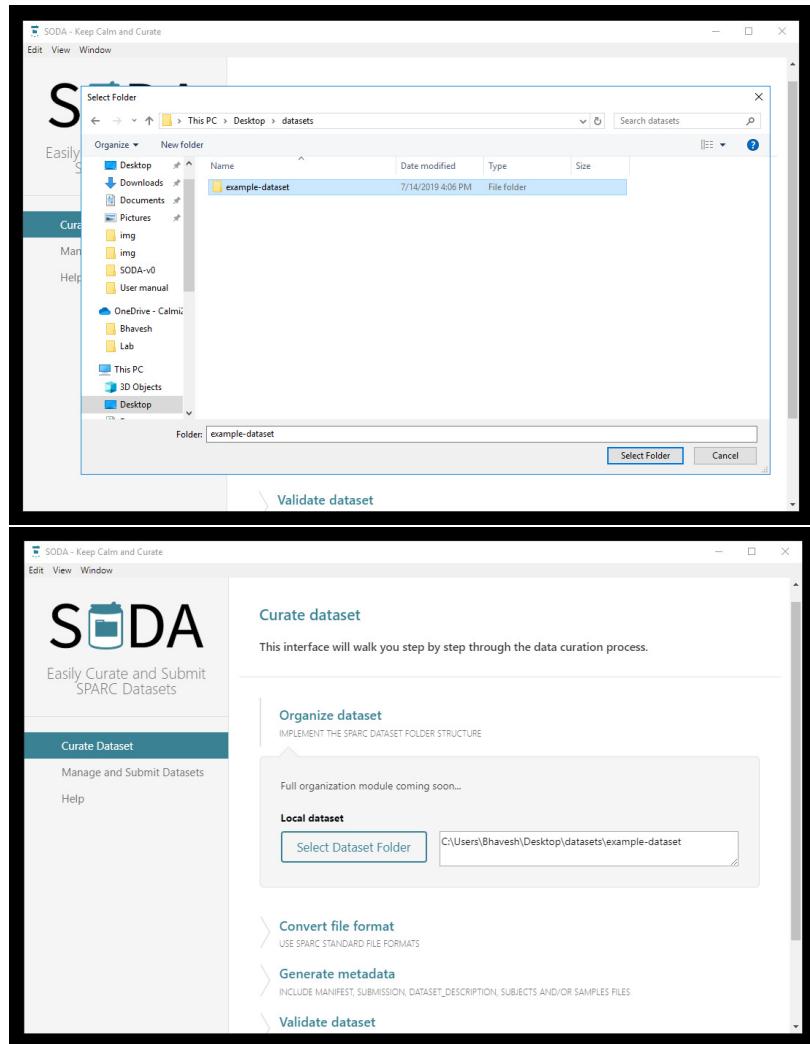
The first four options permit to select desired actions on the dataset which could then be applied under "Curate dataset".

4.1 Organize dataset

Currently, this is where you can choose the dataset you want to curate. It is assumed that the dataset is already organized according to the SPARC defined folder structure.

- Click on "Select Dataset Folder"
- Navigate to your dataset folder and select it
- The path of the selected dataset will appear on the adjacent textbox





Note:

- The text box is editable so you can type the path of the dataset directly (or, e.g., copy/paste it) without going through the navigator.
- The navigator only allows to select a folder.

