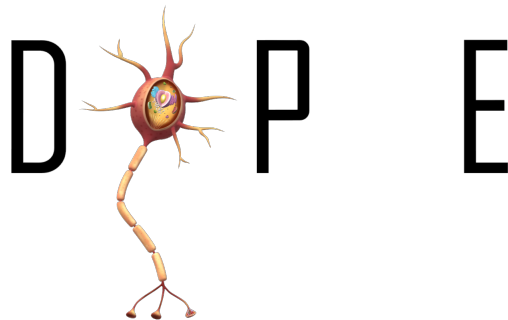


# DOPE User manual v1.1

Bhavesb Patel (bpatel@calmi2.org)  
Parya Aghasafari (paghasafari@ucdavis.edu)

December 2018



# Contents

	Page
1 What is DOPE?	2
2 Current state and future developments	3
3 Installation	4
4 Main interface	7
5 Organize your dataset	7
6 Validate your dataset	8
7 Share your dataset	10
8 BIDS-MRI notebook	14

# 1 What is DOPE?

Once a milestone of a project is reached, SPARC investigators need to format and annotate the corresponding datasets according to standards established by the Data Standard Committee, before sharing them as instructed by the data sharing policy. To achieve that, investigators must identify the applicable guidelines for their specific datasets and generate associated documentations and 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 data analysis, which should ideally remain the primary focus of the investigators.

Data Organization Process Enhancer (DOPE) is a computer program intended to facilitate the data organization process by:

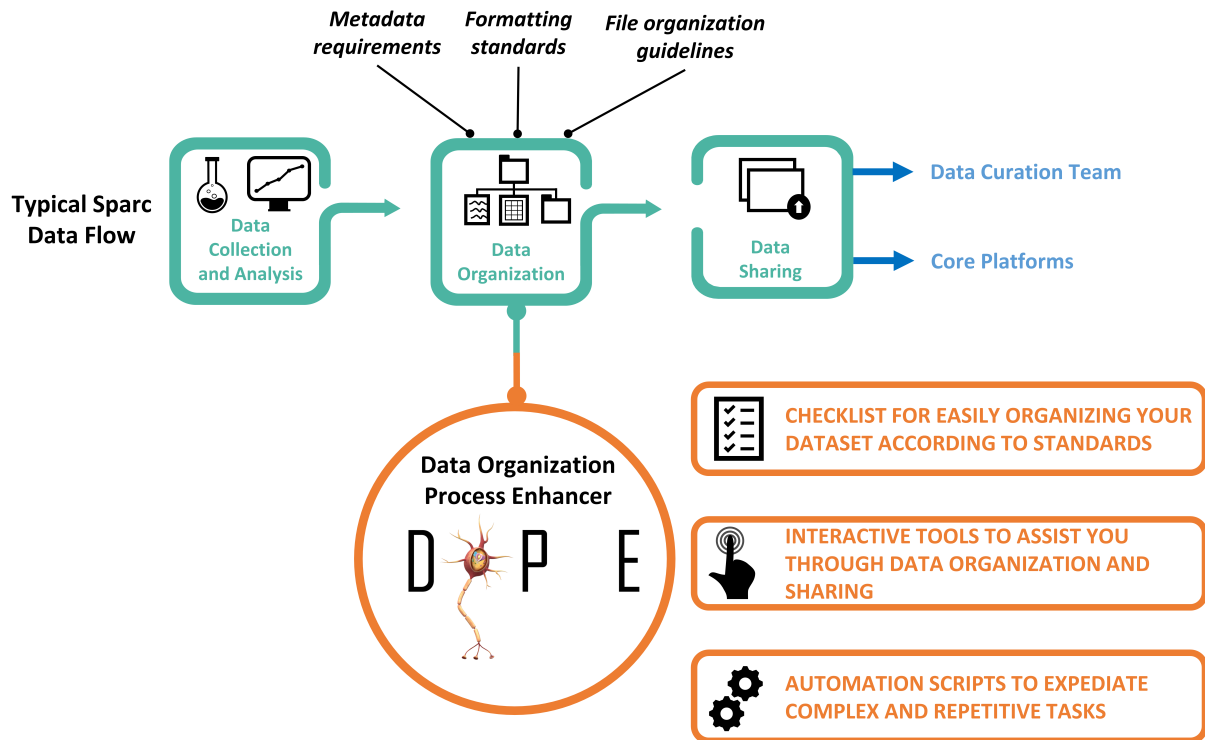
- Centralizing all information related to SPARC standards, guidelines, and recommendations
- Simplifying data organization tasks through interactive tools and automation

If at any point during your data organization process you think "ah, if only this step could be simplified/assisted/automated", it actually could be with DOPE!

With DOPE, users will be able to obtain a simple checklist to meet the SPARC standards applicable to their specific type of dataset. For each item of the checklist, DOPE will provide link to relevant resources (websites, articles, etc.), templates of required forms, and associated computer tools.

DOPE is developed in Jupyter (Python 3) such that is it cross-platform (Windows, Mac, Linux) and easily modifiable by the users according to their needs. All source codes and files are shared with an Open Source license (MIT) to permit user modification and distribution without restrictions, and eventually promote a user-driven development of DOPE.

