



HAL
open science

Epid Data Explorer: A Visualization Tool for Exploring and Comparing Spatio-Temporal Epidemiological Data

Laëtitia Viau, Jérôme Azé, Fati Chen, Pierre Pompidor, Pascal Poncelet,
Vincent Raveneau, Nancy Rodriguez, Arnaud Sallaberry

► To cite this version:

Laëtitia Viau, Jérôme Azé, Fati Chen, Pierre Pompidor, Pascal Poncelet, et al.. Epid Data Explorer: A Visualization Tool for Exploring and Comparing Spatio-Temporal Epidemiological Data. Health Informatics Journal, 2024, 30 (3), 10.1177/14604582241279720 . lirmm-04658031

HAL Id: lirmm-04658031

<https://hal-lirmm.ccsd.cnrs.fr/lirmm-04658031v1>

Submitted on 17 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License



Epid data explorer: A visualization tool for exploring and comparing spatio-temporal epidemiological data

Health Informatics Journal
1–18

© The Author(s) 2024

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/14604582241279720

journals.sagepub.com/home/jhi



Laetitia Viau , Jérôme Azé, Fati Chen, Pierre Pompidor,
Pascal Poncelet, Vincent Raveneau  and Nancy Rodriguez 

LIRMM, Université de Montpellier, CNRS, France

Arnaud Sallaberry 

LIRMM, Université de Montpellier, CNRS, France

AMIS, Université Paul-Valéry Montpellier 3, France

Abstract

The analysis of large sets of spatio-temporal data is a fundamental challenge in epidemiological research. As the quantity and the complexity of such kind of data increases, automatic analysis approaches, such as statistics, data mining, machine learning, etc., can be used to extract useful information. While these approaches have proven effective, they require a priori knowledge of the information being sought, and some interesting insights into the data may be missed. To bridge this gap, information visualization offers a set of techniques for not only presenting known information, but also exploring data without having a hypothesis formulated beforehand. In this paper, we introduce Epid Data Explorer (EDE), a visualization tool that enables exploration of spatio-temporal epidemiological data. EDE allows easy comparisons of indicators and trends across different geographical areas and times. It facilitates this exploration through ready-to-use pre-loaded datasets as well as user-chosen datasets. The tool also provides a secure architecture for easily importing new datasets while ensuring confidentiality. In two use cases using data associated with the COVID-19 epidemic, we demonstrate the substantial impact of implemented lockdown measures on mobility and how EDE allows assessing correlations between the spread of COVID-19 and weather conditions.

Corresponding author:

Laetitia Viau, LIRMM, Université de Montpellier, CNRS, 161 rue ada, Montpellier 34000, France.

Email: laetitia.viau@hotmail.fr



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further

permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Keywords

spatio-temporal epidemiological data, data visualization, data exploration, data comparison

Introduction

In a context of increasing volumes of spatio-temporal data, data analysis is a powerful tool for supporting epidemiological research. There are various approaches to processing this type of data, such as statistics, data mining, machine learning, etc. However, a solid understanding of the data is necessary to choose the most appropriate approach. Visualization satisfies these needs through two approaches: data presentation and in-depth exploration. As data becomes increasingly voluminous, it is becoming crucial to develop suitable, high-quality, interactive tools.

We propose Epid Data Explorer (EDE), a visualization tool enabling fast, simplified exploration of spatio-temporal data. A particular feature of this tool is its ability to facilitate comparisons between indicators for different geographical areas, dates and/or periods. EDE can also be used to visualize and compare trends. Another notable advantage of EDE is the ability to import new datasets securely, which guarantees data confidentiality when required. In short, Epid Data Explorer (EDE) meets an essential need in the field of epidemiology by providing a platform for the visualization and comparison of spatio-temporal epidemiological data. With EDE, the exploration and comparison of epidemic data becomes more accessible and more efficient.

Requirements

This work is conducted within the framework of the European project MOOD (MOnitoring Outbreak events for Disease surveillance in a data science context), engaging 25 partners from 12 countries. The project members include public health agencies, veterinary health agencies, and surveillance practitioners. It is dedicated to developing innovative tools and services through close collaboration with professionals involved in the early detection, assessment, and monitoring of current and potential infectious disease threats in Europe and beyond.