4.2 Convert file format

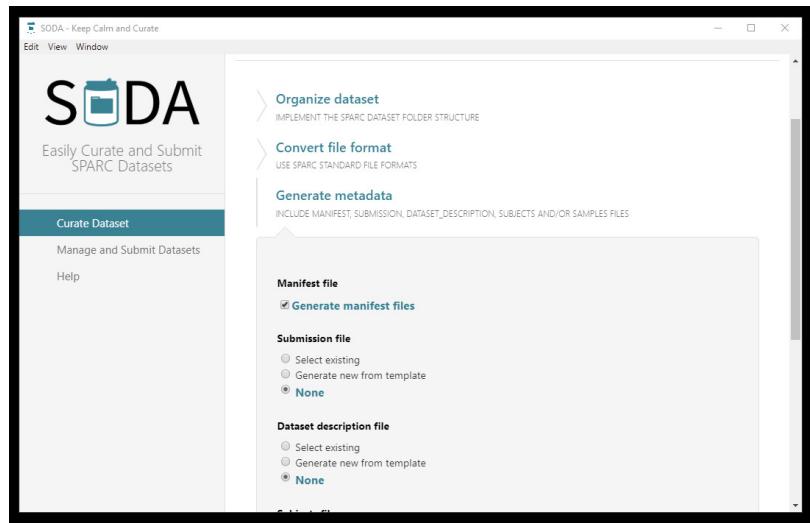
Coming soon

4.3 Generate metadata

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

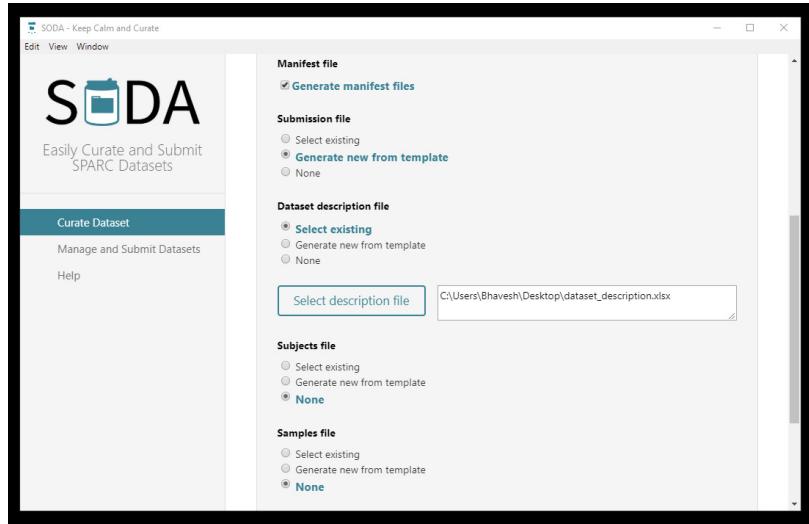
- manifest.xlsx within each first level folders of your dataset (it is not required in sub-folders)
- submission.xlsx
- dataset_description.xlsx
- subjects.xlsx
- samples.xlsx

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.
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 dataset

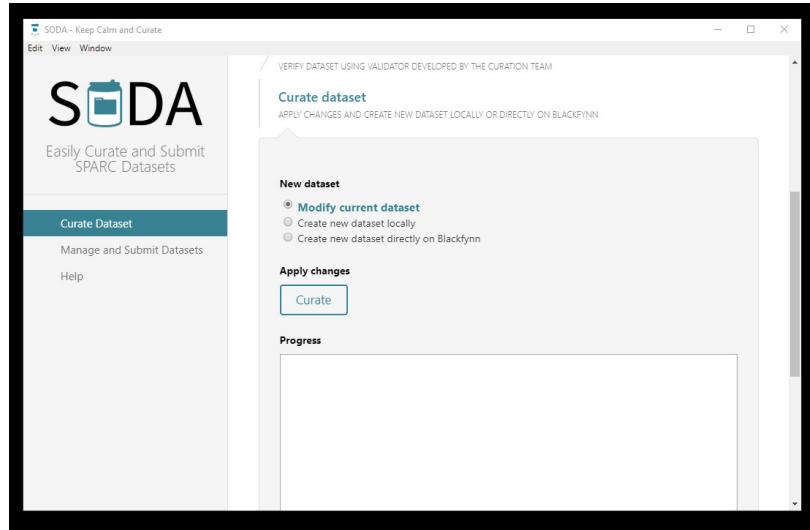
Coming soon

4.5 Curate 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

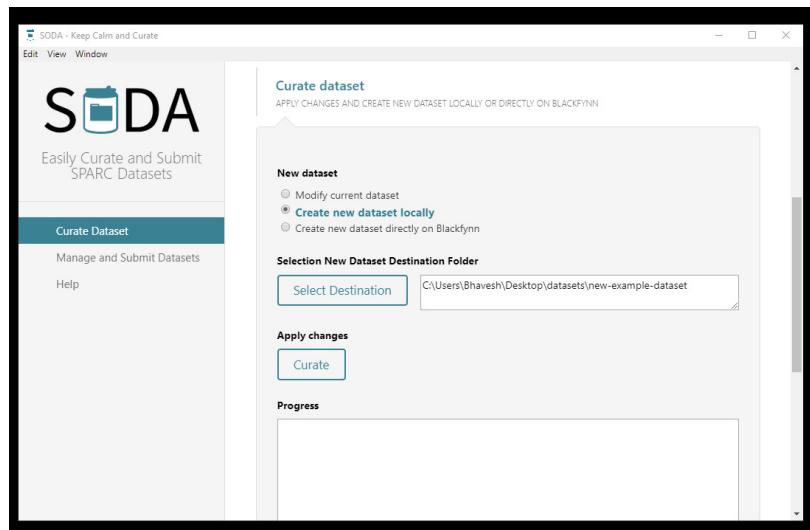
If "Modify current dataset" is select, the requested actions will be applied to the current dataset folder (selected under "Organize dataset"). Simply click on "Curate" to do so.