Thanks to Stefano Morotti, Enrico Ravagli, Roberta Sclocco, and Nicole Thompson for their contributions to the initial version of DOPE during the SPARC Hackathon.



## 2 Current state and future developments

Currently, a checklist generator is available (with sample checklists and templates included for illustration purpose for some data types). A dataset validator is integrated to check if a dataset contains all required files (an example is included for illustration purpose). A Blackfynn notebook is included to manipulate datasets interactively on Blackfynn (create new datasets, upload files without losing the original folder structure, etc.). A BIDS notebook for MRI files is included as well. It includes a step-by-step guide to satisfy the BIDS guidelines along with helpful tools and templates. Details for each functionality are provided in the next sections.

Further development of DOPE will be considered if interest is expressed by the SPARC community, including:

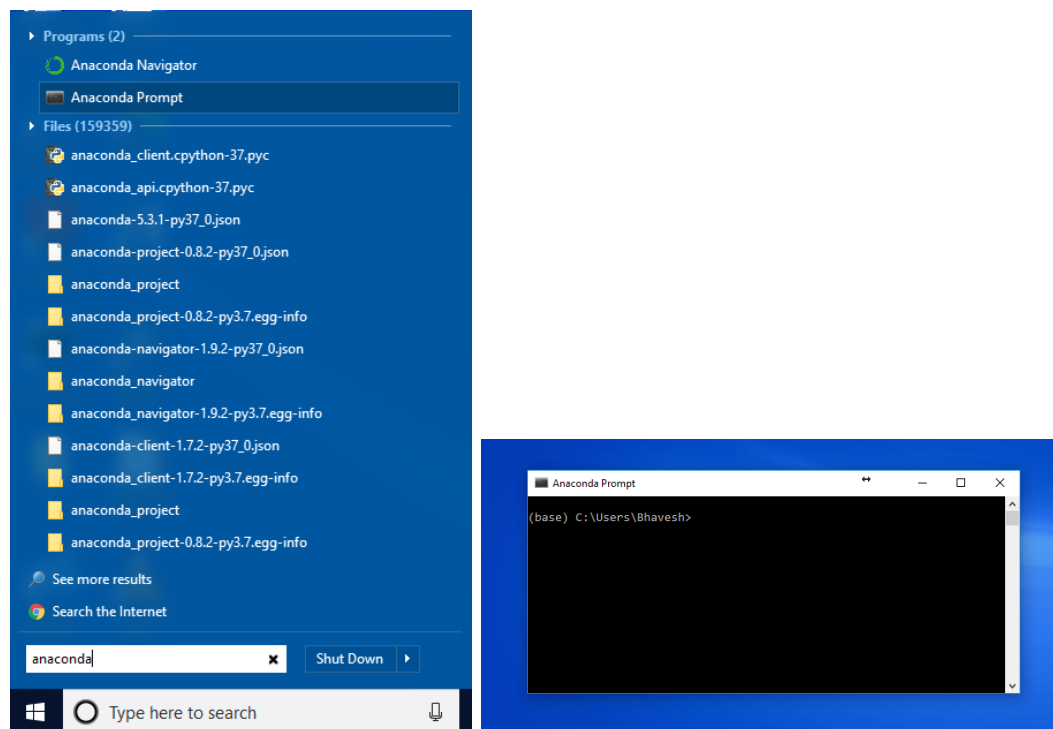
- Actual checklists and form templates based on indications from the Data Curation Team
- Automated file organization based on BIDS guidelines
- BIDS notebooks for data type other than MRI

- Additional functionality in the Blackfynn notebook: direct API login from Jupyter (without requiring to go through the command line), renaming dataset, live upload status, creating data models and records, adding relations, etc.
- Thorough dataset validators
- DOPE video tutorials
- More functionality based on requests from users and DRC teams!

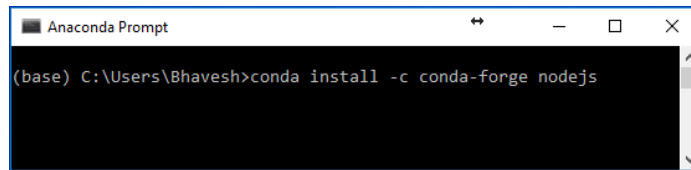
### 3 Installation

The easiest way to use DOPE is described below (illustrations are for Windows platform):

1. Install **Anaconda** for Python version 3 following the Anaconda installation guidelines for your operating system. Anaconda conveniently installs Python, Jupyter Notebook, and JupyterLab.
2. Open the **Anaconda Prompt** (which is simply the Terminal if you are on macOS or Linux).

