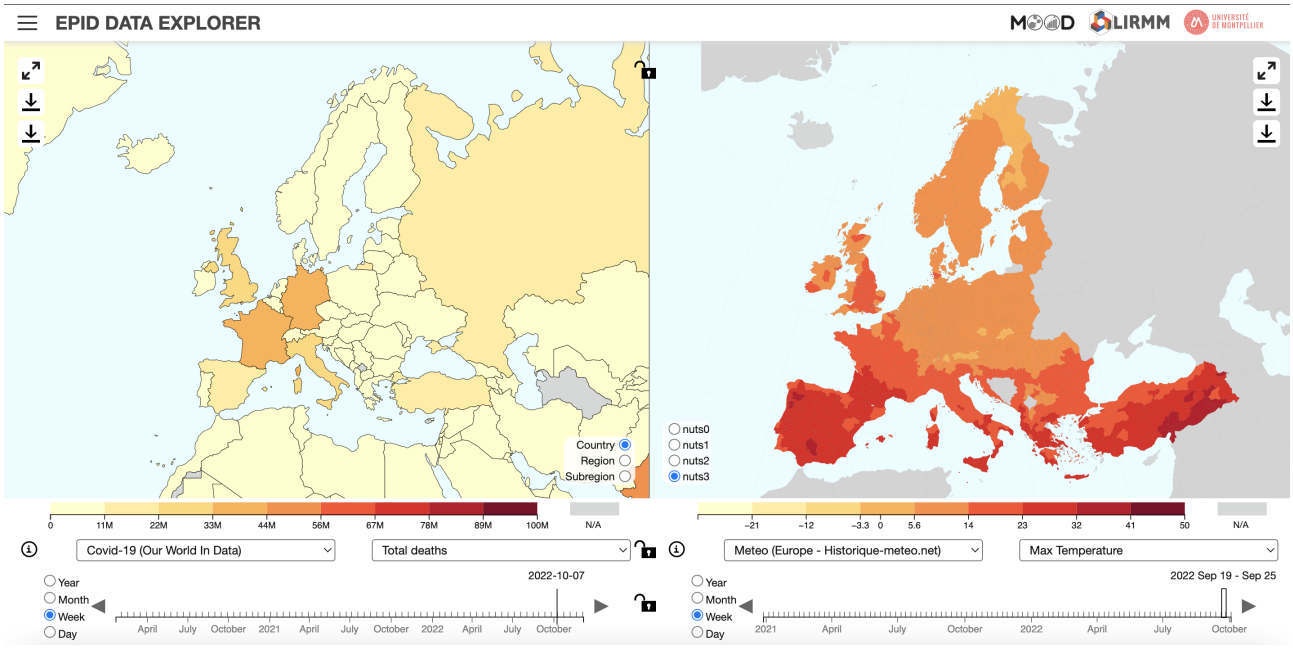
# The Epid Data Explorer Platform




## LIRMM-UM

Version 1.0 - January 2023

Contributors [1]: Jérome Azé, Fati Chen, Pierre Pompidor, Pascal Poncelet, Vincent Raveneau, Nancy Rodriguez, Arnaud Sallaberry, Laetitia Viau.

[1] In alphabetical order.

# Contents


**1 EDE: A General Overview** **3**
- 1.1 The Global and Two-Part Views . . . 4
- 1.2 The Main Components of the Global View . . . 4
- 1.3 The Main Components of the Two-Part View . . . 10

**2 Installation of EDE** **12**
- 2.1 Download and Install EDE . . . 12
- 2.2 Starting EDE . . . 13
  - 2.2.1 Using The Epid Data Explorer - Setup Interface . . . 13
  - 2.2.2 Running EDE Without the Setup Interface . . . 14

**3 Installation of a Data Store for EDE** **16**
- 3.1 Download and Install the Data Store . . . 16
- 3.2 Start the Data Store . . . 16
  - 3.2.1 Using The Epid Data Explorer - Dataset Setup Interface . . . 16
  - 3.2.2 Installing a Data Store Without the Setup Interface . . . 21

# Introduction

The Epidemic Data Explorer (**EDE**) is a user-friendly platform that facilitates the monitoring and analysis of the progression of an epidemic over time and location by leveraging aggregated data. It enables users to perform cross-regional and cross-temporal analyses of epidemiological data, as well as covariate data (e.g. weather data).

The platform provides a comprehensive set of features that enable users to analyze data across various geographical regions and time frames. It offers a range of capabilities, including the ability to view data at different levels of temporal resolution, zoom in on specific regions for closer inspection, compare the behavior of a particular region with the entire dataset, download selected data, and compare two indicators[2] from different datasets on the same interface.

EDE supports common geocoding standards, such as NUTS (2021 revision) and ISO 3166 (2020 revision) codes, making it easy to integrate data from various sources. The platform accepts data in the form of a csv file, including formats used by organizations like ECDC and Our World in Data, as well as custom formats defined by the user.

EDE offers two options for use: through the Mood platform or on a local setup (on a personal computer or with personal data).

- The Mood platform can be accessed at:

  `https://advanse.lirmm.net/EDE/epid-data-explorer/`

- EDE can be downloaded and installed on a personal computer by accessing the code available at:

  The user interface for displaying the data:

  `https://gite.lirmm.fr/advanse/EDE/epid-data-explorer`;

  The data store that provides the data:

  `https://gite.lirmm.fr/advanse/EDE/dataset`

The objective of this report is to showcase the various features and capabilities of EDE and to provide instructions on how to install and set up the platform. The report is structured as follows: Chapter 1 provides a comprehensive overview of the key components of EDE, while Chapter 2 offers detailed instructions on how to install and configure EDE on a personal computer, including how to add new data to the platform. Chapter 3 describes how to create and manage a data store for EDE.

**Note:** This report complements the READMEs that can be found in the git repositories. It is advised that users also refer to the READMEs for additional information.

[2] In the rest of this report, the terms "variable" or "indicator" will be used interchangeably to refer to elements that have a value that can change over time, such as "the number of cases" for a disease.

# Chapter 1

# EDE: A General Overview

Upon accessing the platform, users are greeted with the home screen (as shown in Figure 1.1), which displays a list of available datasets [1]. The list on the left showcases all available datasets, while the "Recommended Groups" section on the right offers data sets organized by topic of interest.

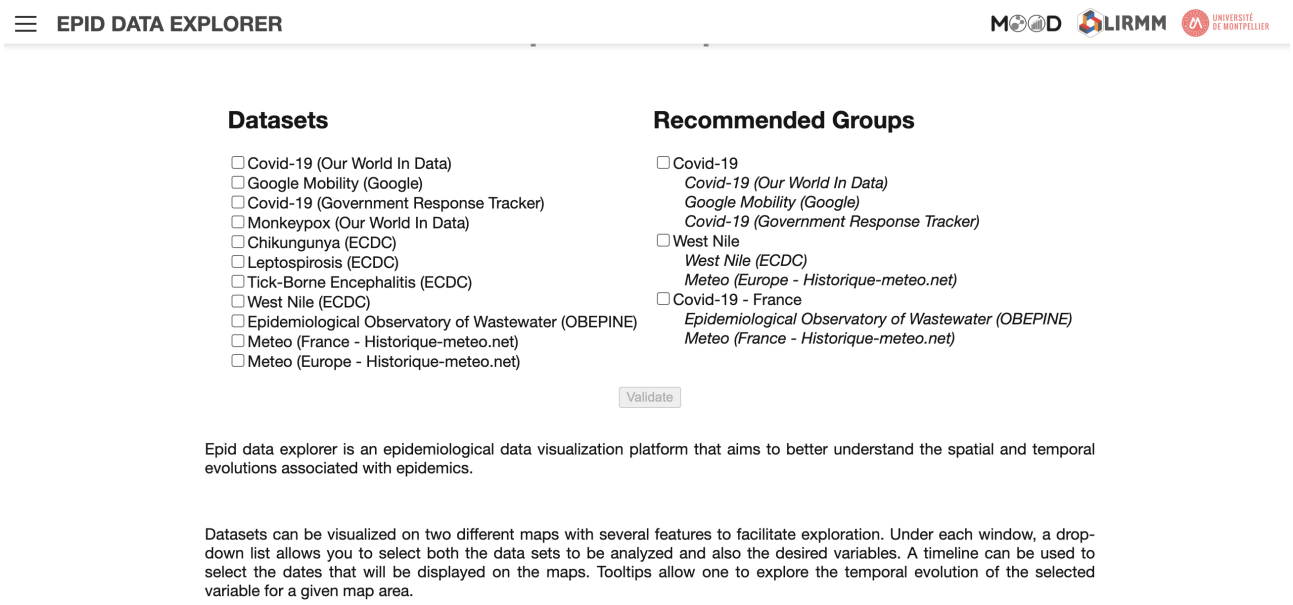



Figure 1.1: A screenshot of the EDE homepage

To choose a dataset, the user just needs to click on the corresponding checkbox in either the datasets list or the "Recommended Groups" section.


Figure 1.2: Selection of datasets

For instance, Figure 1.2 shows the data selection process when selecting the "Covid-19" group. As seen in the illustration, the related datasets in the left column are automatically selected.

[1] The data used to demonstrate the features of EDE is sourced from ECDC (`https://www.ecdc.europa.eu/`), Our World in Data (`https://ourworldindata.org`), Google (`https://www.google.com/covid19/mobility/`), Government Response Tracker (`https://www.bsg.ox.ac.uk/research/covid-19-government-response-tracker`), Obepine (`https://www.reseau-obepine.fr`) and Historique meteo (`https://www.historique-meteo.net`).

## 1.1 The Global and Two-Part Views

EDE offers two types of visualizations after datasets are selected. The first, shown in Figure 1.3, gives a comprehensive view of the dataset, allowing the user to explore it from both temporal and geographical perspectives. The second, shown in Figure 1.4, is a two-part view that enables the user to track and compare various indicators from both temporal and geographical perspectives.

