



# How to run the MSDA infrastructure v2.1

<b>How to run the MSDA infrastructure v2.1</b>	<b>1</b>
STEP 1: UPLOAD YOUR COVID-19 CORE DATASET	1
STEP 2: RUN THE PIPELINE	2
STEP 3: Interacting with UI	3
Navigate to browser	3
Type of analysis	3
Running the Analysis	3
STEP 4: Zipped Result	4
STEP 5: Stopping the container	5
STEP 6: Removing the container	5

## STEP 1: UPLOAD YOUR COVID-19 CORE DATASET

Please create a new folder (wherever you prefer on your local and you can name it as you prefer) and put your COVID-19-core data (csv file) into this new folder.

**Important note 1: Please make sure your COVID-19 data file has the correct name (`msda_data_clinicians.csv` or `msda_data_patients.csv`).** The analysis task force had decided to analyze and interpret patient-reported and clinician-reported data separately for now. Therefore, if your data source combines records reported by clinicians as well as reported by patients, please split your cohort into two different ones:

- **`msda_data_clinicians.csv`: data source containing reports of clinicians**
- **`msda_data_patients.csv`: data source containing reports of patients**

The data should be called `msda_data_patients.csv` and/or `msda_data_clinicians.csv` and be contained in the folder you just created. As the name suggests, `msda_data_patients.csv` should contain your patient-reported data. While `msda_data_clinicians.csv` should contain your clinician-reported data.

**Important note 2:** Please make sure your data is properly transformed to the formats as described in the COVID-19 in MS core data dictionary (the most recent version of the COVID-19 in MS core data dictionary can be found on this [link](#). Please check out “a detailed dictionary describing the COVID-19 core dataset is provided” underneath Step 1: DATA COLLECTION).

**Important note 3:** there are two new variables in the covid19 dataset

**Introduce new variable “covid\_wave” in two ways:**

**Covid\_wave1** Have the registry categorize it themselves and provide us with the definition

**Covid\_wave2** using hard cuts (real first wave 1, real first wave 2)

`covid_wave2 = 1` when `covid19_date_suspected_onset` before 31st of May 2020

`covid_wave2 = 2` when `covid19_date_suspected_onset` after 1st of October 2020



## STEP 2: RUN THE PIPELINE

- Open your **Powershell/Terminal** and navigate the terminal to the folder you created and in which your data is located. You can use the 'cd' command to change the directories.
- Easiest way to find the directory path is to just type cd followed by a space and then drag and drop the location of the folder you created (and which contains the data) put between " " and press enter.

```
PS C:\Users\lucp10534> cd "G:\My Drive\federated architectures\MSDA Toolkit with UI - test LG"
PS G:\My Drive\federated architectures\MSDA Toolkit with UI - test LG>
```

- You can easily check where you are by typing 'pwd' and check the content by 'ls'.

```
PS G:\My Drive\federated architectures\MSDA Toolkit with UI - test LG> ls

Directory: G:\My Drive\federated architectures\MSDA Toolkit with UI - test LG

Mode                LastWriteTime         Length Name
----                -
-----          16/05/2021    04:28         25700 msda_data_clinicians.csv
```

- Copy and paste the below command to the terminal.

```
$ docker run -it -p 8080:8080 -v ${PWD}:/app/wwwroot/Scripts/Python/MSDA_Query3/Data
msdaalliance/covid19gdsi-ui
```

It may take some minutes for the first time to download the container and extract it; after pulling, you should see an output as demonstrated below.