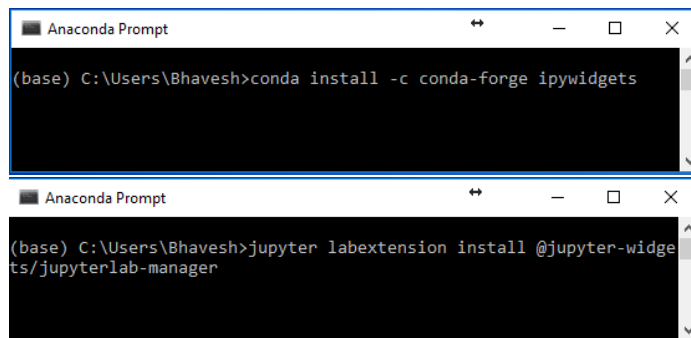


3. Install **nodejs** from the Anaconda Prompt following the conda nodejs installation guide.



```
(base) C:\Users\Bhavesh>conda install -c conda-forge nodejs
```

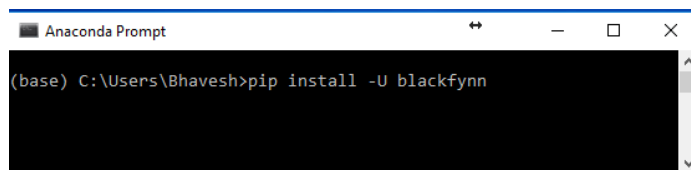
4. Install **ipywidgets** from the Anaconda command prompt following the conda ipywidget installation guide. Make sure to follow the additional step to install the JupyterLab extension.



```
(base) C:\Users\Bhavesh>conda install -c conda-forge ipywidgets

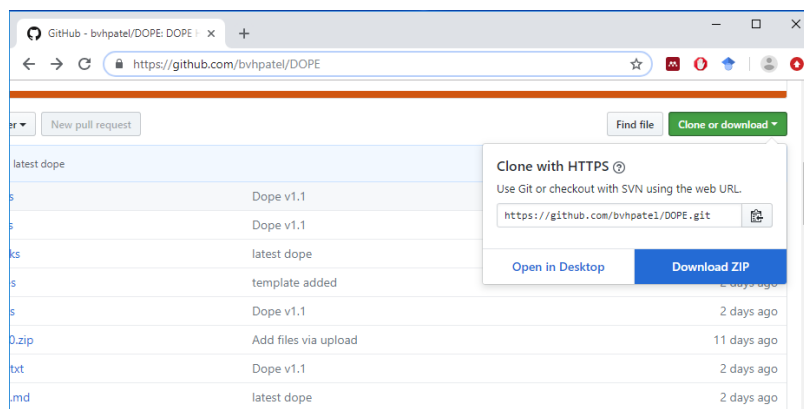
(base) C:\Users\Bhavesh>jupyter labextension install @jupyter-widgets/jupyterlab-manager
```

5. Install the **Blackfynn python API** from the Anaconda Prompt following the pip Blackfynn installation guide.

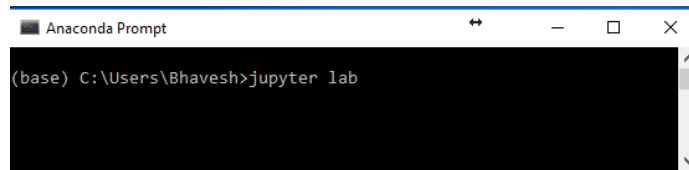


```
(base) C:\Users\Bhavesh>pip install -U blackfynn
```

6. Download all **DOPE** files from the Github repository (click "Clone or download" and select "Download ZIP").



7. In the Anaconda Prompt, type "jupyter lab" to launch **JupyterLab**.

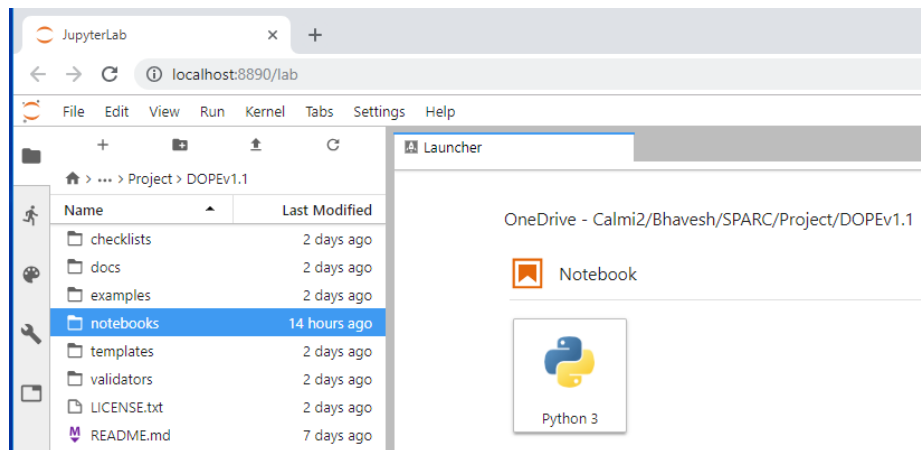


```

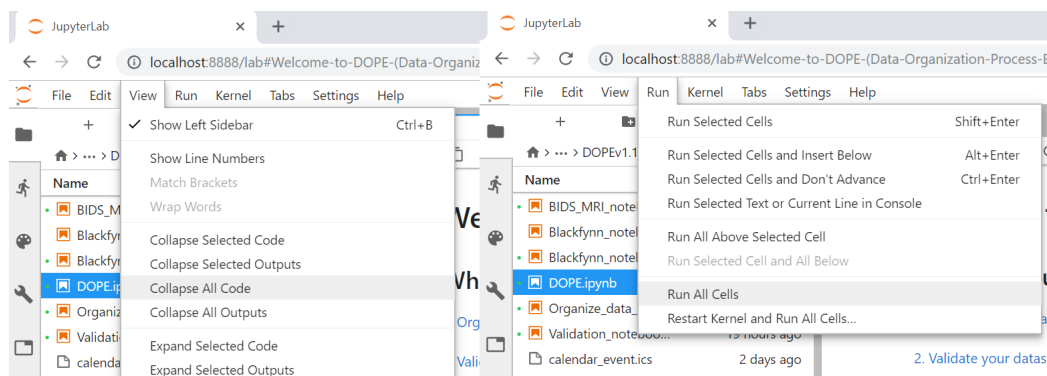
Anaconda Prompt
(base) C:\Users\Bhavesh>jupyter lab