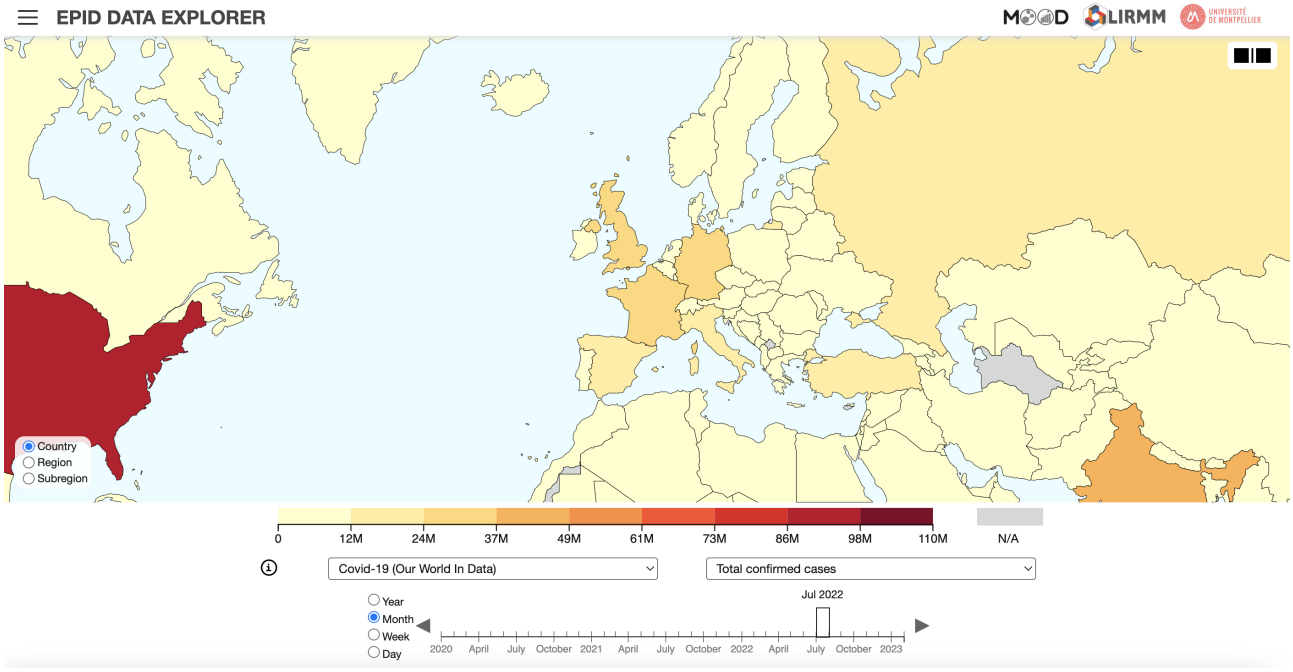



Figure 1.3: The global view

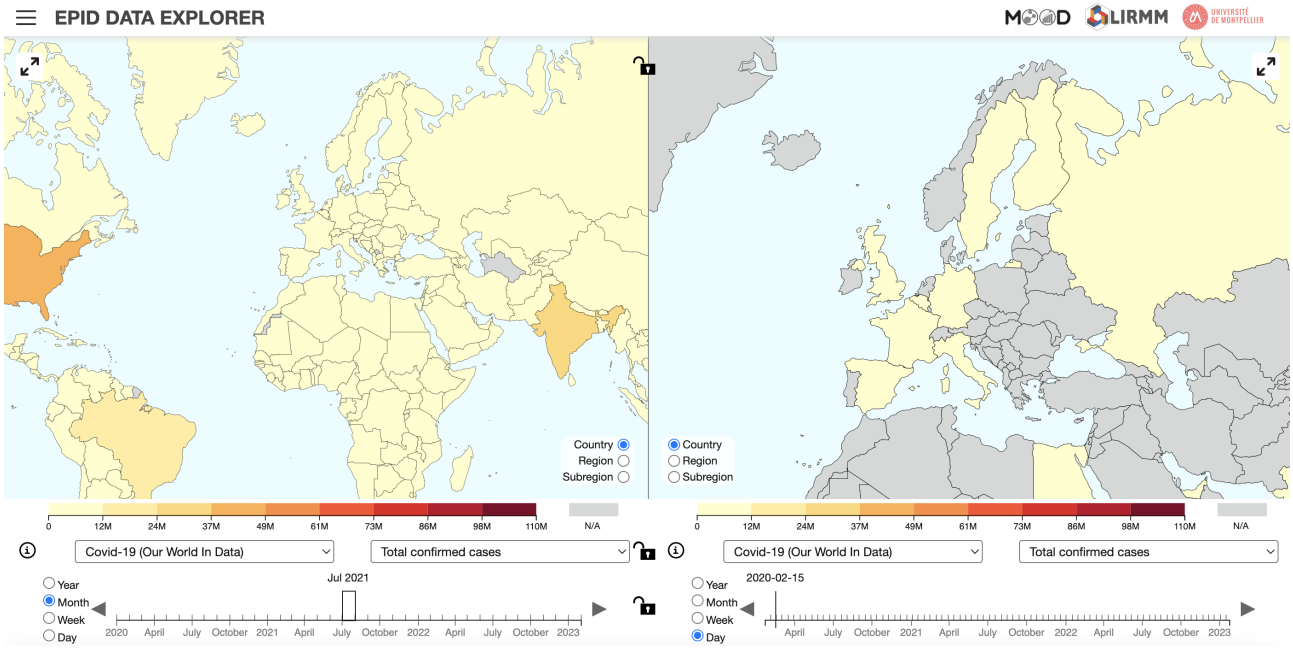



Figure 1.4: The two-part view

## 1.2 The Main Components of the Global View

The main components of the global view are illustrated in Figure 1.5.

The EDE platform provides users with a list of available datasets (as shown in Figure 1.5-ⓐ) to select from. By clicking on the information icon (Figure 1.5-ⓑ), users can access additional information such as the

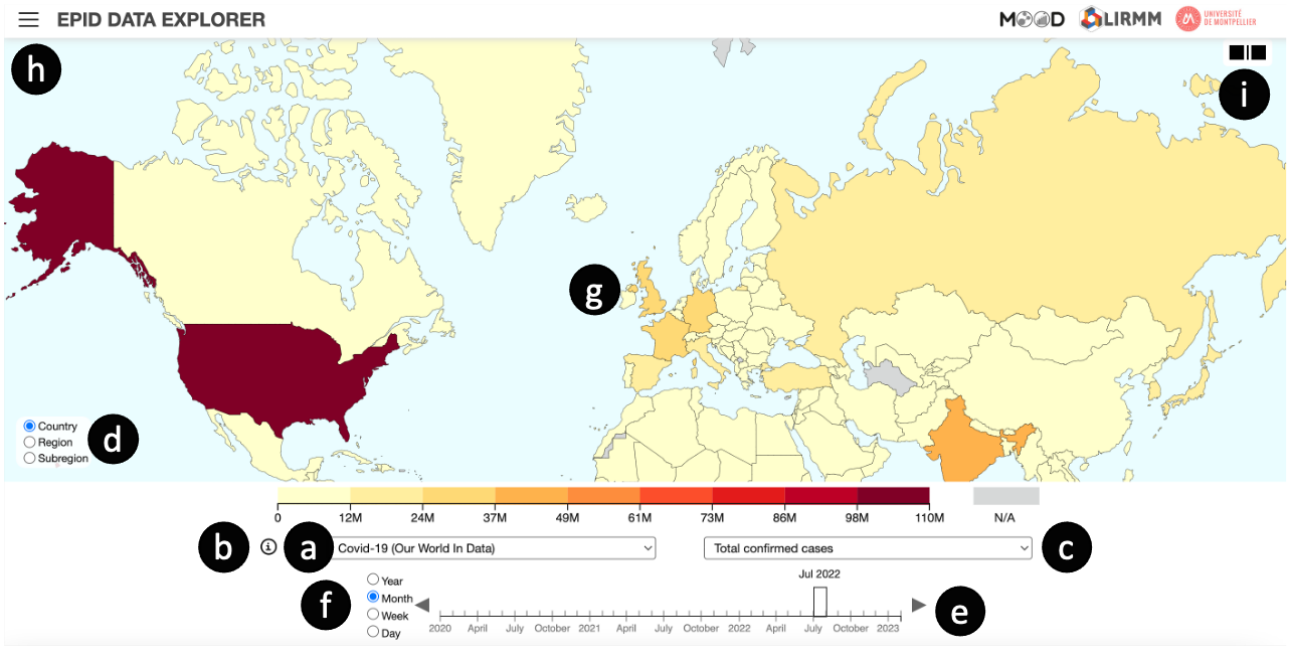



Figure 1.5: The main components of the global view

dataset's description and URL.

Once a dataset is selected, users can easily choose the desired indicator to track by selecting from a list of available options (Figure 1.5-ⓒ). In the example shown in the figure, the "Total Confirmed Cases" indicator is selected from the "Covid-19" dataset from "Our World in Data".

To adjust the geographic granularity of the map, users can select different levels of detail (Figure 1.5-ⓓ). The data from Our World in Data is based on ISO 3166, which enables users to switch between viewing data at the country to subregion level. If the original data is in NUTS format, the granularity can also be adjusted to display data at NUTS0, NUTS1, NUTS2, and NUTS3 levels, as demonstrated in Figure 1.6.

It is important to note that the level of geographical granularity displayed on the map is dependent on the granularity of the data. The map can only display values at the level provided in the dataset. For example, if the data uses ISO 3166 geocoding, the granularity will range from country to subregion, and if the data uses NUTS geocoding, it can range from NUTS0 to NUTS3.

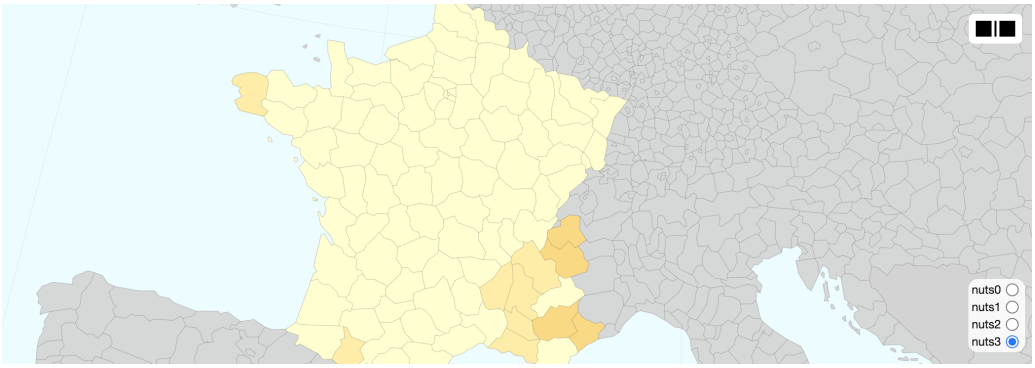



Figure 1.6: An example of the global view with NUTS geocoding

EDE makes it easy to navigate through time with its timeline feature. The map updates automatically to display values for the selected indicator as shown in Figure 1.5-ⓔ. The user can adjust the temporal granularity, ranging from daily to yearly, as shown in Figure 1.5-ⓕ. The minimum granularity level is determined by the granularity of the dataset. The system also automatically aggregates data from the lowest level (day) to the desired higher level, updating the legends accordingly, as demonstrated in Figure 1.7.

The map in Figure 1.5-ⓖ enables easy navigation and exploration of different regions and provides a more detailed view of a specific area. It displays the value of the selected indicator for the selected time and geographic

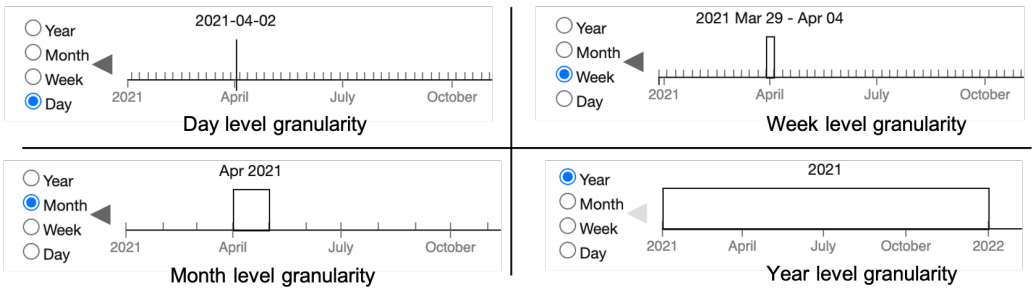



Figure 1.7: The different time granularities

location. The legend below the map clearly displays the values of the indicator and adjusts automatically based on the selected data type. EDE is capable of handling both ordinal and quantitative data types, as shown in Figure 1.8 where the examples of legends for continuous quantitative data (top two) and ordinal qualitative data (bottom one) are displayed.

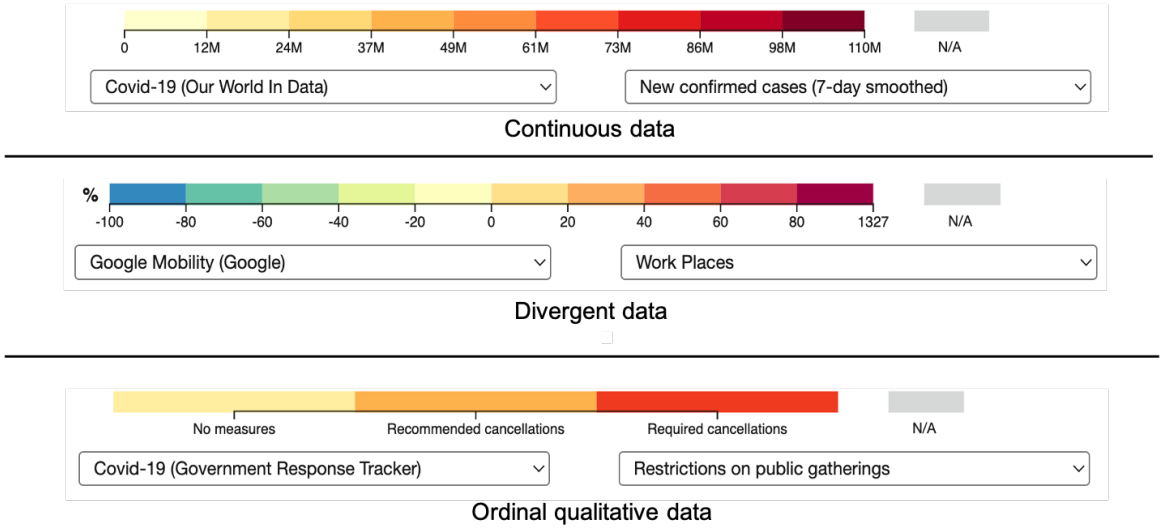



Figure 1.8: Examples of ordinal qualitative data and continuous quantitative data

EDE generates the legend by analyzing the dataset to determine the minimum and maximum values, and then dividing them into ranges for optimal distribution within the visualization. As shown in Figure 1.9, the legend adapts to the values of the selected indicator, using data from Our World in Data since the start of the Covid outbreak. For instance, the legend for "Total Confirmed Cases" ranges from 0 to 100 million, while that for "New Confirmed Cases (7 days smoothed)" ranges from 0 to 900 thousand cases.

With EDE's legend, users can easily find specific values on the map by clicking on it. This action will highlight the corresponding geographical regions on the map. As shown in Figure 1.10, when clicking on the legend for "Total confirmed cases" with values between 98 million and 110 million, the associated geographical areas will be highlighted on the map.