4.5.2 Create new dataset locally

If "Create new dataset locally" is select, 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.
- Click "Curate" to apply requested actions

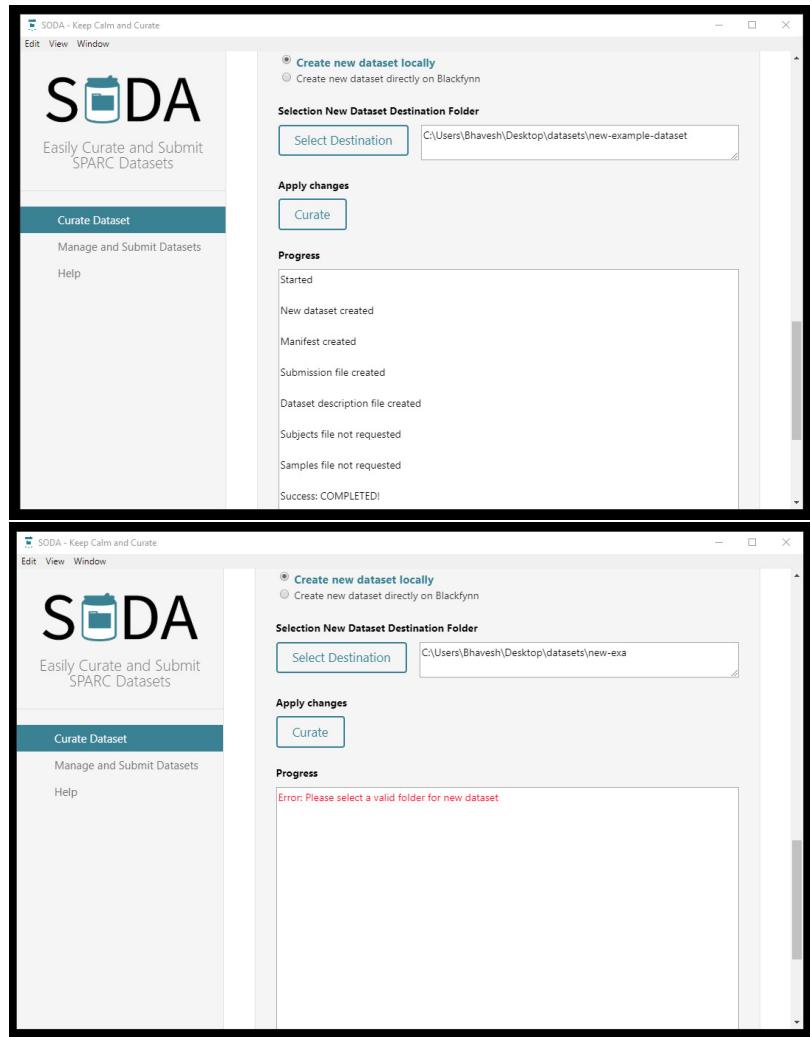


4.5.3 Create new dataset directly on Blackfynn

Coming soon