```

8. Locate your DOPE files from the JupyterLab navigator. Open DOPE.ipynb from the "notebooks" folder of DOPE and **get started!**



Whenever you open a notebook, we recommend to do "View → Collapse All Codes" to hide all the ugly code lines, then do "Run → Run All Cells" to activate all the functions. Repeat these steps if some unexpected error occur.

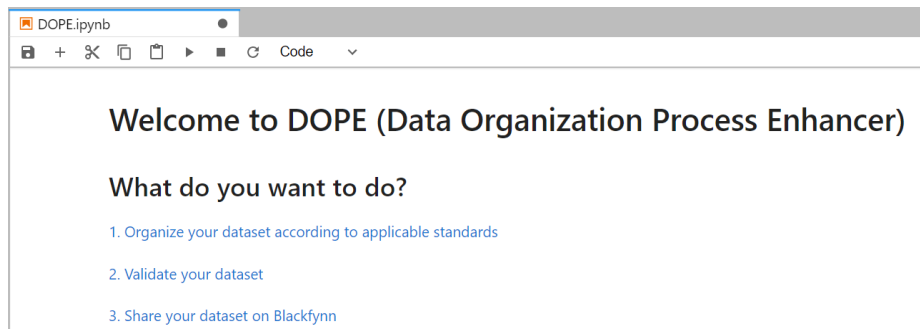


In the future, we are hoping DOPE will be integrated in the SPARC Docker developed by Dr. Horn.

## 4 Main interface

Open DOPE.ipynb from the "notebooks" folder to access the main interface. Three options are available:

1. Organize your dataset according to applicable standards
2. Validate your dataset
3. Share your dataset on Blackfynn



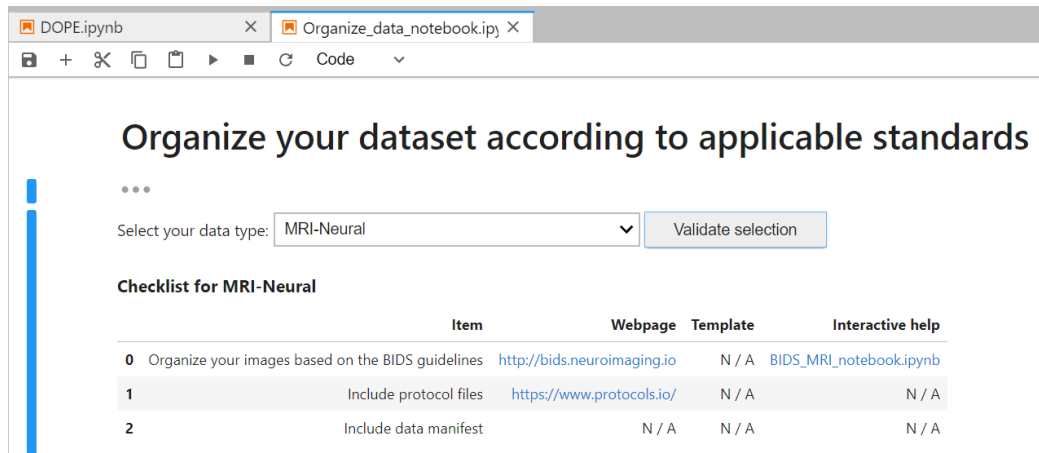
Each of the options are clickable. A click on any option will open the corresponding notebook. Details on their functions are provided in the next sections.

## 5 Organize your dataset

This notebook (Organize\_data\_notebook.ipynb) generates a checklist of steps to follow in order to meet the applicable SPARC standards for the selected data type.

- Do "View → Collapse All Codes" to hide all the code lines, then do "Run → Run All Cells" to activate all the functions of the notebook (see Installation section).
- Select your data type from the drop-down list.
- Click on the "Validate selection" button.
- The corresponding checklists are displayed in a table.





- For each item, the table contain link to external resources, templates, and interactive notebooks when applicable.