**Notes**:

- The aim of EDE is to allow users to easily navigate through the data over time and view the corresponding indicator values on the map. The values displayed are based on the data set's period, such as the minimum and maximum values and discretization. As a result, when navigating with the timeline, certain geographical areas may not appear in relation to specific values in the legend. For instance, as demonstrated in Figure 1.11, using Our World in Data and "New Confirmed Cases (7 days smoothed)" as the indicator, the legend only shows values between 0 and 100K in November 2021. This means that within the dataset, for each value in the legend, there is a period and location where the indicator reached that value. Users can explore the time and geography to find these periods and areas. For example, for the periods associated with 800K-900K, we can see that "New Confirmed Cases (7 days smoothed)" appear in North America in January 2012, as shown in Figure 1.12.
- It's worth noting that the values displayed in the legend are based on the highest level of geographical

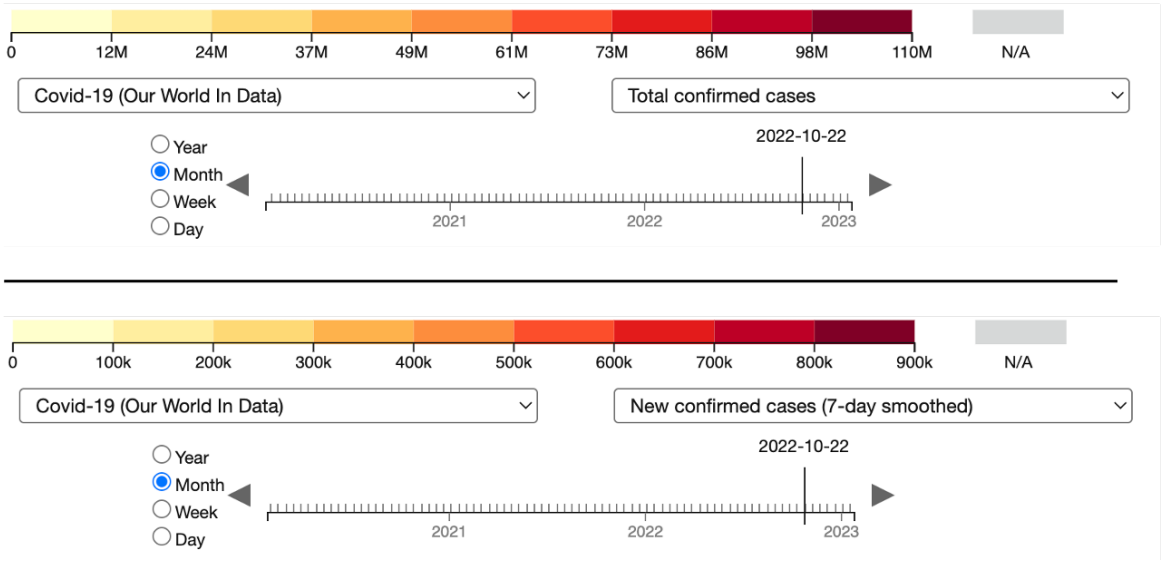



Figure 1.9: An example of the adoption of the legend according to the indicator

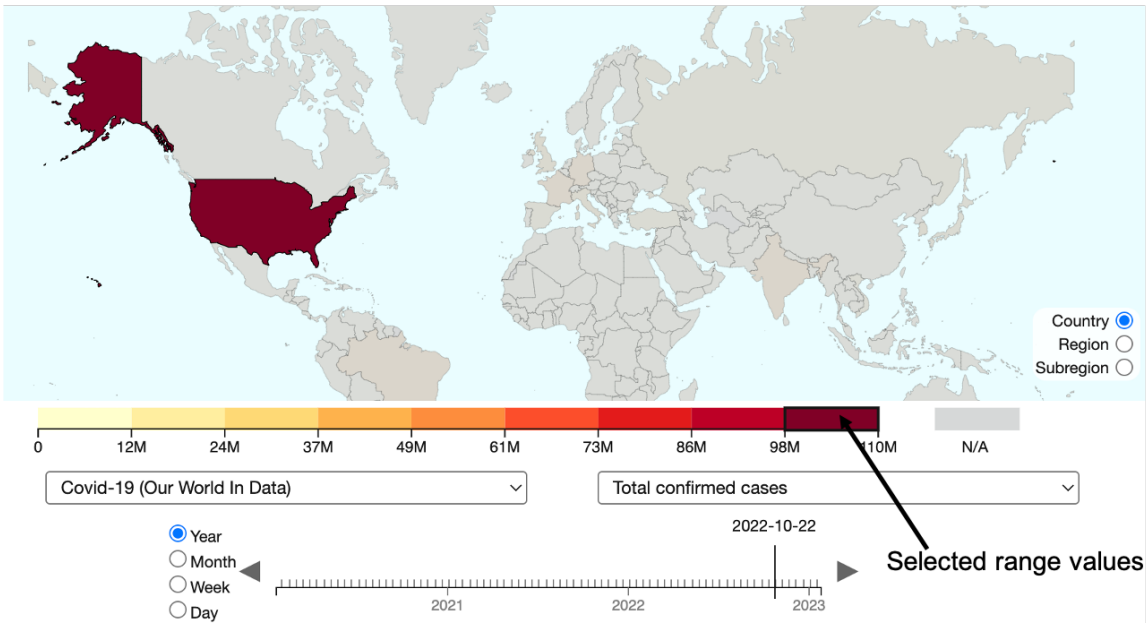



Figure 1.10: Geographic areas highlighted by selecting a part of the legend

granularity available in the data. If the data is available at different levels of granularity, such as NUTS3 to NUTS0, the values displayed will be at the highest level (NUTS0).

Navigating the map in EDE allows the user to easily access detailed information about a specific area. Simply click on an area to view its information in a tooltip. The tooltip can be repositioned around the map and its size can be adjusted by dragging its window borders, as demonstrated in Figure 1.13.

The tooltip in Figure 1.14 showcases the key elements of this feature. It displays the values of the selected indicator for the chosen area throughout the entire dataset period. In our example we selected "New Confirmed Cases". if the user wants to view data for a different indicator, the information in the tooltip will be updated automatically. To provide context, a vertical line is displayed in conjunction with the timeline, highlighting the date being viewed (as shown in Figure 1.14-ⓐ). The line also displays dots for key statistical measures such as the maximum, 3rd quartile, median, 1st quartile, and minimum values.

Users can navigate the tooltip to view values in different time periods. While navigating, another line displays information about the values (as seen in Figure 1.14-ⓑ). The tooltip also includes information about the indicator for the selected country or area, as well as for all geographic zones (maximum, minimum, and quartile values), to provide context (Figure 1.14-ⓒ). To further help the user understand the selected geographical area's indicator value in relation to others, the tooltip displays global values for all areas in blue (Figure 1.14-ⓓ).

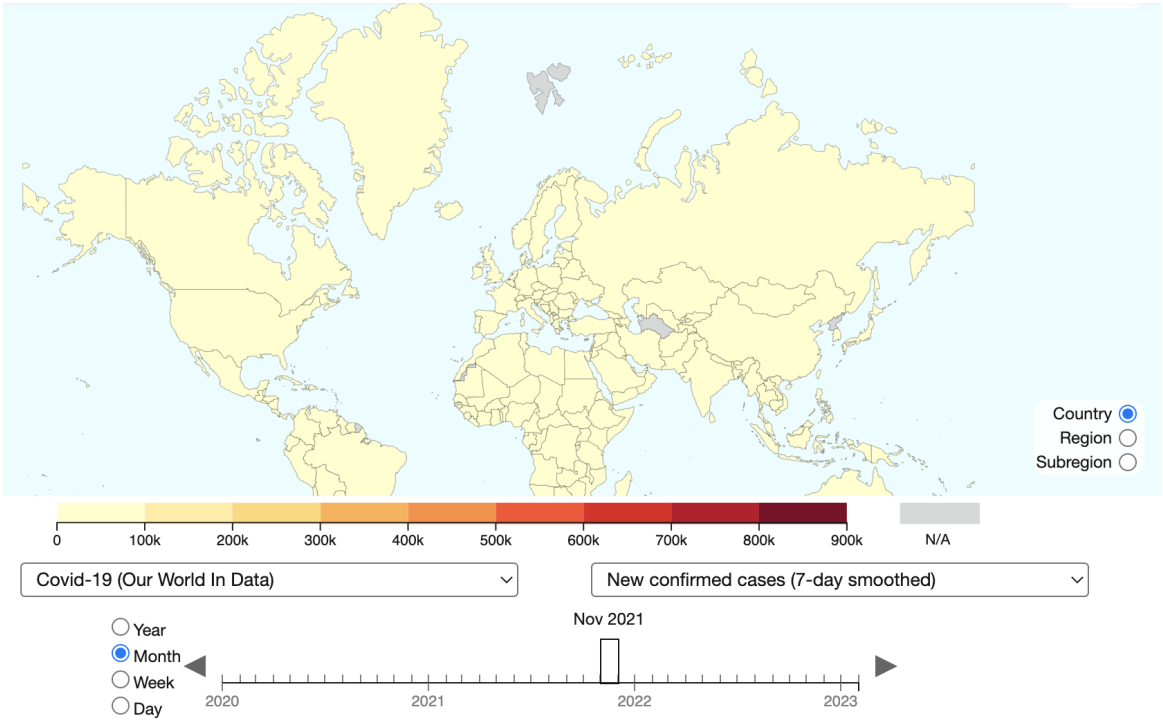



Figure 1.11: An example of an indicator where not all possible values of the legend are represented in the map

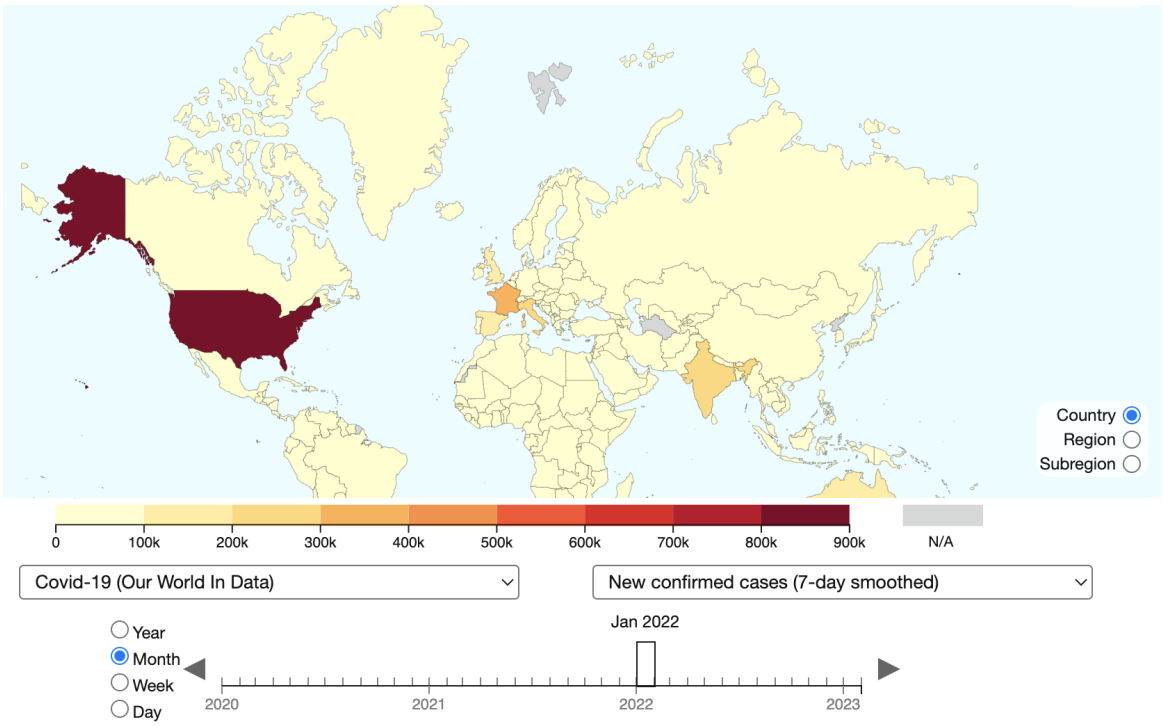



Figure 1.12: A location and period corresponding to the largest value between 800K and 900K for "New Confirmed Cases (7 days smoothed)"

The tooltip feature in Figure 1.14 provides various ways to analyze data. The user can select different options, as indicated by the switches in Figure 1.14-ⓔ, to tailor the displayed chart:

- The **dataset/area** switch allows the user to choose the data range displayed on the ordinate axis. The user can view the entire dataset or select a specific area of interest.
- The **linear/log** switch enables the user to switch between a linear and logarithmic scale.
- The **7-day avg.** option provides a smoothed curve by calculating the average value over a 7-day sliding

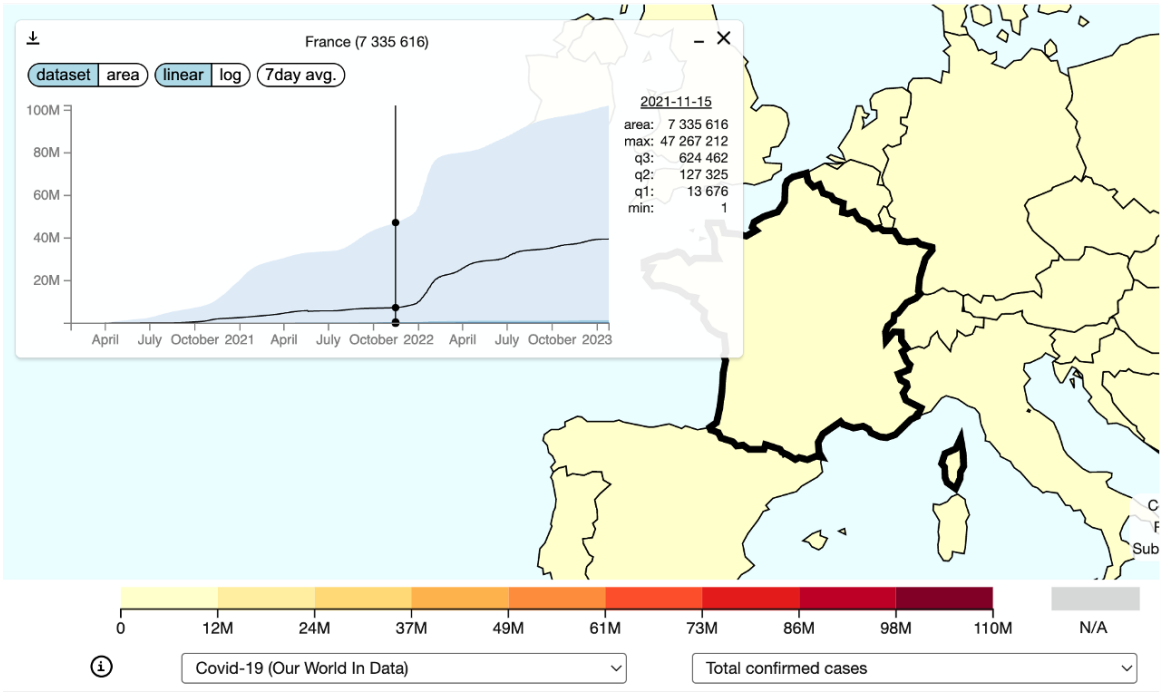



Figure 1.13: An Example of a tooltip when clicking on France

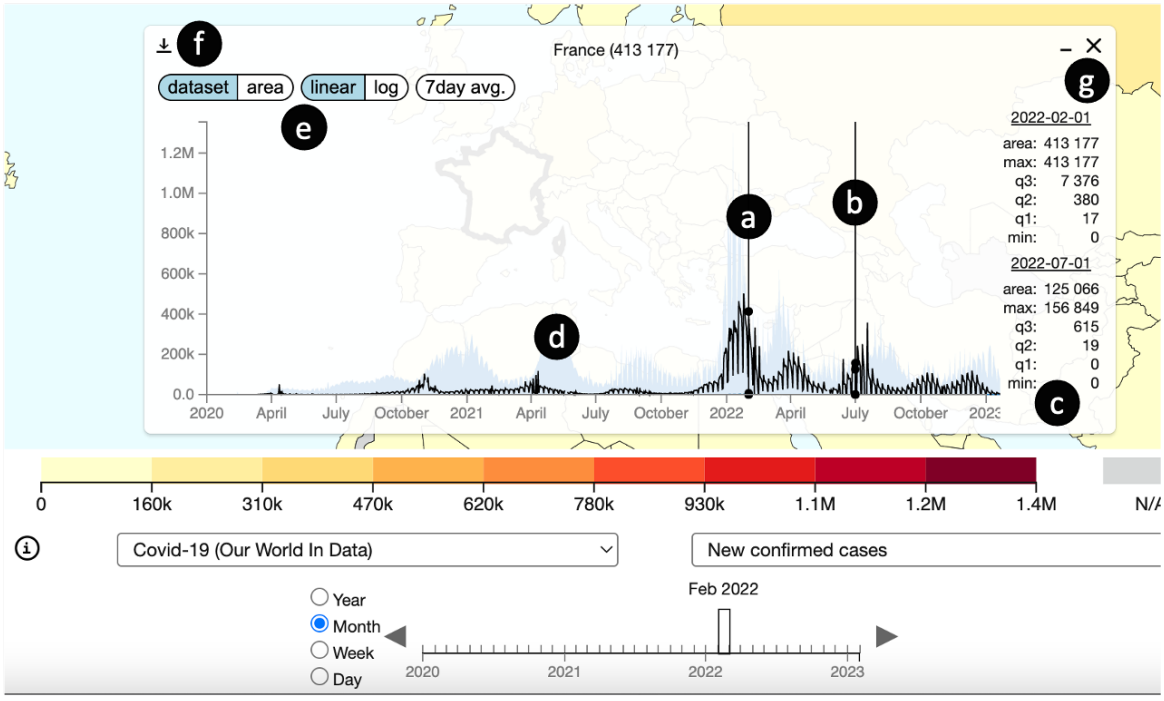



Figure 1.14: The main components of the tooltip

window.

If users want to conduct further analysis on the selected data, they can easily download it in csv format by clicking the download icon (see Figure 1.14-ⓕ). The downloaded data will include the area name and the indicator.

The tooltip can be closed by clicking the close icon (see Figure 1.14-ⓖ).

For help with the functionality of EDE, users can access assistance and visual aids by clicking the help icon located in the top left corner of the screen (see Figure 1.5-ⓗ).

For comparison purposes, the user can switch to the two-part view by clicking the switch icon at the top right corner of the map (see Figure 1.5-ⓘ).

FRA_new_cases

| id | date | new_cases |
|---|---|---|
| FRA | 2020-01-24 | 2 |
| FRA | 2020-01-25 | 1 |
| FRA | 2020-01-26 | 0 |
| FRA | 2020-01-27 | 0 |
| FRA | 2020-01-28 | 1 |
| FRA | 2020-01-29 | 1 |
| FRA | 2020-01-30 | 0 |

Figure 1.15: An example of exported csv file

## 1.3 The Main Components of the Two-Part View

The global view of EDE allows the user to explore and observe the evolution of a selected indicator over space and time. On the other hand, the two-part view enables the user to compare values of different indicators. This view facilitates comparison of different time periods and geographical regions, as well as the ability to compare indicators or indicators in relation to other covariates (e.g. the spread of the West Nile Virus in relation to weather).

As shown in Figure 1.16-ⓐ, the two-part view consists of two parts, each of which offers the same functionalities as in the global view (such as selection of the dataset, indicator, timeline, map, tooltip, etc.) to facilitate comparisons. When comparing data, the user must consider three main parameters: location, time, and indicator. This task can be challenging as these parameters vary from dataset to dataset. To help the user, EDE provides the option to lock in one, two, or three of these parameters, allowing the user to focus on their areas of interest. This is achieved by using padlocks on the platform (as seen in Figure 1.16 ⓑ, ⓒ, and ⓓ).

The three padlocks (see Figure 1.16-ⓑ, ⓒ, and ⓓ) allow for easier analysis of the visualizations by synchronizing or comparing the maps. The padlock between maps synchronizes navigation, the padlock on the datasets displays the same variable on both maps, and the padlock on the timeline displays data at the same date.

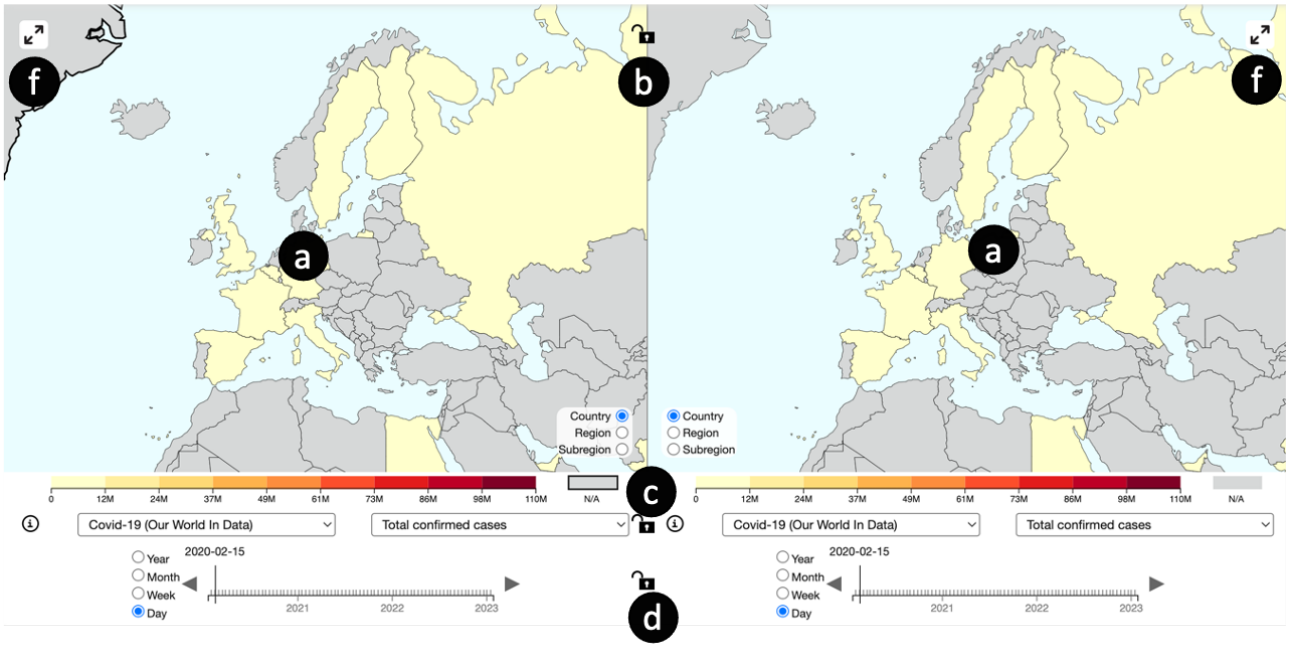



Figure 1.16: The main components of the two-part view

As an example [2], consider users who want to track the "New Confirmed Cases" indicator globally (see Figure 1.17). They start by selecting a specific region on the map and locking it in place. Then, they select the desired indicator and lock it in place as well. This automatically adjusts the map and indicator on both maps, enabling the users to navigate through different time periods and observe the evolution of the chosen indicator in the selected region. In another scenario, users may want to examine the correlation between Covid rates in wastewater and rainfall in France (see Figure 1.18). To do so, they simply select France on the map, lock it in place, choose the time period, and then observe the evolution of both indicators on the map during the selected period.