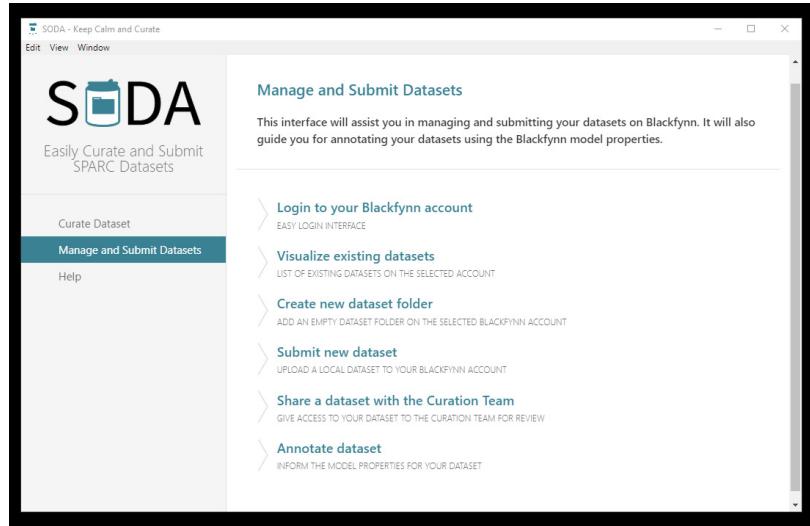
4.5.4 Progress window

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



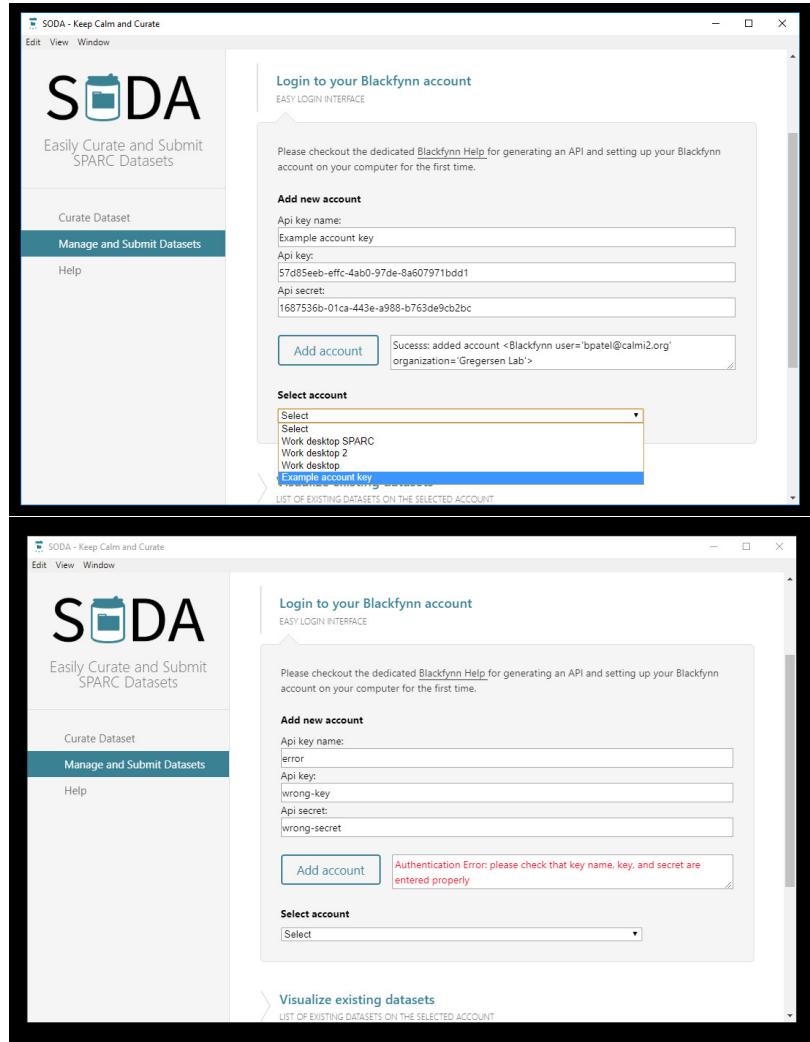
5.1 Login to your Blackfynn account

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".

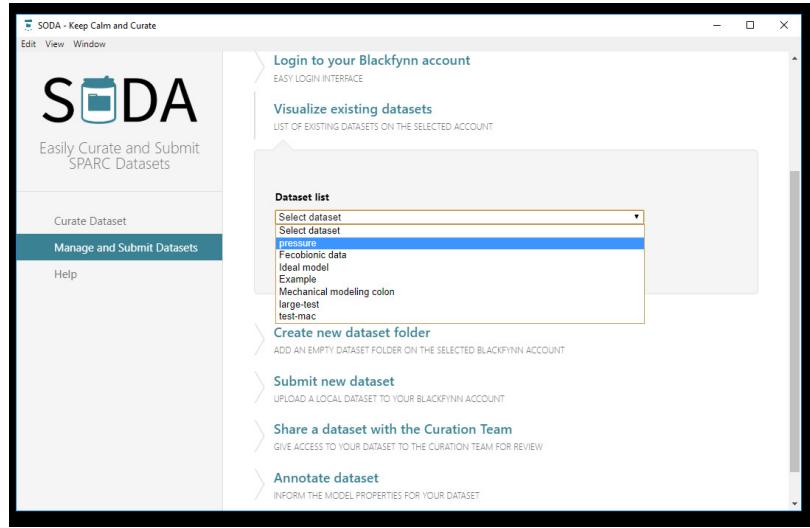
Note:

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



5.2 Visualize existing datasets

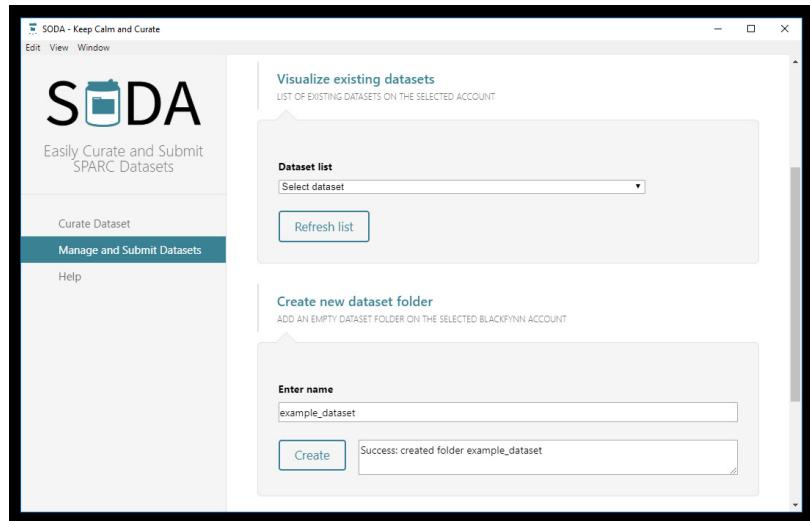
When an account is selected in the previous menu, all the existing datasets will be displayed under the "Dataset list" dropdown list. If a new dataset is created through the blackfynn web page, click on "Refresh" to see it appear in this list.

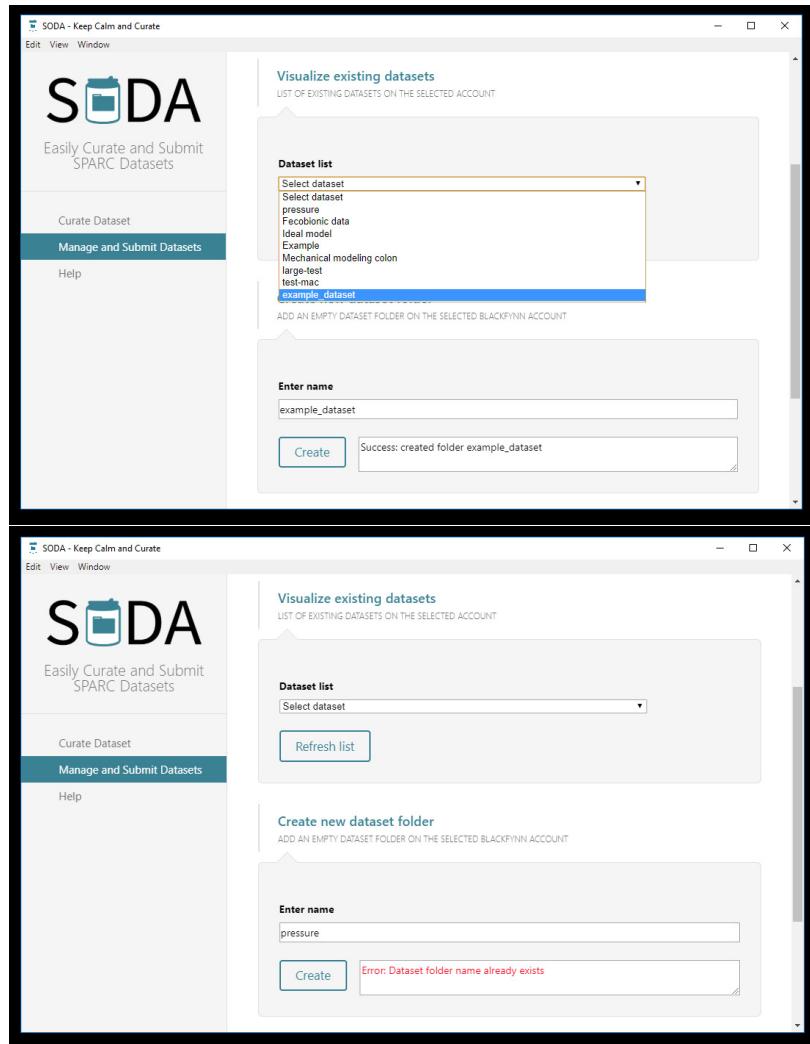


5.3 Create new dataset folder

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 under "Visualize existing datasets"). Otherwise, an error message will be displayed with the nature of the error.