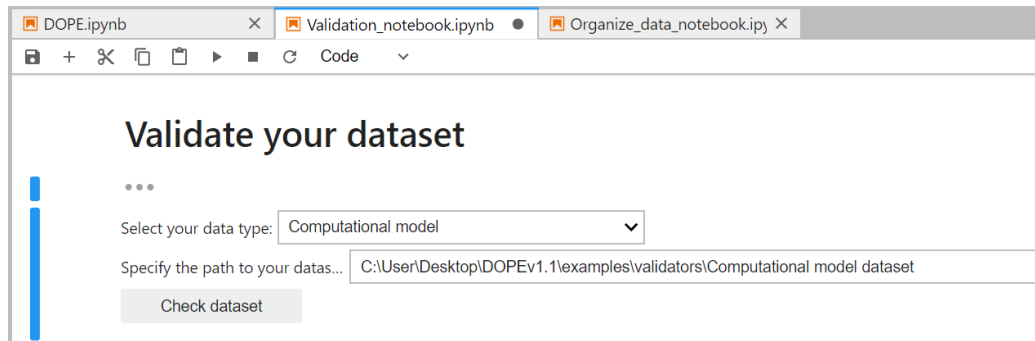
Notes:

- The checklists are based on Excel files contained in the "checklists" folder of DOPE.
- Adding an Excel file to this folder for a new data type would automatically add it to the drop-down list.
- When clicked on their links, the templates are downloaded from the DOPE repository on Github. They are also accessible directly in the "templates" folder of DOPE.
- The interactive notebooks are located in the "notebooks" folder and open when clicked in a new Jupyter tab.

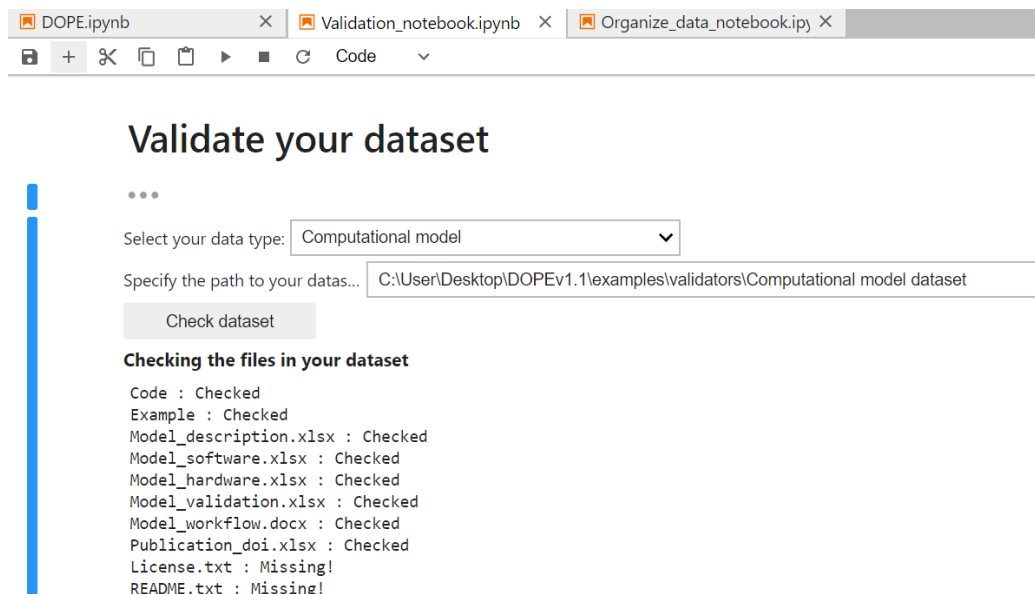
## 6 Validate your dataset

This notebook (Validation\_notebook.ipynb) allows to verify that required files are included in a dataset and the files are organized according to guidelines.

- Do "View → Collapse All Codes" to hide all the code lines, then do "Run → Run All Cells" to activate all the functions of the notebook (see Installation section).
- Select your data type from the drop-down list.
- Enter the path to your dataset folder in the text box.



- Click on the "Check dataset" button.
- A list of the required files and folders in your dataset is generated with either a "Checked" indication if the file/folder is found in the dataset (at the right location) or "Missing!" if it is not.



Notes:

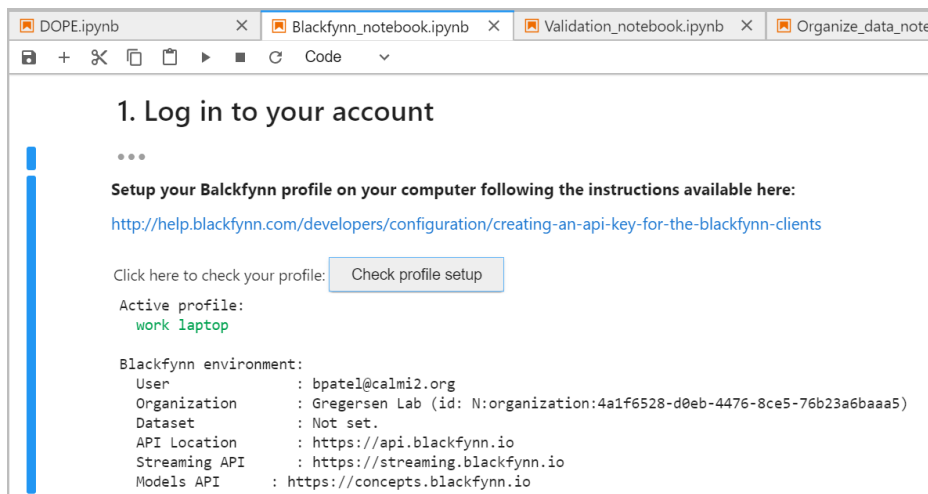
- The list of expected files and folders is generated based on the Excel files contained in the "validators" folder.
- Adding an Excel file to this folder for a new data type would automatically add it to the drop-down list (currently only "Computational model" is included as example).
- For now, only first level of files and folders are checked. In the future, a more thorough check (subfolders and files) will be included.