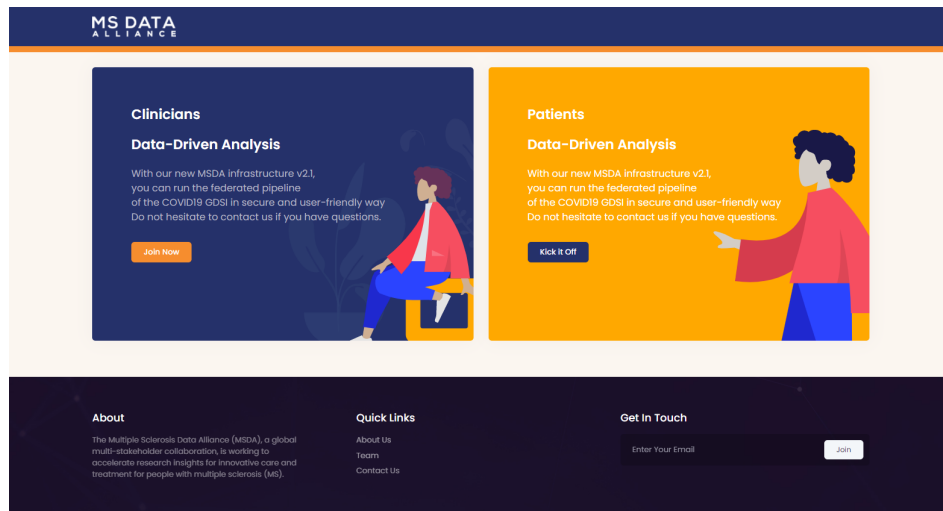
```
warn: Microsoft.AspNetCore.DataProtection.Repositories.FileSystemXmlRepository[60]
      Storing keys in a directory '/root/.aspnet/DataProtection-Keys' that may not be persisted outside of the container.
      Protected data will be unavailable when container is destroyed.
warn: Microsoft.AspNetCore.DataProtection.KeyManagement.XmlKeyManager[35]
      No XML encryptor configured. Key {dcdfee98-e38c-4c2f-b21c-2ee5cedced05} may be persisted to storage in unencrypted
      form.
info: Microsoft.Hosting.Lifetime[0]
      Now listening on: http://[::]:5000
info: Microsoft.Hosting.Lifetime[0]
      Application started. Press Ctrl+C to shut down.
info: Microsoft.Hosting.Lifetime[0]
      Hosting environment: Production
info: Microsoft.Hosting.Lifetime[0]
      Content root path: /app
```



## STEP 3: Interacting with UI

### *Navigate to browser*

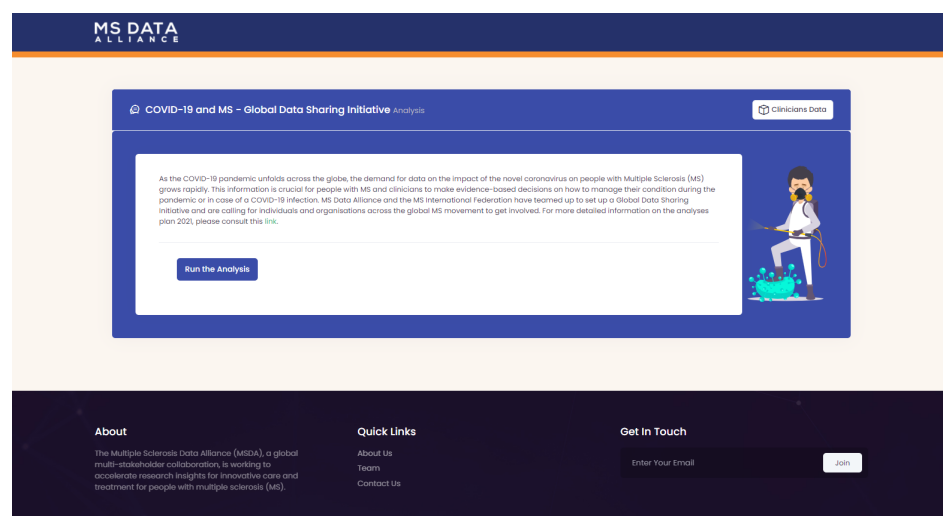
Open your preferred browser and navigate to **localhost:8080**. You will see the MSDA Toolbox landing page.