In this context, we defined, in collaboration with a diverse group of users, including public health officials who monitor disease outbreaks and coordinate responses, data analysts who process large datasets and identify trends, and individuals responsible for health surveillance to prevent crises, a set of requirements that have guided the design of our platform. This process was carried out iteratively: users defined a set of requirements, we then proposed a prototype, which led users to specify or redefine new requirements, prompting us to make a new proposal, and so on. This approach was strongly inspired by the recommendations of Munzner¹ and the design process proposed by Sedlmair et al.²

These requirements cover both the functionality of the platform and the system architecture itself. End-users have expressed the desire to visualize multiple datasets in a unified platform so as to identify potential correlations between indicators through comparative analysis. They also want the ability to track changes in indicator values over time. These tasks can be facilitated by supporting simultaneous comparisons within a single view. Additionally, as certain data may be sensitive and require confidential storage, the tool must implement robust data security measures. This includes the ability for users to import their own datasets while maintaining strict data privacy. Based on these needs, the following requirements have been identified for the Epid Data Explorer tool:

R1. Users must have the ability to make various comparisons within a single view,³ encompassing all possible combinations of three dimensions: (1) indicators (the value of the indicator visualized), (2) time (the corresponding date), and (3) space (the geographical area). For instance, let us consider the indicators "Total confirmed cases" and "Total confirmed cases per 1,000,000 people". It should be possible to compare the values of these two indicators for the same date and the same geographical area. It should also be possible to compare the values of the same indicator in the same geographical area for two different dates, or to compare the same indicator in two different geographical areas for the same date.

R2. A user may need to access different spatial and temporal granularities. Some datasets may have different geographic scales, such as country, region or sub-region. If the data is not at the highest level of the scale, it must be possible to perform aggregation. For example, if a dataset contains values by region, the values can be aggregated for each country so that the user can view the data by region or by country if desired. The same is valid for temporal granularity, with levels such as day, week and month.

R3. It must be possible to visualize and compare the evolution of an indicator for multiple geographical entities. For instance, let us consider the indicator "Max temperature" and two regions: Andalusia in Spain and Nord-Norge in Norway. Users should have the ability to view the evolution of the maximum temperature value in each of these regions and make comparisons. They should be able to quickly identify significant differences between regions.

R4. Users must be able to compare their own datasets with each other and with public datasets if needed. Users often have their own datasets, and they would like the capability to effortlessly integrate them into a user-friendly system, exploring and leveraging them in the same manner as publicly available data.

R5. Private data must be managed securely. The users emphasized the need to protect their own datasets to ensure that they cannot be accessed by third parties.

Related work

There are many web tools available for epidemiological data surveillance visualization. In this work, we only focus on those that enable tracking epidemics in both geographical and temporal dimensions.

We identified two types of such tools: event-based surveillance and indicator-based surveillance tools. The former processes unstructured event reports, while the latter directly visualizes structured data.

We don't include simulation tools such as GLEaMviz.⁴ Even if they provide numerous functionalities for visualizing epidemiological data (GLEaMviz contains dynamic maps and charts describing the geo-temporal evolution of diseases), they require an epidemic model and a simulation scenario to simulate the spread of infectious diseases, and do not focus on the visualization of actual epidemiological data per se.

Event-based surveillance

For event-based platforms, there are specific tools dedicated to monitoring particular diseases. For instance, Monitoring Rabies in Media⁵ is an alert system designed to monitor the daily circulation of press articles related to rabies. The EpidNews analytical visualization tool⁶ tracks source data related to animal epidemiology to observe the spread of epidemics. It extracts information on locations, dates, and symptoms. The EpidVis tool⁷ simplifies web searches for animal-disease

detection and monitoring. It is a visual query tool specifically designed for animal health experts. A visual analytics interface featuring coordinated views has recently been developed by Kuo et al.⁸ This interface also focuses on animal epidemics and allows the investigation of epidemics by identifying the relationships between livestock farms. The consequences of a disease outbreak, its severity, and its scale are estimated using unsupervised machine learning methods. Another automated system, HealthMap,⁹ queries, filters, integrates, and visualizes unstructured reports on disease outbreaks using text processing algorithms. Users can also propose alerts. Similarly, BioCaster¹⁰ detects and tracks the distribution of infectious diseases by continuously analyzing RSS feeds.