## 7 Share your dataset

This notebook (Blackfynn\_notebook.ipynb) allows to manipulate files on the Blackfynn platform. Before continuing, do "View → Collapse All Codes" to hide all the code lines, then do "Run → Run All Cells" to activate all the functions of the notebook (see Installation section).

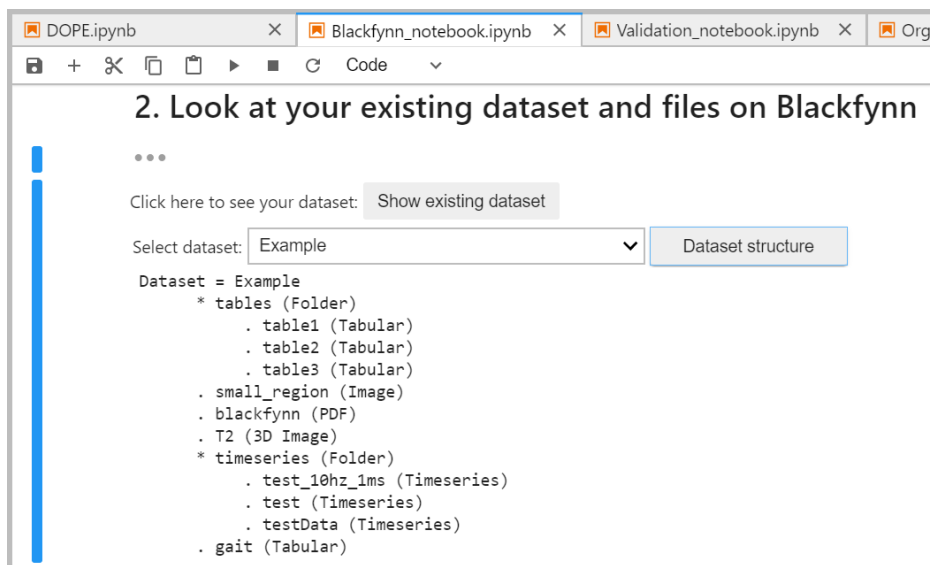
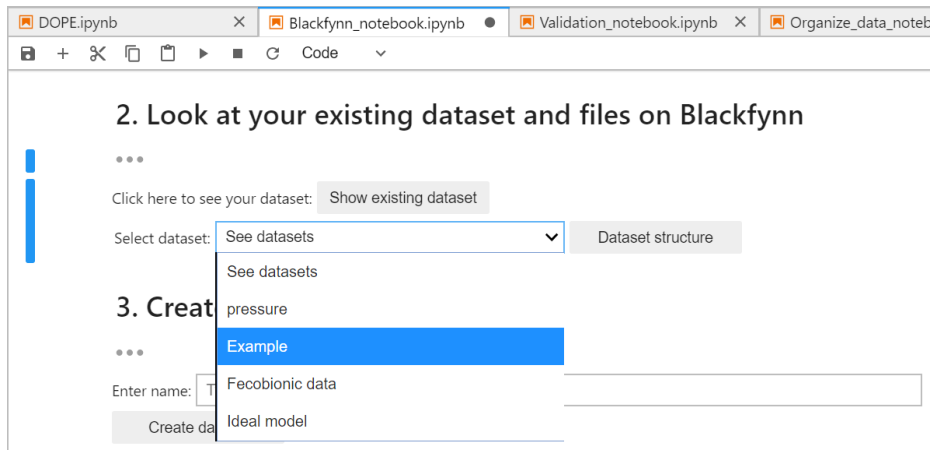
### 1. Log in to your account

Your Blackfynn profile must be setup on your computer according to the instruction provided on the platform. Click on the "Check profile setup" button to verify that your profile is set up properly.



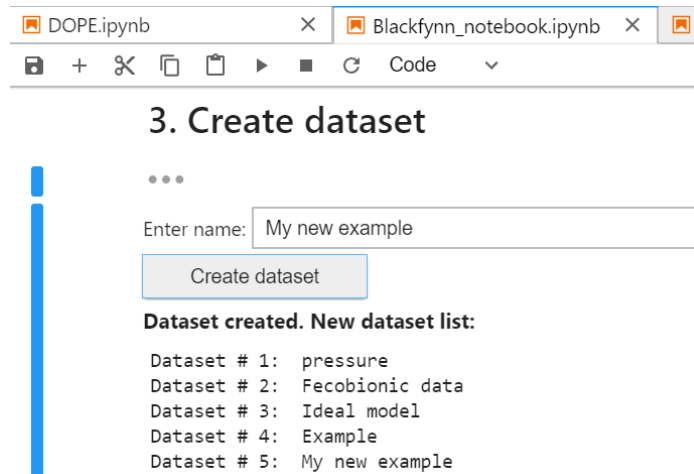
### 2. Look at your existing dataset and files on Blackfynn

Click on "Show existing dataset" to generate a drop-down list with all your datasets. Select a dataset from the drop-down list and click on "Dataset structure" to see the content of the dataset displayed with a nice structure along with information about file type.