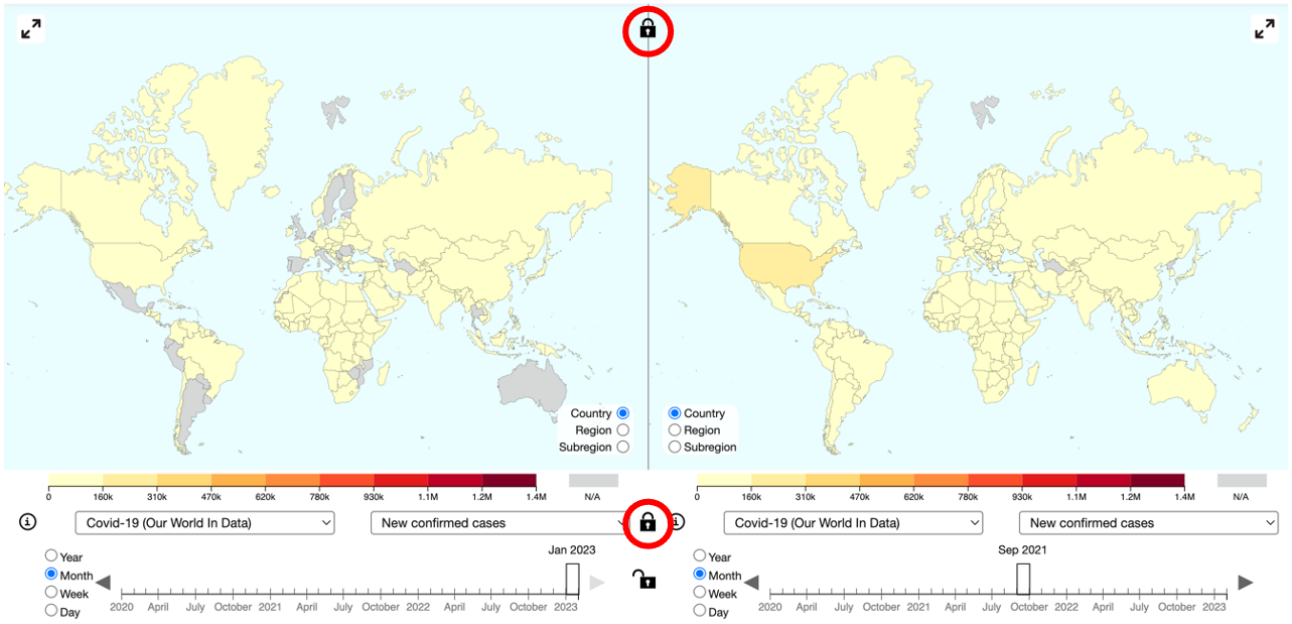



Figure 1.17: An illustration of a user interested in the same indicator on the same geographical area

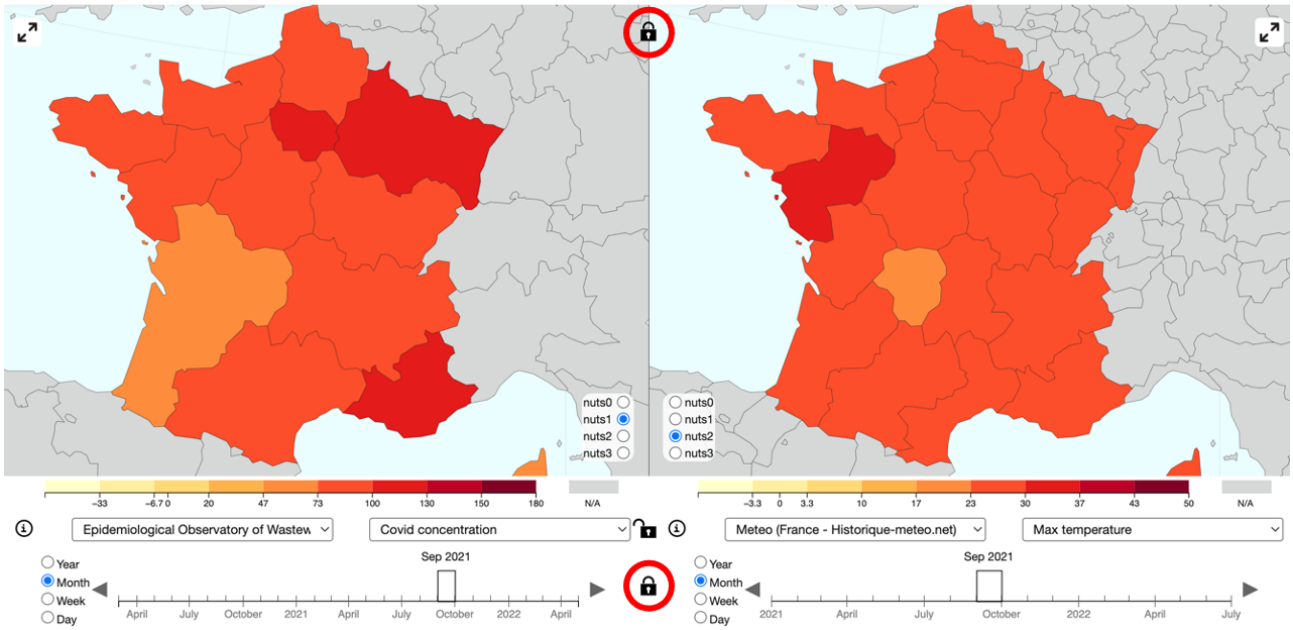



Figure 1.18: An illustration of a user interested in different indicators, on the same periods on the same geographical area

[2]Please note that the examples given are fictional and are used for illustrative purposes only.

# Chapter 2

# Installation of EDE

EDE is a platform developed using Javascript and requires data to be in Json format. To ease deployment and ensure data privacy, it has been designed to access data from data servers. This allows users to access data from their own computer, their organization, or from public sources, such as the platform's demonstration data obtained from the data server at `https://advanse.lirmm.net/EDE/dataset-advanse`.

EDE can be installed on a user's personal computer (see Figure 2.1-ⓐ) or a computer within their organization (see Figure 2.1-ⓑ).

Users can access external public data (as shown in Figure 2.1-ⓒ), data on their own computer (Figure 2.1-ⓓ), or private data from their organization with the appropriate rights and access (e.g. via VPN, as shown in Figure 2.1-ⓓ).

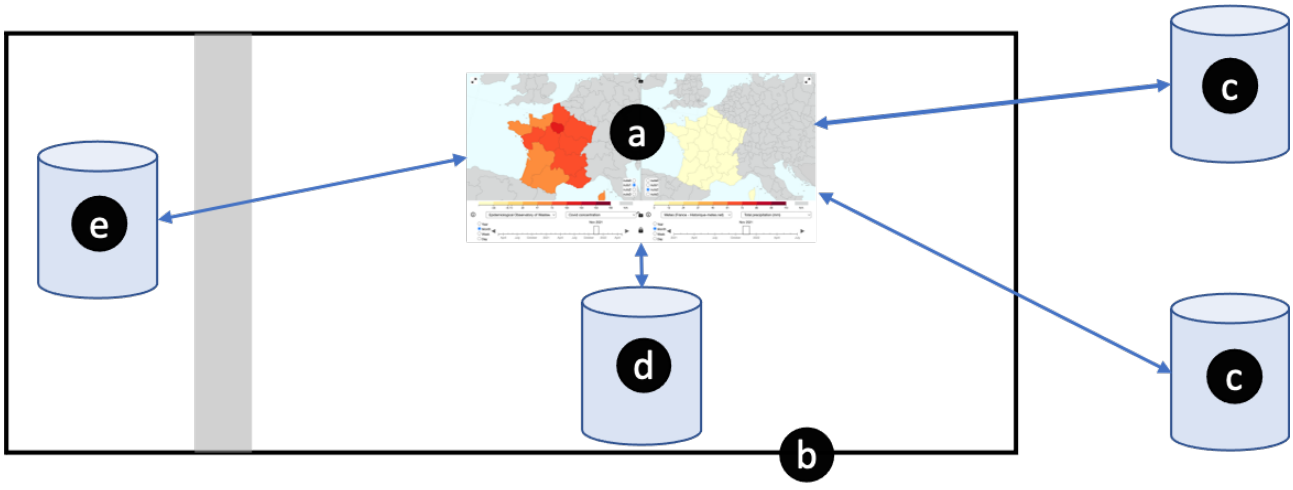



Figure 2.1: EDE and servers of datasets

In this section we mainly focus on the EDE installation. Installation of data servers is discussed in Chapter 3.

The package can be downloaded at the following URL:

`https://gite.lirmm.fr/advanse/EDE/epid-data-explorer`

It contains the sources as well as executable codes (Linux, Windows, MacOS) to easily install the interface.

## 2.1 Download and Install EDE

After accessing `https://gite.lirmm.fr/advanse/EDE/epid-data-explorer`, the user can download the folder either by using the download option or by cloning the git repository.

In the rest of the report, we assume that the user has installed EDE in the "**MyEDE**" directory.

If the user has downloaded a compressed file, they can simply extract it in the MyEDE directory. If they have cloned the git repository, they can open a terminal in the MyEDE directory and run the following command:

`git clone https://gite.lirmm.fr/advanse/EDE/epid-data-explorer.git`.

In either case, a directory named "epid-data-explorer" will be created in the MyEDE directory.

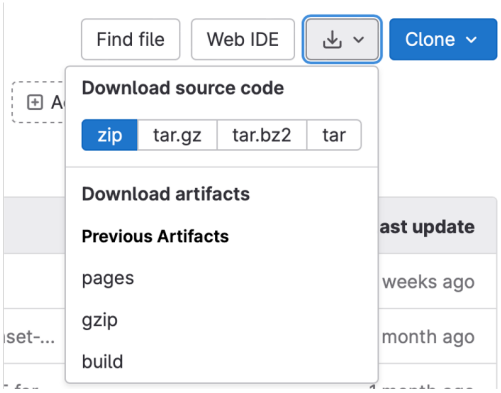



Figure 2.2: Download and install EDE

## 2.2 Starting EDE

EDE is a web-based application, so it requires a server to be running. To make the process easier for the user, a graphical setup tool interface, called **Epid Data Explorer - Setup**, is provided for Linux, Windows, and MacOS. This interface not only allows the user to activate the server, but also manage (add, modify the display order, and remove) the data servers so they can be viewed.

For Linux (resp. Windows), it suffices to run the executable file "EDE Dataset Setup" (resp. EDE Dataset Setup.exe).

**MacOS specificity:** On MacOS, there are some challenges in installing the interface. To launch the interface, the user must open a terminal, navigate to the directory "MyEDE/epid-data-explorer", and run the command: `python3 gui/main.py`.

**Important note: It is important to note that the "Epid Data Explorer - Setup" window must remain open, as closing it will result in the shutdown of the EDE user interface instance. The main reason is that the interface launches the server which allows access to the data. However, for experienced users, there is an option to launch the server without the interface (C.f. Section 2.2.2).**

### 2.2.1 Using The Epid Data Explorer - Setup Interface

The Figure 2.3 shows the screen when you launch the Epid Data Explorer - Setup interface. You can choose the port for the server (see Figure 2.3-ⓐ) and start the server (see Figure 2.3-ⓑ). A data server is available by default at `https://advanse.lirmm.net/EDE/dataset-advanse"`. To access EDE, you can either click on the interface button or directly enter the address in your browser. For example, if the selected port is 8001, the address would be `http://localhost:8001`. Figure 2.4 shows an example of EDE running on a localhost using data from the Mood server.

The "Manage Data Sources" tab, as shown in Figure 2.5, gives the user the ability to manage their data sources. This includes deleting existing data servers, adding new ones by entering the URL (using the "Add Data Source" button), and changing the order of servers. These modifications will automatically appear in the visualization when the page in the browser is refreshed.

**MacOS specificity:** On MacOS, there are some challenges in installing the interface. To launch the interface, the user must open a terminal, navigate to the directory "MyEDE/epid-data-explorer", and run the command: `python3 gui/main.py`.