**NOTE:** You can always redirect to the landing page by clicking on the “MS DATA ALLIANCE” logo on the top left corner.

### *Type of analysis*

Select your type of analysis: patient-driven analysis, or clinician-driven analysis. After clicking on your desired type of analysis, the browser will redirect you to a new page. On this page, you should also check that you choose the correct type of analysis by checking out the upper right attribute.



### *Running the Analysis*



By clicking on the “Run the Analysis” button, the analysis will be executed locally. You can follow the descriptive logs by referring to your terminal. You should find the output of the different steps in your terminal, as shown below.

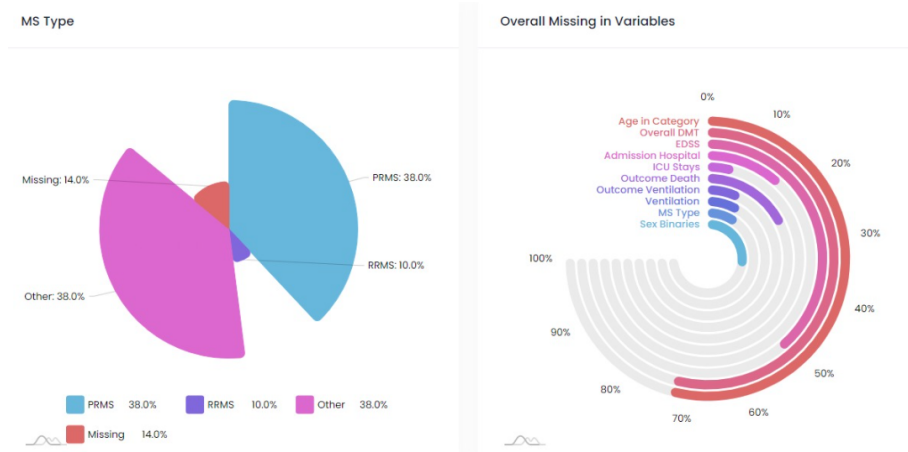
```
Clinician Data Driven Analysis have been started
Cleaning data ....
Unable to clean year_reporting as the column is not in the dataframe
Creating new variables ....
Computing tables ....
Done !
Zip file is ready to send
```

As soon as you get “Done!” in your terminal, the browser will redirect you automatically to the result page.

There will be two different sections that you can navigate through: Results and Charts.



These Results and Charts are solely for your personal use. The MSDA/MSIF will not be able to see your results and charts, because they don't have access to your local. Please feel free to check and interpret your own data by using these Results and Charts tools. The provided charts are interactive. You can try different features within the charts, e.g. within the Charts you can click on certain MS Type categories (in the legend), and in the Overall Missing in Variables chart you can drag your pointer.

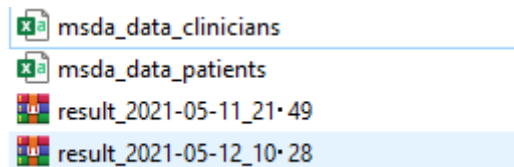


Depending on the chart type (e.g. bubble chart, bar chart,...) you can try different features. We recommend you to look into these charts so that you would be able to quickly notice possible errors in the data.



## STEP 4: Zipped Result

After running the analysis (described in 3-2-3-3), you should as well check the folder you created. The MSDA Toolkit will have taken care of zipping the result, and you should find the zipped result in the folder you are in (and which you created in step 3).



## STEP 5: Stopping the container

In your terminal, you can stop the container by pressing ( **Ctrl + C** ). This will terminate the container, which is recommended because you don't need it anymore unless you would want to run the container on top of your new core dataset. Additionally, it's better to terminate the container, because it consumes power and energy of your computer (within this container a linux operating system is launched with many other components: Python, web application, which is interpreting the results... ).

## STEP 6: Removing the container

For the next execution, we highly recommend removing the container after you did your analysis. In this way, we can avoid conflicts.

7-1. By entering this command, all the containers will be removed from your cache.

```
docker rm $(docker ps -a -q)
```

7-2. By entering this command, all the images will be removed from your cache.

```
docker rmi --force $(docker images -a -q)
```