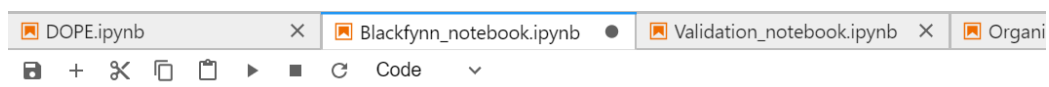
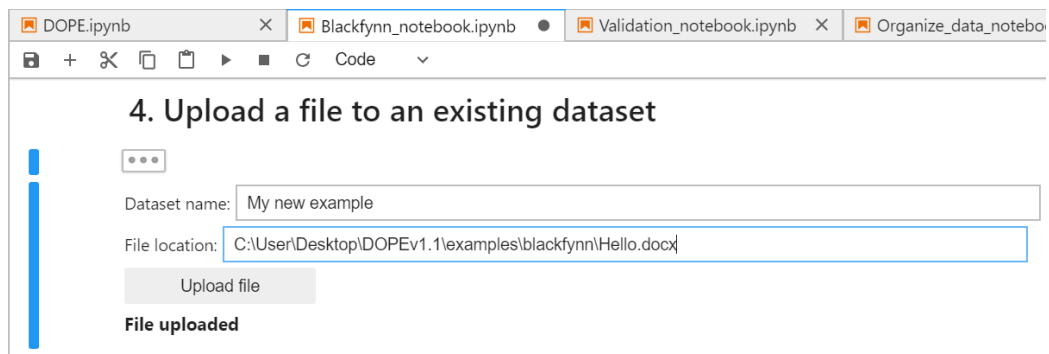
### 3. Create dataset

Enter a dataset name in the text box and click on "Create dataset" to create the new dataset. A list of your datasets is printed to show that your new dataset has been added.

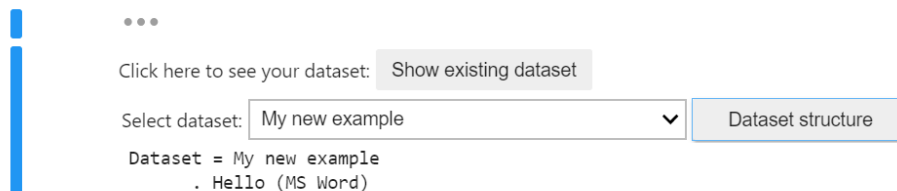


#### 4. Upload a file to an existing dataset

Type the name of the dataset where you want to upload the file in the first text box. Enter the path of the file to be uploaded in the second text box. Click "Upload file" to upload. You can go back to step #2 and check that your file has indeed been uploaded.