Indicator-based surveillance

One of the platforms used for indicator-based epidemiological data visualization is Epi Visualization.¹¹ This tool allows for the exploration of health data and of estimates related to epidemics. The available data is quite extensive, covering 369 diseases across 204 countries and territories from 1990 to 2019. With Gapminder,¹² users can visualize data spanning various subjects, not only in the field of health but also in economics, the environment, demography, and many others. The platform Empres-i¹³ visualizes data sourced from various international animal health authorities including ministries of agriculture and health and the Food and Agriculture Organization of the United Nations. The European Centre for Disease Prevention and Control (ECDC) provides interactive databases that offer tabular and map views of ECDC data. One of the tools provided is the Surveillance Atlas of Infectious Diseases, which allows the manipulation of data collected via the European surveillance system TESSy. The ECDC also provides Dashboards. One such dashboard is the Polio dashboard, which provides an overview of the global poliovirus situation.

The global COVID-19 pandemic that started in 2020 led to the development of many visualization tools. One of the best known is the COVID-19 Dashboard,¹⁴ developed by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. The visualizations proposed by the COVID-19 Data Explorer tool from Our World in Data,¹⁵ in the form of diagrams, maps, and tables, are also widely recognized. Additionally, other applications such as COVID-19 CG¹⁶ and CoV-Spectrum¹⁷ focus specifically on mutations of the SARS-CoV-2 virus. These applications offer functionalities to explore and analyze data on virus sequences through phylogenetic trees, as well as mutation and lineage tracking by locations and dates of interest. Moreover, there are other applications with similar functionalities such as Nextstrain¹⁸ and VirusViz,¹⁹ which can address different types of viruses. There is also a dedicated dashboard for monitoring the status of COVID-19 vaccines called the COVID-19 Vaccine Tracker. For further exploration, interested readers can refer to,²⁰ which provides a survey of COVID-19-related visualization dashboards.

The performance of indicator-based platforms with regard to the requirements is summarized in Table 1. We consider R1 to be fulfilled if the approach uses one of the strategies described by Gleicher et al.³ for each of the configurations described in the requirement, ie. if the comparison can be made within a single visualization showing both objects to be compared. Consequently, visualizations that rely on animation, or any user action, to switch from one object to another, without displaying them simultaneously, are not considered satisfactory. This is because they require significant cognitive load, which can hinder accurate comparison.²¹ Using one of the strategies described by Gleicher et al. for some, but not all, configurations is considered only a partial fulfillment of the requirement. R2 is fulfilled if different granularities are available for both space and time. Having only one of these options is considered a partial fulfillment of the requirement. R3 is fulfilled if the evolution of several indicators can be displayed at the same time. If the tool only

Table 1. Fulfillment of requirements by indicator-based platforms. ✓ indicates that the platform fully meets the requirement, ○ indicates partial fulfillment, and X indicates that the approach does not address the requirement at all.

	R1	R2	R3	R4	R5
Epi visualization	○	○	✓	X	X
Gapminder	✓	○	✓	✓	✓
Empres-i	○	✓	✓	✓	✓
Surveillance atlas of infectious diseases (ECDC)	○	✓	✓	X	X
COVID-19 vaccine tracker (ECDC)	○	X	○	X	X
COVID-19 dashboard	○	✓	○	X	X
COVID-19 data explorer (our world in data)	○	○	✓	X	X
COVID-19 CG	○	○	○	✓	✓
CoV-spectrum	○	○	✓	✓	✓
Nextstrain	○	✓	○	✓	✓
VirusViz	○	○	○	✓	X

displays the evolution of a single indicator at a time, the requirement is considered partially fulfilled: the indicators are indeed displayed, but their comparison is tedious for the reasons explained in R1. R4 is fulfilled if users are able to use the tool with their own data, in addition to any data that could be provided by the tool. R5 is fulfilled if the users' data can be loaded securely, either because they are not stored by the tool or because any data storage elements can be deployed locally.

This analysis shows that none of the indicator-based platforms presented fully meet all the requirements. We are therefore proposing an optimized solution that addresses all user needs.

The EDE platform

The Epid Data Explorer (EDE) web platform offers two ways to visualize data on maps. The first, which is displayed by default, consists of the juxtaposition of two maps. The second is a view with a single map. The data used to demonstrate the functionality of EDE comes from the ECDC,²² Our World in Data,²³ Google,²⁴ Government Response Tracker,²⁵ Obépine²⁶ and Historique météo.²⁷

Homepage

Before accessing the visualizations, the user must select the datasets of interest using the list of datasets on the left (Figure 1(a)) and/or the list of dataset groups on the right (Figure 1(b)). Groups allow the datasets to be assembled by theme. In Figure 1, the COVID-19 group is selected, which means that by clicking on “Validate”, the user will be able to visualize all the datasets available in EDE related to the COVID-19 pandemic. The “Meteo (Europe - historique-meteo.net)” dataset is also selected, so it will also be displayed. The homepage also provides an overview of the purpose and principle of EDE (Figure 1(c)). There is a menu (Figure 1(d)) in the top-right corner of the platform.