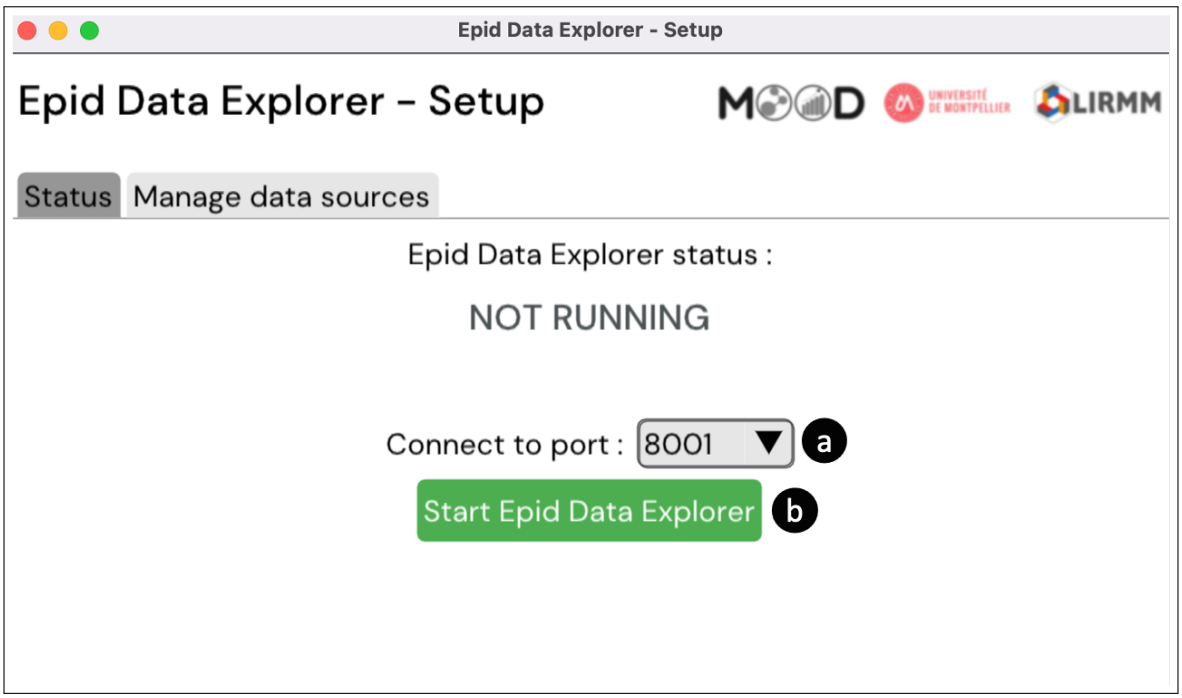



Figure 2.3: The Epid Data Explorer - Setup interface - the Status tab

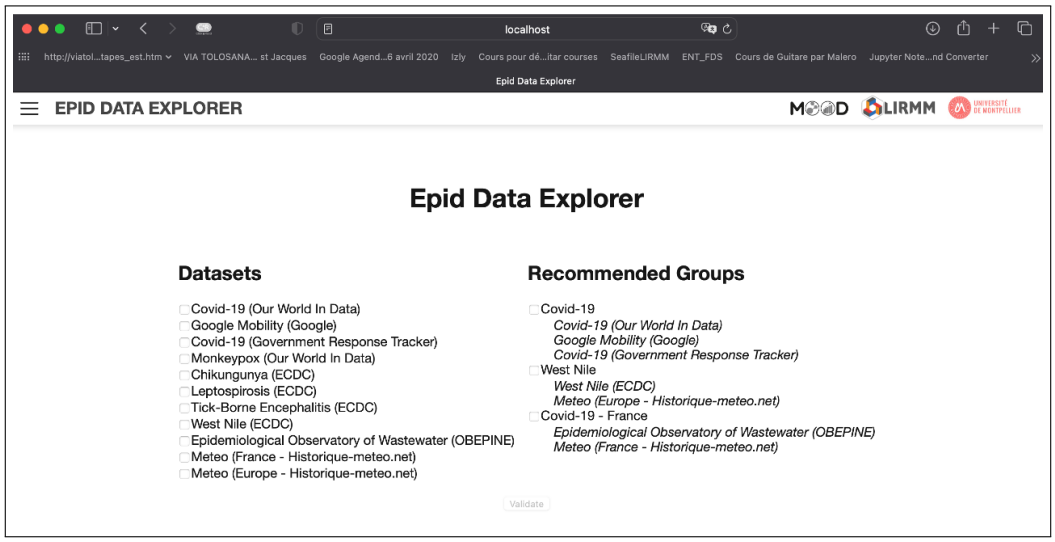



Figure 2.4: The Epid Data Explorer platform running on localhost

### 2.2.2 Running EDE Without the Setup Interface

For advanced users, it is possible to bypass the use of the graphical setup interface. This section provides a brief overview of the main components and updating procedures for experienced software installers. For more in-depth information, refer to the README file in the git repository.

To start the server, navigate to the "MyEDE/epid-data-explorer/gui" directory and launch python3. Then, import the server file, set the desired port number, and start the server. The following example shows how to set the port number to 8010:

```
...MyEDE/epid-data-explorer/gui % python3
Python 3.9.12 (main, Apr 5 2022, 01:53:17)
[Clang 12.0.0 ] :: Anaconda, Inc. on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>>import server
>>>server.PORT = 8010
>>>server.start()
```

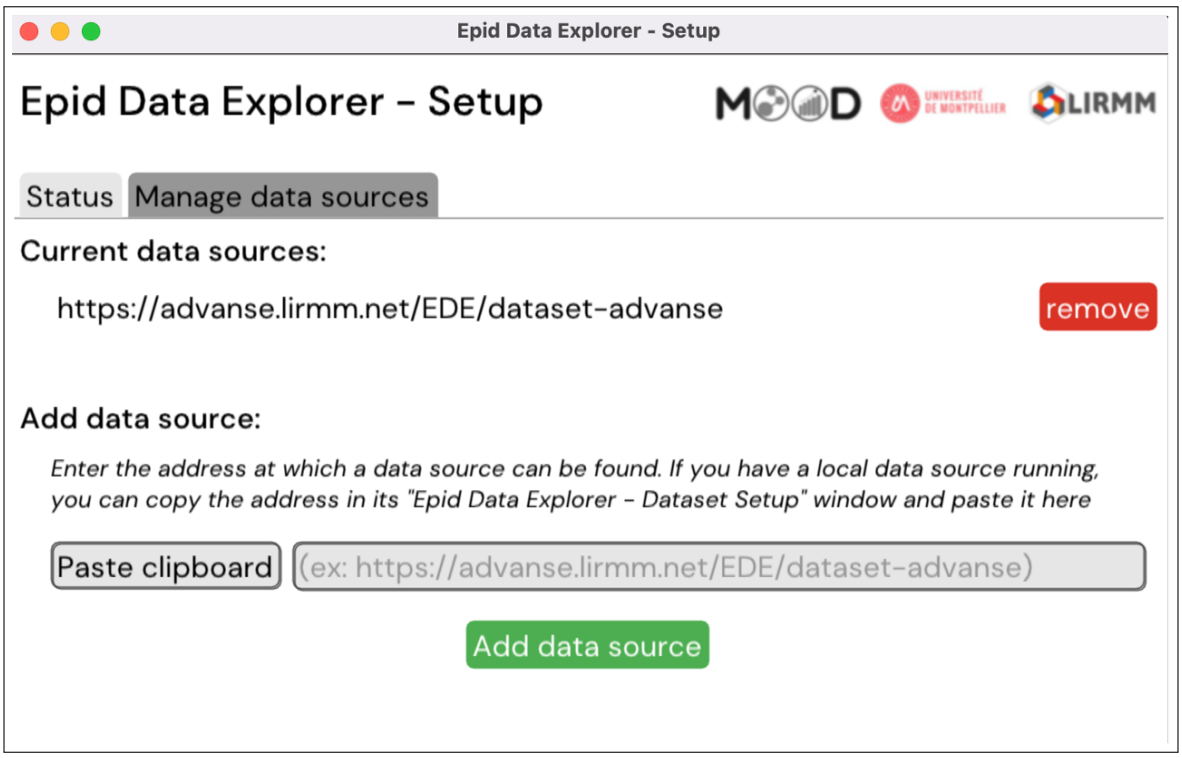



Figure 2.5: Epid Data Explorer - Setup Interface - the Manage Data Sources tab

To start a new instance of EDE's user interface, make the "public" directory located at the root of the project accessible through a web server. To update the list of available data servers, simply edit the "sources.json" file in the "public" directory.

# Chapter 3

# Installation of a Data Store for EDE

In Chapter 2, we learned about the servers that store the data displayed by the Epidemic Data Explorer (EDE) interface. In this section, we will show you how to create a new data store with different sources and add its address to the Epid Data Explorer - Setup Interface, so that new data can be viewed.

The data store package can be downloaded at the following URL: `https://gite.lirmm.fr/advanse/EDE/dataset`. It contains the sources as well as executable codes (for Linux, Windows, and MacOS) to facilitate the installation and management of the server.

## 3.1 Download and Install the Data Store

To download the data store, visit the URL `https://gite.lirmm.fr/advanse/EDE/dataset`. You can download the folder either by clicking the download option or cloning the git repository. If you choose to download a compressed file, extract it in the MyEDE directory (as outlined in Section 2.1). If you prefer to use git, type the following command in the MyEDE directory:

`git clone https://gite.lirmm.fr/advanse/EDE/dataset.git`.

In either case, a directory named "dataset" will be created in the MyEDE directory.

## 3.2 Start the Data Store

To make the data accessible remotely, it is necessary to start a server. To make this process easier for the user, a graphical setup tool interface, called **Epid Data Explorer - Dataset Setup**, is provided for Linux, Windows, and MacOS. This interface not only allows you to start the server, but also to manage the data sources (add, modify the display order, and remove).

For Linux, simply run the executable file "EDE Dataset Setup." For Windows, run "EDE Dataset Setup.exe."

**MacOS specificity:** On MacOS, there are some challenges in installing the interface. To launch the interface, the user must open a terminal, navigate to the directory "MyEDE/epid-data-explorer", and run the command: `python3 gui/main.py`.

**Important note: It's important to note that the Epid Data Explorer - Dataset Setup window must remain open as closing it will result in the shutdown of the EDE user interface instance. The main reason is that the interface launches the server which allows access to the data. However, for experienced users, there is an option to launch the server without the interface (C.f. Section 3.2.2). In this case, however, the interface is still necessary to add data to the server.**

### 3.2.1 Using The Epid Data Explorer - Dataset Setup Interface

Figure 3.1 illustrates the main components of the Epid Data Explorer - Dataset Setup interface. It has four tabs that are described below.

#### The "Status" tab

Figure 3.2 shows an example of the "Status" tab, when the user selects port 8002 and clicks the "Start EDE server" button (see to Figure 3.1). The window displays that the server has been launched at the address

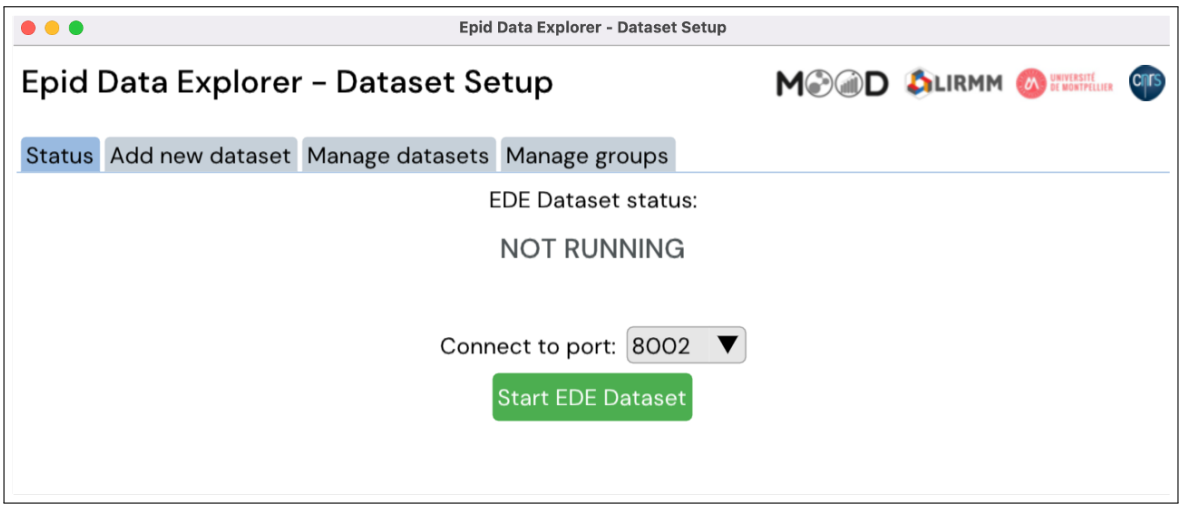



Figure 3.1: Epid Data Explorer - Dataset Setup interface

`https://localhost:8002` (as seen in Figure 3.2-ⓐ). To make the new data server accessible to EDE, the user can copy the address and add it to the "Epid Data Explorer - Setup" interface (see to Figure 2.5). Finally, the user can shut down the server (refer to Figure 3.2-ⓒ), but it is important to note that doing so will make the data inaccessible. Therefore, the user should not close the window as it will automatically stop the server.