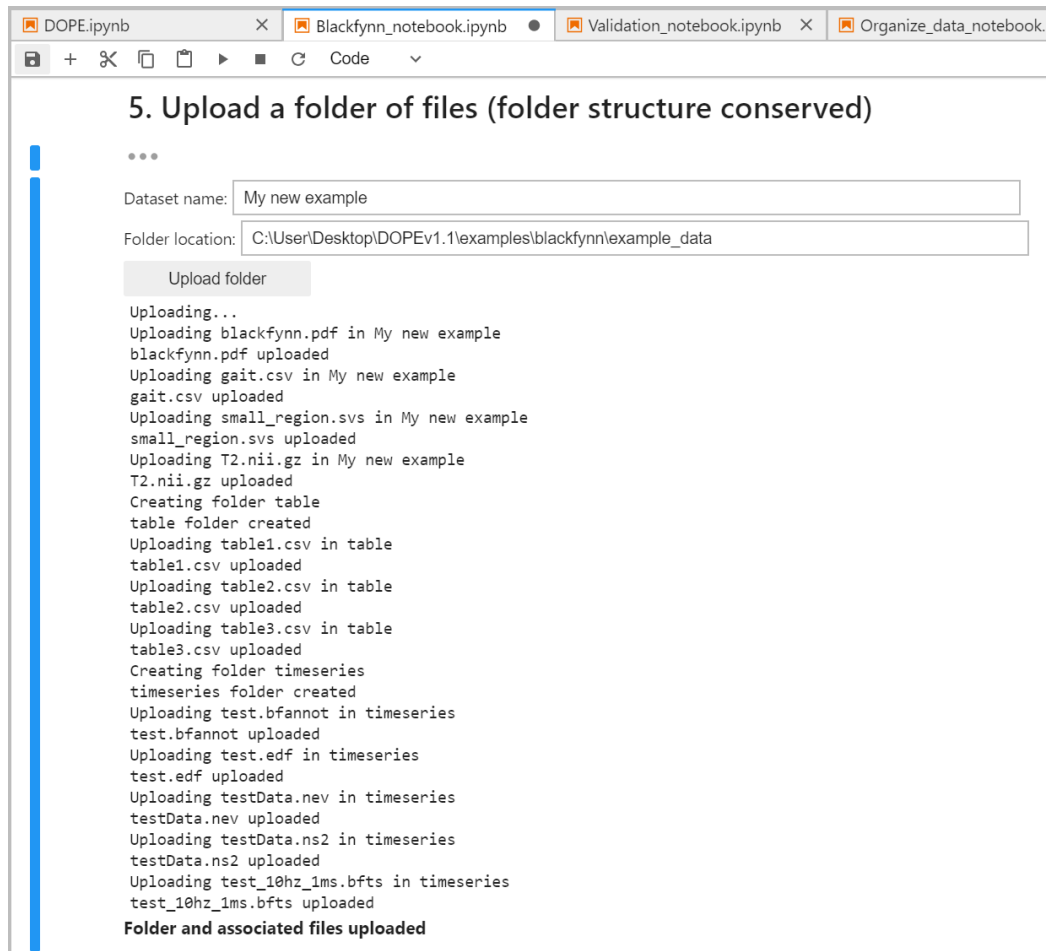


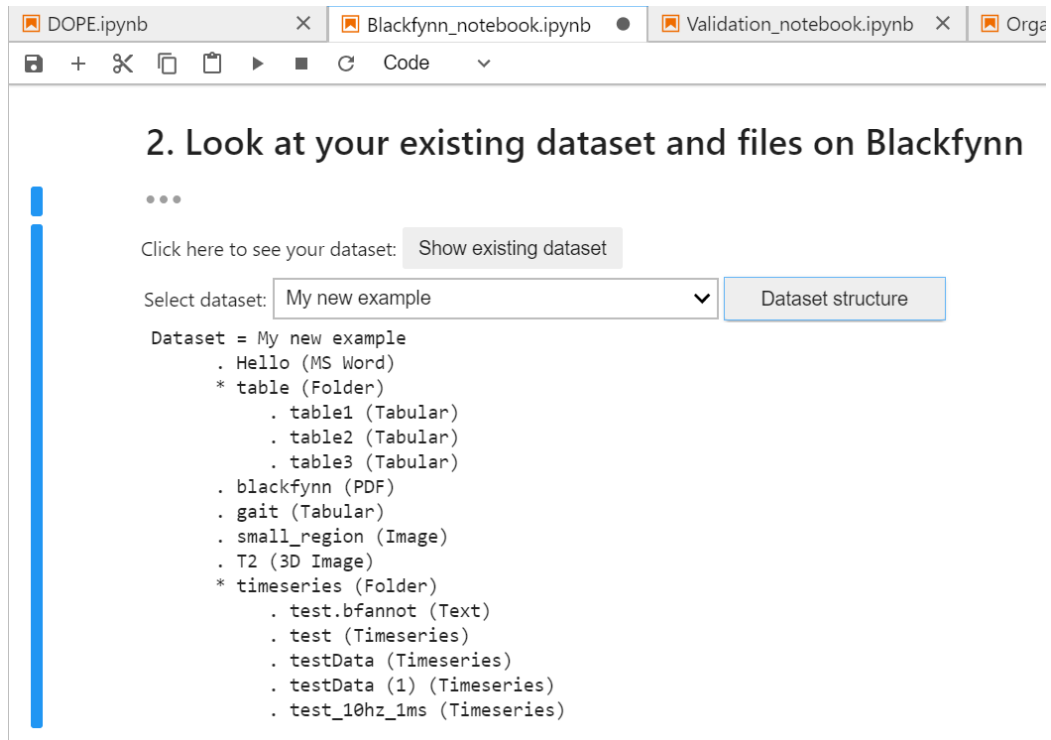
#### 2. Look at your existing dataset and files on Blackfynn



## 5. Upload folder of files (folder structure conserved)

Type the name of the dataset where you want to upload the files in the first text box. Enter the path of the folder to be uploaded in the second text box. Click "Upload folder" to upload. Status of the file being uploaded is provided until all the files are uploaded. You can go back to step #2 and check that your files have indeed been uploaded with the original folder structure.





## 6. Create a calendar reminder.

Enter your calendar reminder event in the first text box (e.g. Share dataset "Great data" with the Data Curation Team by tomorrow). Enter year (yyyy), month (mm), day (dd), hour (hh, 24h format), and min (mm) in the successive boxes. Click on "Create calendar event". This action generates a "calendar\_event.ics" file (called iCalendar format file) in the "notebooks" folder of DOPE. This file could be imported to any calendar (Google, Outlook, etc.) to add the reminder.

Notes:

- The "example\_data" used for illustration is provided by Blackfynn and is available [here](#).

## 8 BIDS-MRI notebook

This notebook (BIDS\_MRI\_notebook.ipynb) provides a step-by-step guide to meet the BIDS specifications for neural MRI files along with useful tools. It is currently at an early stage of development with few functions. For instance for those who are unfamiliar or uncomfortable with json files (required by BIDS) a function is included to enter your information in an Excel file and generate the corresponding json file automatically.

DOPE.ipynb

Blackfynn\_notebook.ipynb

BIDS\_MRI\_notebook.ipynb

Validation\_notebook.ipynb

+

✂

▶

■

↺

Code

▼

# BIDS v1.1.1 Notebook

## 1. Describe your dataset in a json file

...

Download file [dataset\\_description.xlsx](#), fill it out, and save it in your dataset

Indicate path of dataset\_description.xlsx folder:

Click here to generate json file: Create json

Creating json file  
json file is created at C:\Users\Calmi2\Desktop\DOPEv1.1\examples\BIDS\dataset\_description.json