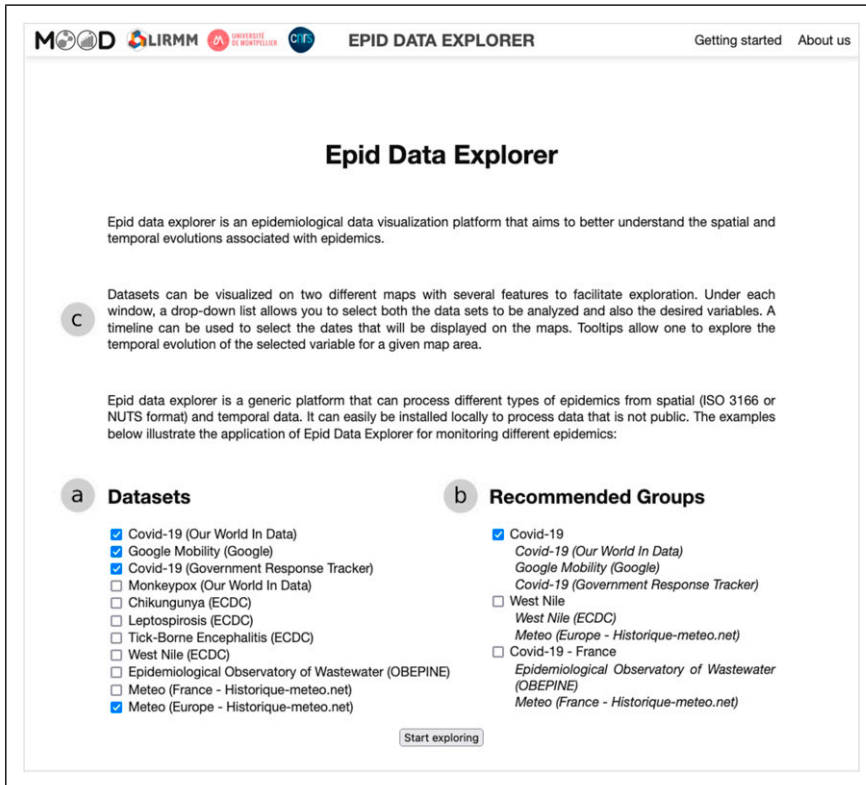


Figure 1. Epid Data Explorer homepage.

Proposed views

The 1-map view, as shown in [Figure 2\(a\)](#), is displayed by default. In this view, the user is able to focus on one element, being provided with an overview of an indicator that can be explored from a temporal and geographical perspective without the clutter of a second map. If the user needs to compare two maps, the 2-map view ([Figure 2\(b\)](#)) is more suitable. It provides two juxtaposed maps, allowing easy comparisons. Switching between the two views is very simple. To go from a 2-map view to a 1-map view, the user simply clicks on the button located in the upper-left corner of the desired map ([Figure 2\(b\)-Ⓐ](#) or [Figure 2\(b\)-Ⓑ](#)). To add a second map, the user simply clicks on the button ([Figure 2\(a\)](#)) located in the upper-left corner of the map.

Main components

Map components. The map ([Figure 3\(a\)](#)) facilitates the exploration of various geographical features. In [Figure 3](#), the country granularity is selected. It is possible to modify the level of detail via the input field composed of radio buttons ([Figure 3\(b\)](#)). In this example, the possible granularities are country, region and subregion and correspond to the ISO 3166 nomenclature.²⁸ Depending on the dataset, the nomenclature is either ISO or NUTS.²⁹ In the latter case, the granularity has 4 levels: NUTS0,