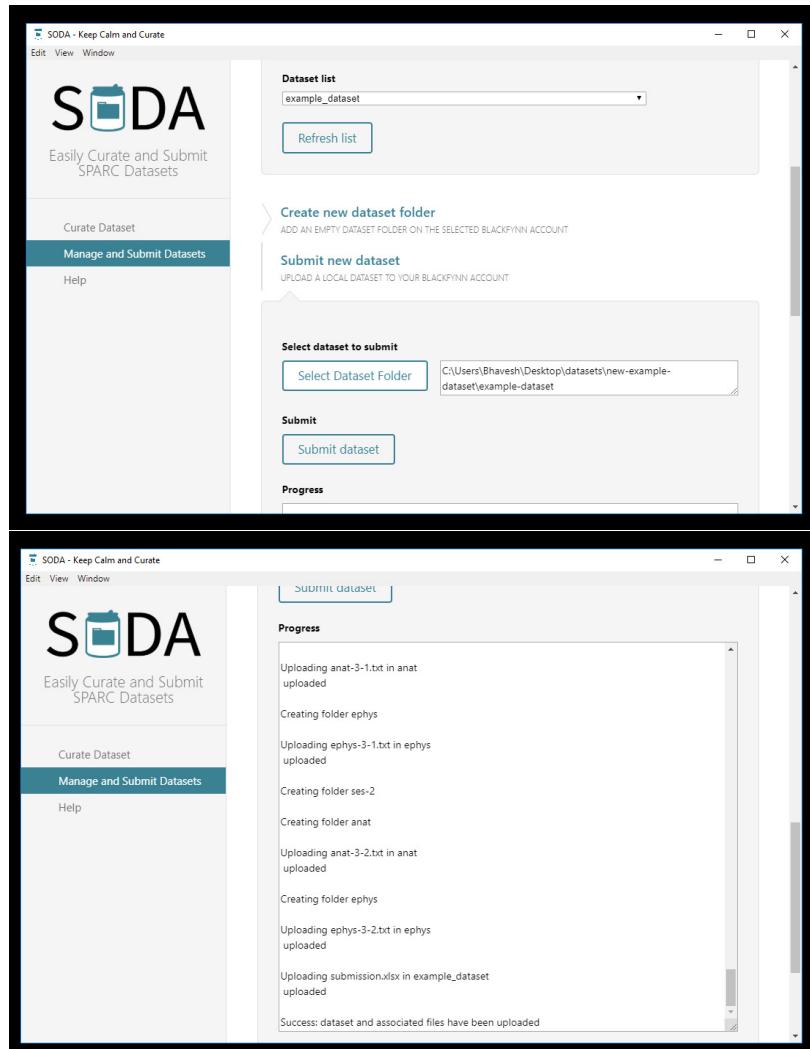


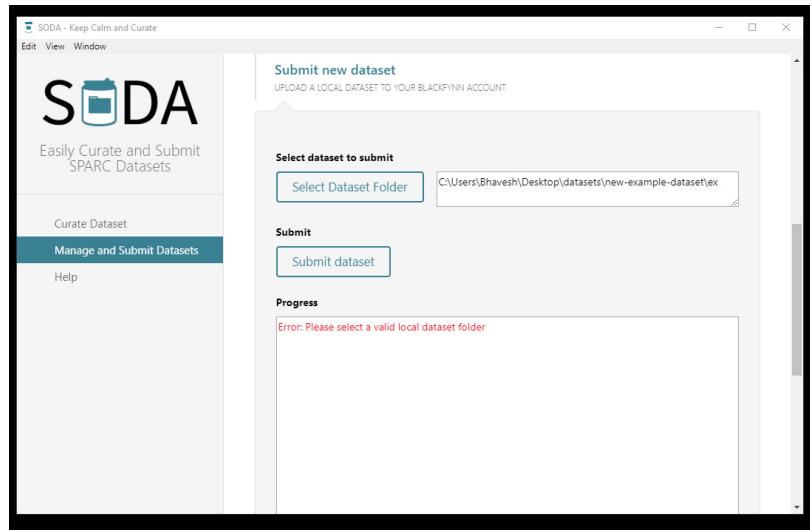


5.4 Submit new dataset

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 textbox
- 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





5.5 Share a dataset with the curation team

Coming soon

5.6 Annotate dataset

Coming soon

6 Help

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

