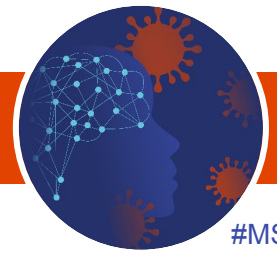


MANUAL

How to run the federated pipeline in 2021?



What has changed compared to 2020?



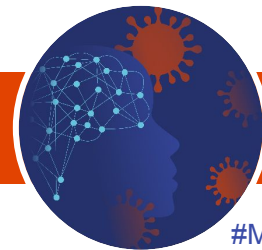
#MSCOVID19

Last year, we used dockerized Jupyter notebooks to run the federated pipeline of the GDSI (**MSDA infrastructure v1.0**). However, since then we have been working on improving the MSDA federated architecture (**MSDA infrastructure v2.1**) to incorporate some of the feedback, comments and suggestions we have received from you. The main differences between the MSDA infrastructure v1.0 and v2.1 are the following:

- ***We have reduced the “black-box” experience.*** More specifically, the MSDA infrastructure v1.0 resulted in a bunch of csv files that were very difficult to assess. In the MSDA infrastructure v2.1, we have worked on developing a *user-friendly end-user interface (UI)* that will allow you to walk through the pipeline more intuitively as well as to assess the results of the queries using a dashboard visualization of the key results.
- ***We have eliminated the need to exchange scripts using email, to reduce issues with “version control”.*** Last year, scripts were sent back and forth continuously, leading to the use of not-up-to-date scripts. Now, the script is shared via a [github repository](#), making it possible to ensure that you always have the latest version of the script.
- ***Increased security:*** using a new operating system ([Alpine](#)) and additional hardening
- ***Resource efficiency:*** reducing the size of the docker container with 91% (from 1.31 to 0.12 GB)
- ***Coding language agnostic:*** different coding languages are now possible (compared to only Python before)

More details about the MSDA infrastructure v1.0 and v2.1 are provided in [using this link](#). Next to this, [this video](#) provides you with a demo of the MSDA infrastructure v2.1. Because of these two main reasons listed above, we recommend you to use the MSDA infrastructure 2.1. However, if you prefer to stick to the old architecture, this is still possible.

Steps to follow

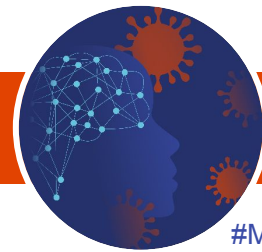


#MSCOVID19

1. Transform your data to the core dataset
2. Choose which infrastructure you want to use and run it
3. After checking the results, provide them to us
(edward.debrouwer@esat.kuleuven.be)

Consult [this link](#) for more information on the analyses plan 2021

Step 1: transform your data to the core dataset



#MSCOVID19

Similarly as last year, run your transformation code to transform the [covid19 core dataset](#).

Important note: there are 2 new variables in the covid19 core 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)

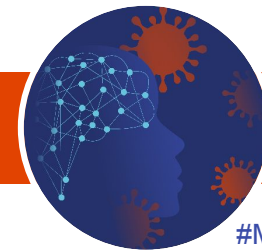
covid_wave2 = 1 when covid19_date_suspected_onset before 31st of May 2020

covid_wave2 = 2 when covid19_date_suspected_onset after 1st of October 2020

We are working with these two different variants of the covid_wave because in our analyses round 1 we are exploring what is the best approach

- covid_wave1 does not reduce the number of records, but might results in reviewer and community comments on the intuitive categorization and heterogeneity across the cohorts
- covid_wave2 could reduce the number of records, but provides us with a more clearer “cut” between what can be referred to a the ‘real first wave’ and the ‘real second wave’ at global scale.

Step 2: Choose which infrastructure you want to use and run it



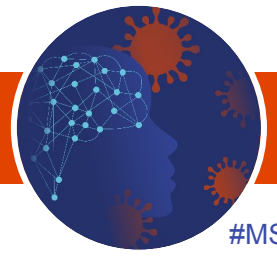
#MSCOVID19

First : [Initial set - up \(manual\)](#)

Secondly, when choosing the [MSDA infrastructure v1.0 \(manual\)](#)

Secondly, when choosing the [MSDA infrastructure 2.1 \(manual\)](#)

Step 3: Check the results and provide them to us



#MSCOVID19

Edward.debrouwer@esat.kuleuven.be

THANK YOU

MS DATA
ALLIANCE

acting under the umbrella
of the European
Charcot Foundation