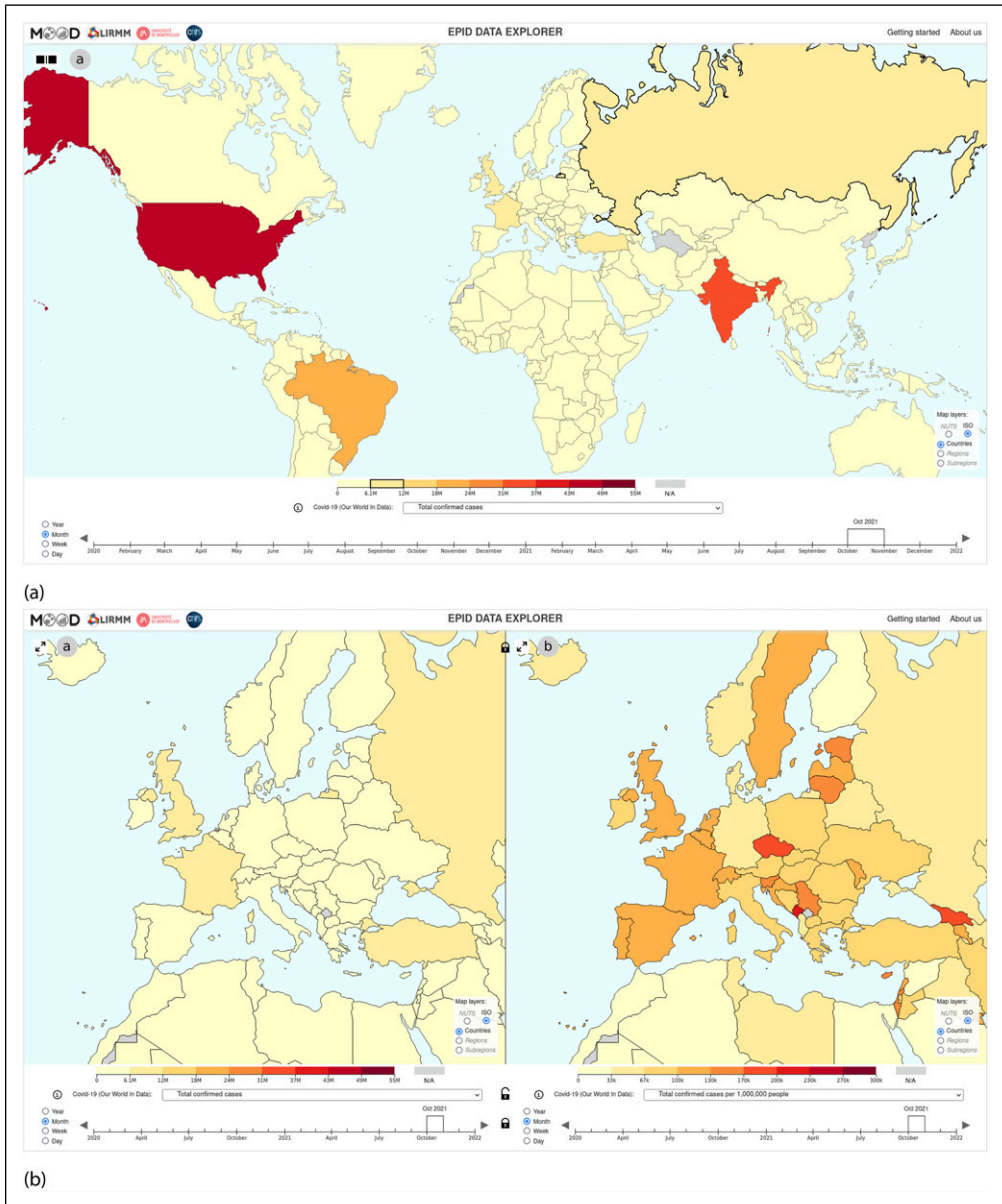


Figure 2. Epid Data Explorer views. (a) 1-map view. (b) 2-map view.

NUTS1, NUTS2 and NUTS3. The map can display values at the level provided in the dataset and at higher levels if aggregation was carried out during the importation of the data. For example, if the data uses ISO 3166 geocoding and the granularity is set to region, aggregation can be carried out to view the data on the country level. This satisfies requirement R2, i.e. providing access to different spatial granularities and aggregating data.