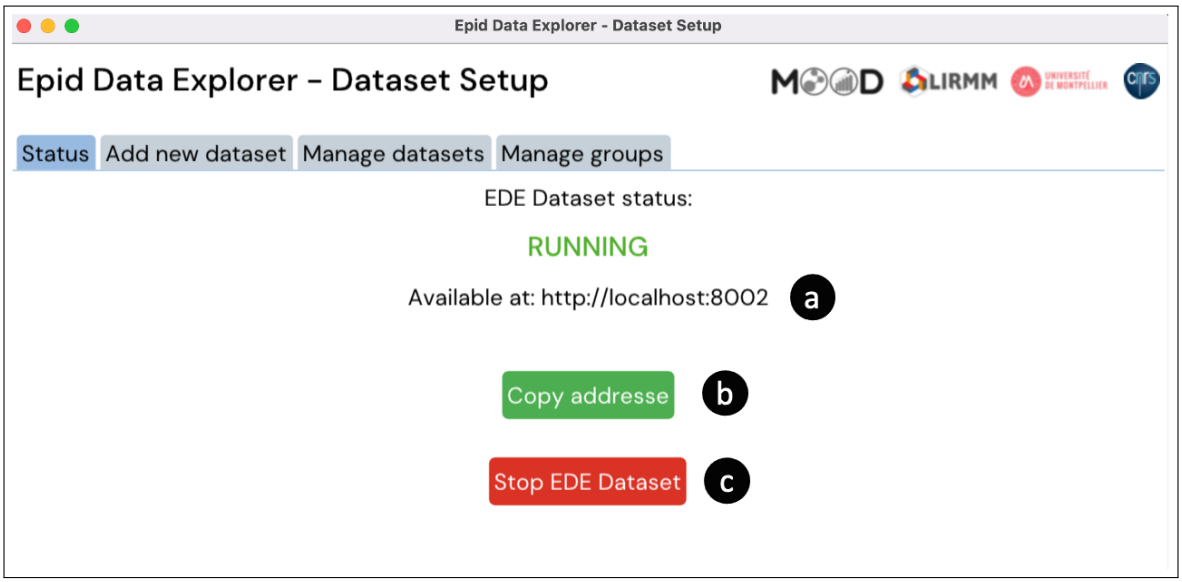



Figure 3.2: Epid Data Explorer - Dataset Setup - the Status tab

### The "Add New Dataset" tab

The "Add New Dataset" tab allows the user to import data into the server. The data must be in csv format and various data formats are accepted, including:

- **ECDC**: the format must follow the names and order of the following columns: "HealthTopic", "Population", "Indicator", "Unit", "Time", "RegionCode", "RegionName", "NumValue", "TxtValue". For EDE, only "Indicator", "Unit", "Time", and "RegionCode" are required. It is essential that the data in "RegionCode" is in NUTS format.
- **Our World in Data**: The format must follow the names and order of the following columns: "iso_code", "continent", "location", "date", the name of the first indicator, the name of the second indicator, and so on. For EDE, the essential information is "iso_code", "date", and the columns associated with all indicators. The iso_code must be in ISO format.
- **User-defined (variables by row)**: this format is inspired by ECDC's but allows for more freedom: not all columns are required, and other geocodes can be used. The user must provide a location, a time, all indicators in one column, and their corresponding value in another column. The geocode can be either

NUTS or ISO. When adding the dataset in the interface, the user must specify the contents of each column.

- **User-defined (variables by column)**: this format is inspired by Our World in Data's but also allows more freedom: not all the columns are required, other geocodes can be used. The user must provide a location, a time, and one column per indicator. The geocode can be either NUTS or ISO. Again, when adding the dataset, the user must specify the contents of each column.

Recommendations: To make the integration of data in EDE easier, it is highly recommended to follow these guidelines:

- **separator**: use a semicolon.
- **date**: the format must be "YYYY-MM-DD", "YYYY-MM", or "YYYY".
- **missing data**: avoid keeping missing data in the dataset.
- geocode: the current version of EDE uses NUTS version 2021 for NUTS, and the ISO 3166-1 Alpha2 code for country level regions and the official ISO 3166-2 code for regions and subregions. However, a list of exceptions can be found at `https://observablehq.com/@vraveneau/covid-mobility-shapes`.The 2020 revision of ISO 3166 codes is used.

When adding a new dataset to the system, the first three points are evaluated and corrected. However, an error is detected when it comes to the geocode part and the file cannot be added.

To illustrate the process, we will use data examples from the "dataset_examples" directory in the git repository. The user starts by selecting their file (refer to Figure 3.3-ⓐ). In our example, we have selected the file "ECDCFormat_YEAR_NUTS3_example.csv" which contains ECDC format files with a temporal granularity of year and a NUTS3 geocode. The "Specify the name of the dataset" field is mandatory and determines how the dataset will be named in the EDE platform (refer to Figure 3.3-ⓑ). In our example, we have labeled it "An example of a file in ECDC format (NUTS3)". The "Specify the URL of the data source" and "Add a description of the dataset" fields are optional (refer to Figure 3.3-ⓑ). If they are filled in, they will be displayed in EDE when the user clicks on the information icon.

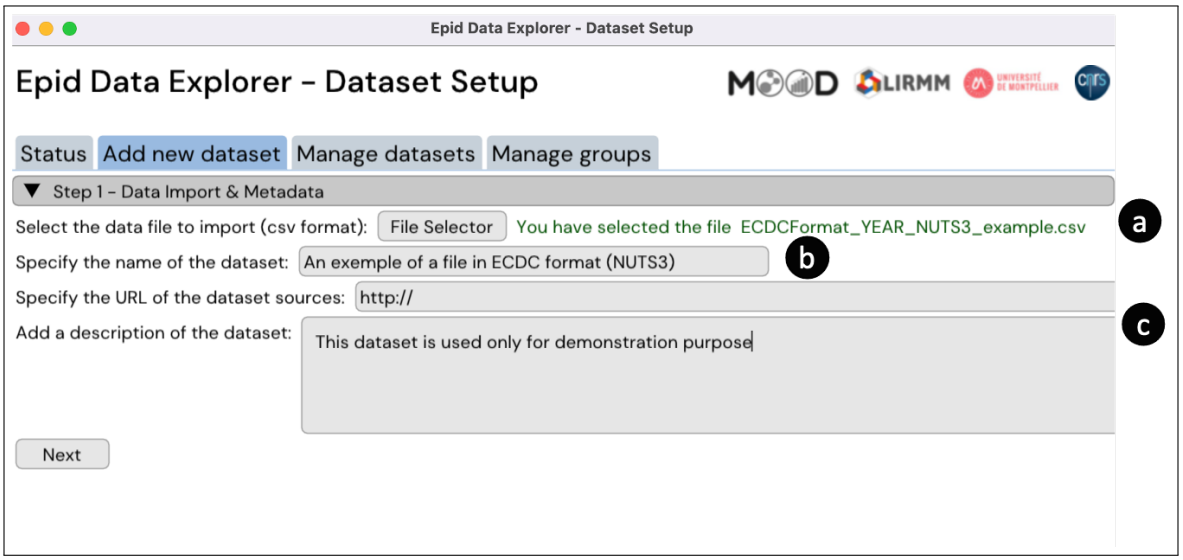



Figure 3.3: The file selector

By clicking the "Next" button, the user will see a new area to specify the file format, as shown in Figure 3.4. By clicking the "Next" button again, the user can select the geocode used. For example, if the ISO format is selected (refer to Figure 3.5), the user can specify if the data is at the country, region, or sub-region level. The "Mix" option indicates that the data is at different geographical granularities. The same applies to NUTS, where the interface offers different levels of NUTS3 to NUTS0 (refer to Figure 3.6).

In our example, as the sample file contains only NUTS3 granularity data, we select NUTS3. By clicking the "Next" button again, the user can select the indicators from the dataset that will be displayed in EDE. If the data has a lower geographical granularity, the user can aggregate the data to higher levels by clicking the "Enable Geographic Aggregation" area (refer to Figure 3.7-ⓐ). If the data is already at the highest level, this

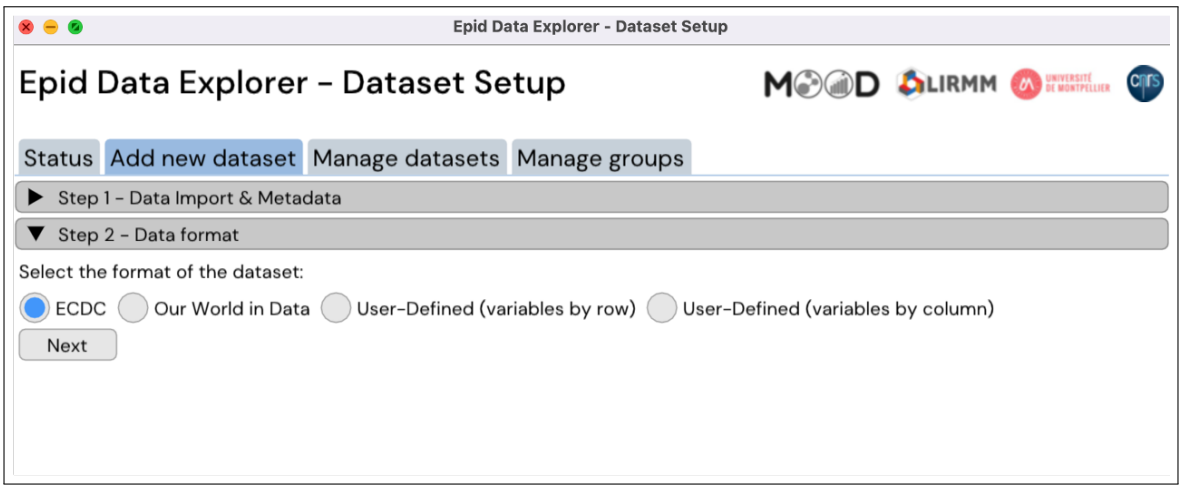



Figure 3.4: The format file selector

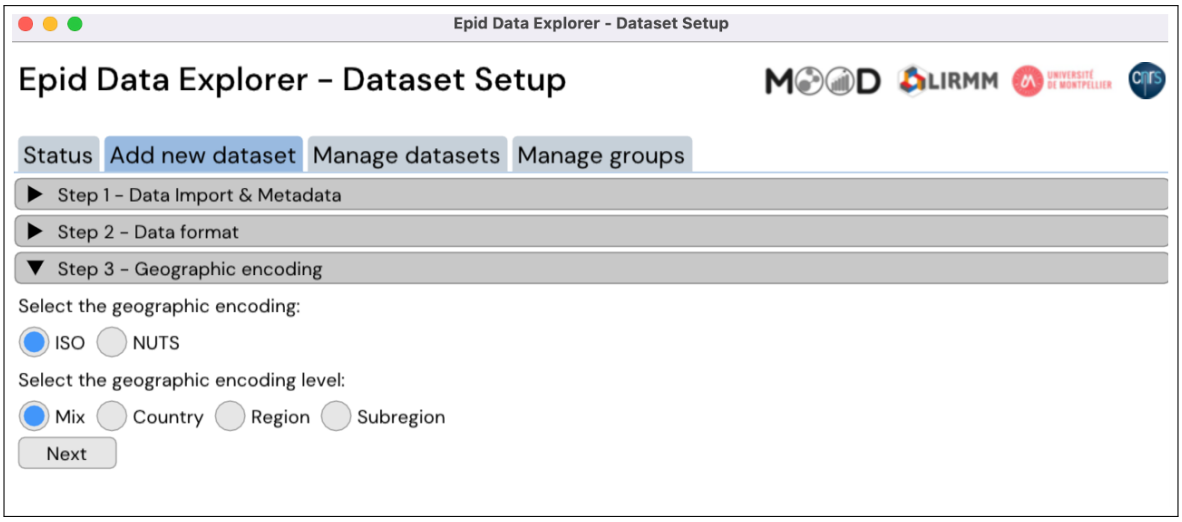



Figure 3.5: The geocode selector for ISO

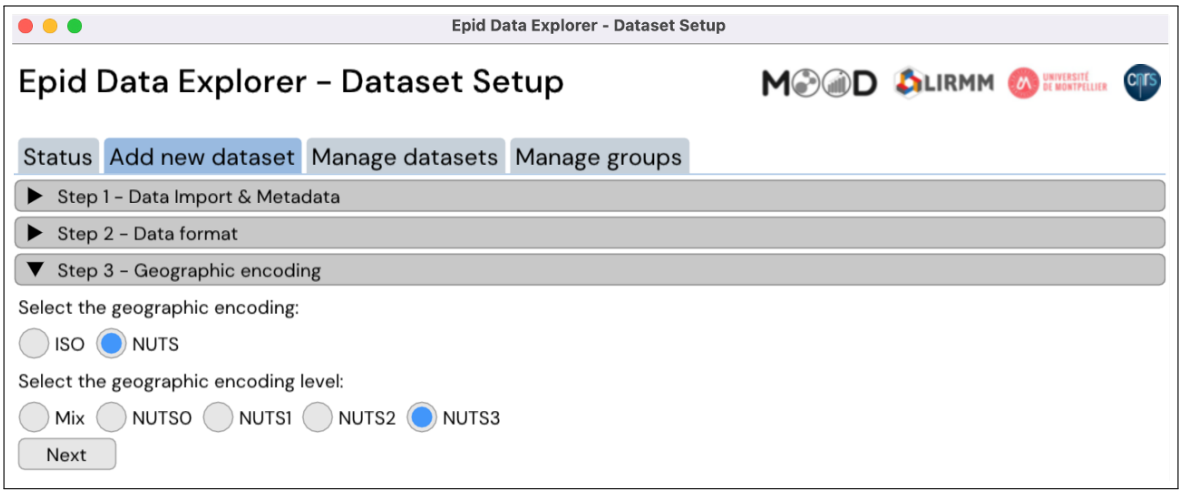



Figure 3.6: The geocode selector for NUTS

operation will have no effect. EDE supports different types of indicators, and the legend under the maps adapts accordingly (refer to Figure 1.8).

As previously discussed (refer to Figure 1.8), EDE supports various types of indicators and the legend on the maps adjusts accordingly. The "Select the type of variable" section (Figure 3.7-ⓑ) enables the user to select the type of indicators.

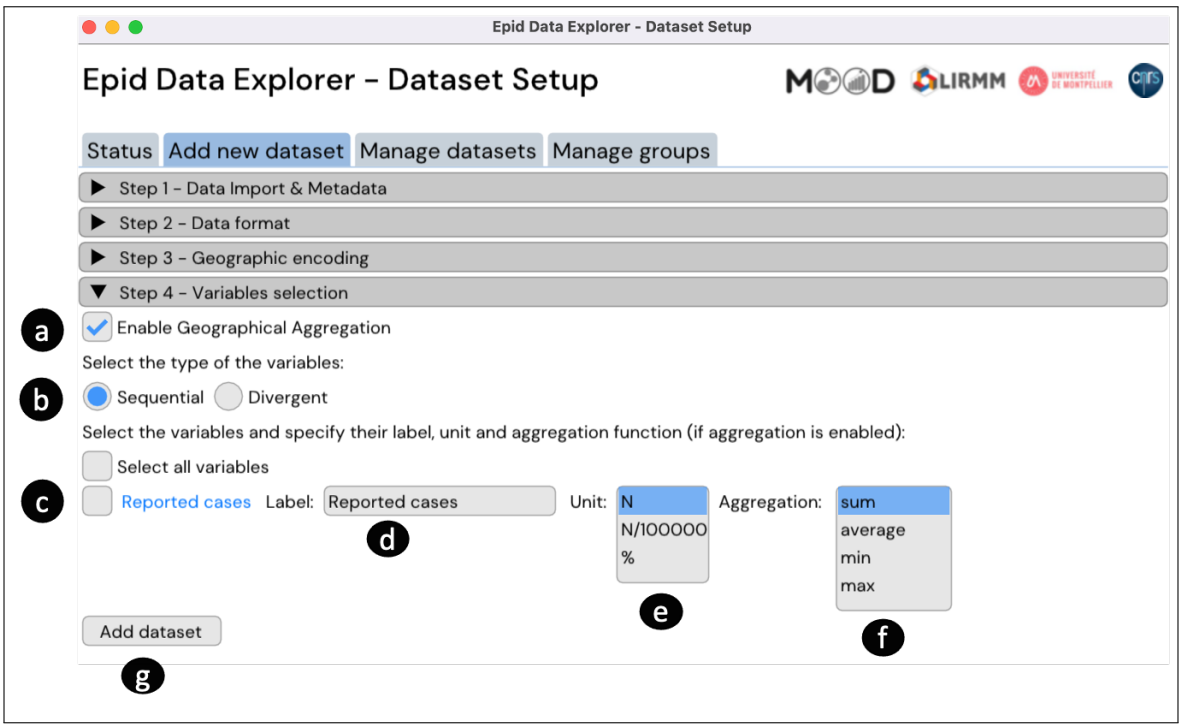



Figure 3.7: The variable selector

The list of all the indicators in the dataset is displayed and the user can choose which ones to include by clicking next to the indicator name. The user can also modify the label name (Figure 3.7-ⓓ). The default value is the one from the dataset, and the label name value will be used in EDE. Finally, the user can specify the unit of the indicator (Figure 3.7-ⓔ). As previously noted, data can be aggregated at different geographical levels, and the user can select the aggregation function to calculate the values of the indicator at a higher level (Figure 3.7-ⓕ).

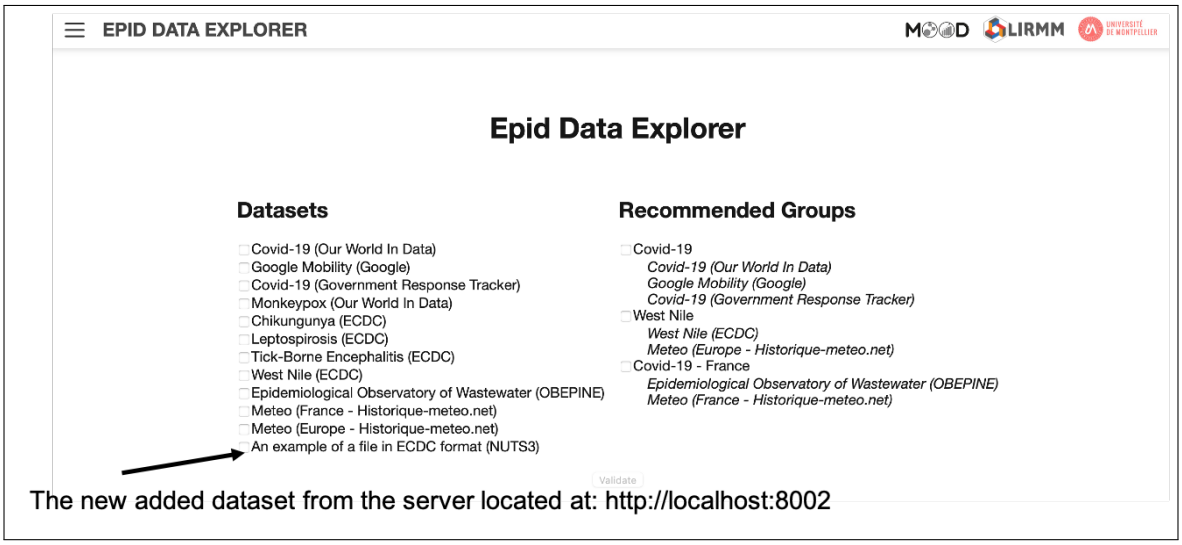



Figure 3.8: The new added dataset in EDE

Then by clicking on "Add dataset" as illustrated in Figure 3.7-ⓖ, the dataset is automatically added to the Data Store. Figure 3.8 shows that the new dataset is thus now available for EDE.

As the data is at a NUTS3 level of granularity, the user wished to aggregate the data to higher levels of granularity and selected, for example, the sum of indicator values (see Figure 3.7 -ⓐ, ⓐ. The interface generates all levels and Figure 3.9 illustrates the result obtained.

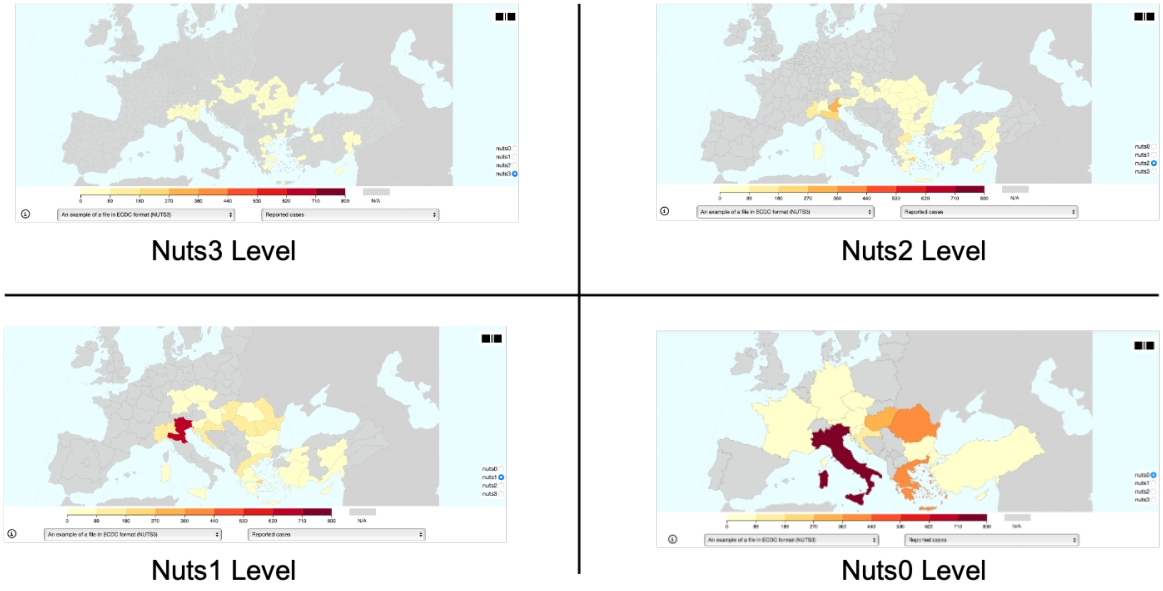



Figure 3.9: The aggregation performed at all the geographical level

**The "Manage Datasets" tab**

The "Manage Datasets" tab (see Figure 3.10) allows the user to delete datasets or reorder them to change their appearance in EDE.

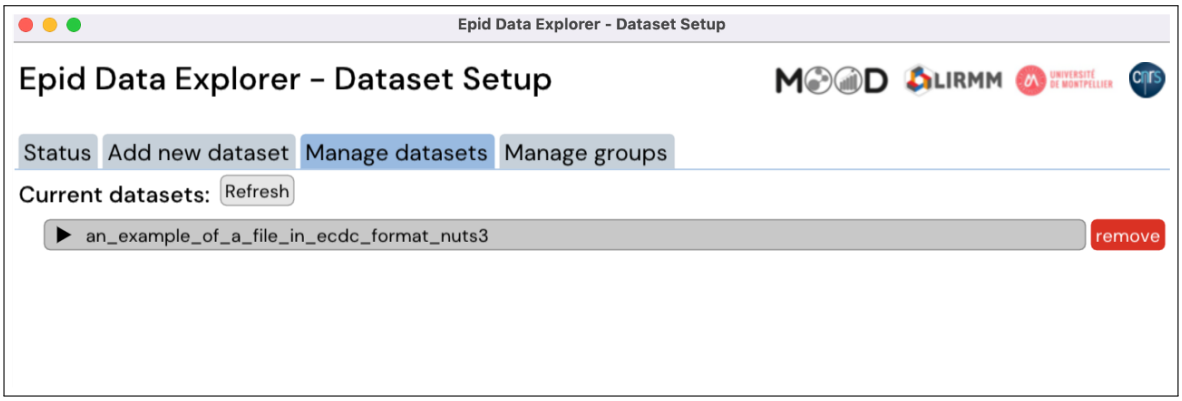



Figure 3.10: The Manage datasets tab

**The "Manage Groups" tab**

As we have seen, when launching the EDE platform, on the right-hand side, "Recommended groups" are offered. The purpose of this tab is precisely to be able to group data according to the user's topic of interest (see Figure 3.11).

### 3.2.2 Installing a Data Store Without the Setup Interface

In addition to using the setup interface, it is also possible to install the Data Store without it. This requires launching the server, similar to the process described in Section 2.2.2. The main difference is that the user must go to the "MyEDE/dataset/" directory. However, the interface provides many data verification and transformation operations, as well as converting the cleaned data into an internal format stored in the "data" directory with files that are used by EDE. The "datasets.json" file is also created to store the necessary information for EDE. For more information, refer to the README in the git repository. **We thus highly recommend using the interface to add or manage data.**

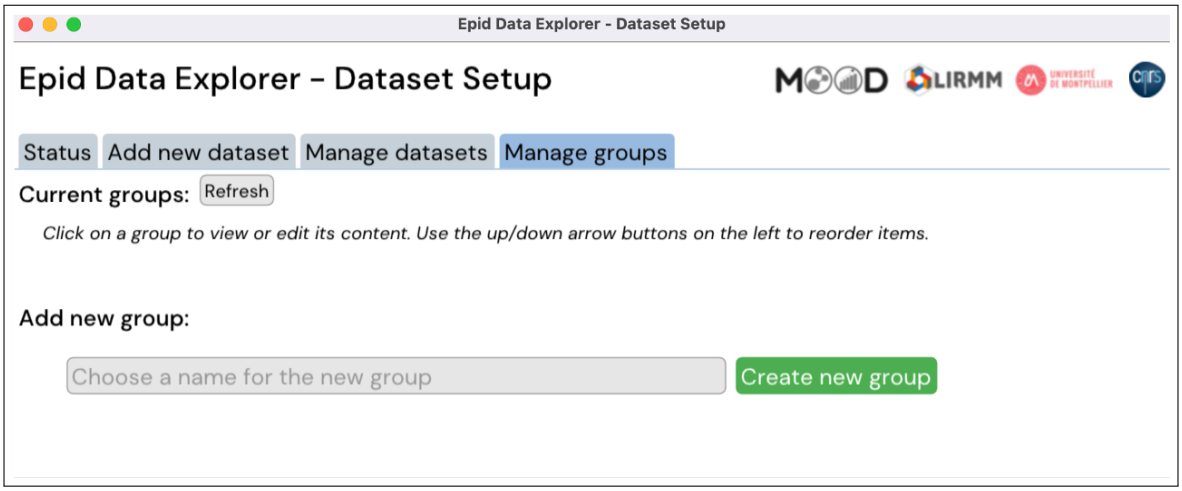



Figure 3.11: The Manage groups tab