The map shows the value of the indicator for each geographical entity at a certain date or period. The color of each geographical entity is defined using a color scale according to the value of the indicator. The legend below the map (Figure 3(c)) clearly displays the indicator values and automatically adjusts to the selected data type. EDE can handle both ordinal and quantitative data types, as shown by Figure 4. Two examples of legends for quantitative data are provided: Figure 4(a) for a continuous scale and Figure 4(b) for a divergent scale. A divergent scale is used when data values move in two opposite directions. In the example, the values are positive and negative, and the pivot

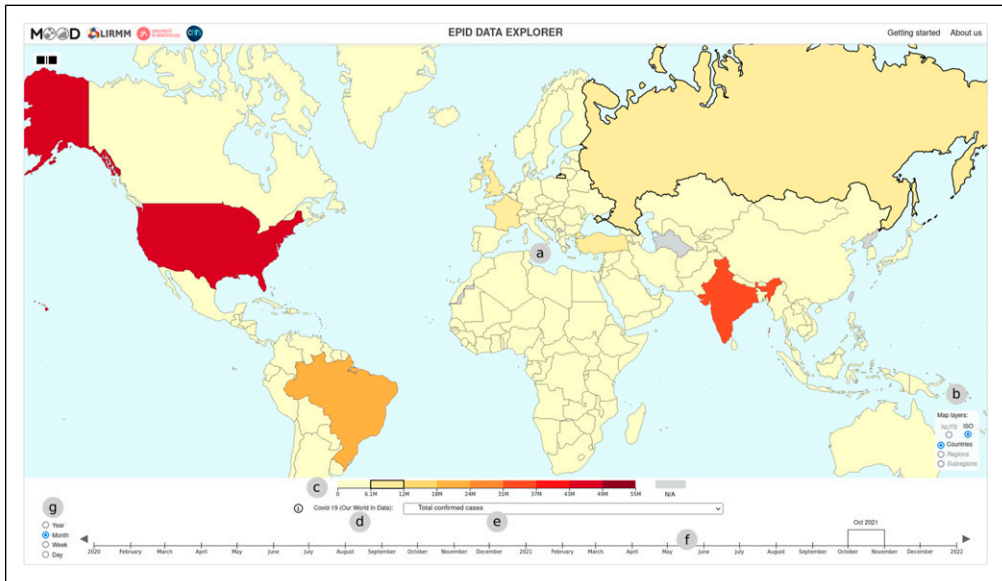


Figure 3. Main components of a map.

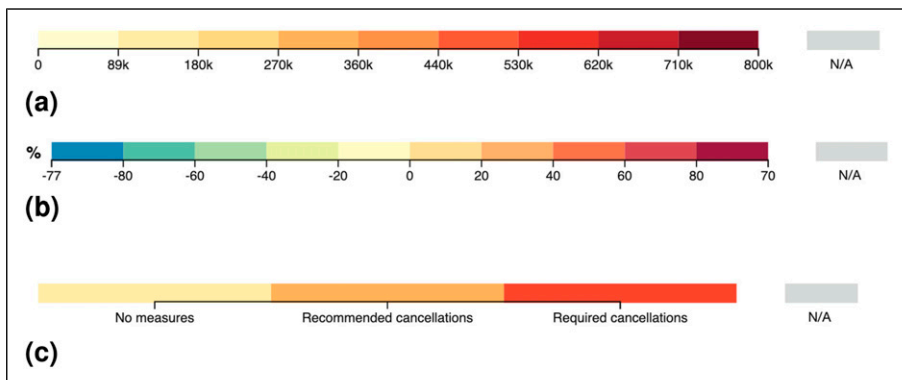


Figure 4. Legends according to the type of data to be visualized. (a) Legend for quantitative data based on a continuous scale. (b) Legend for quantitative data based on a divergent scale, data values go in two opposite directions. (c) Legend for ordinal qualitative data.

value is 0. Another example of a legend, for ordinal qualitative data, is shown in Figure 4(c). EDE generates the legend by analyzing the dataset values for adapting the legend to the selected indicator.

Several datasets are available, and each of them is composed of several indicators. In the example in Figure 3, the dataset “COVID-19 (Our World In Data)” is selected (Figure 3(d)) and the indicator is “Total confirmed cases”. The selection of an indicator (and thus the associated dataset) can be easily changed using the drop-down list (Figure 3(e)).

In addition to selecting an indicator from a dataset, a user can also select a date via the timeline (Figure 3(g)). Like the geographical entities, the time dimension also has several granularities: day, week, month and year. It can be modified by selecting the time level via the radio buttons (Figure 3(g)). This satisfies requirement R2: providing access to different time granularities. The minimum granularity level is determined by the granularity of the dataset. The system also automatically aggregates the data from the lowest level (by day) to the desired higher level and updates the legends accordingly, as shown in Figure 5. This functionality satisfies requirement R2.

Tooltips. While navigating on a map, the user can easily access more detailed information on the selected indicator for a geographical entity. Tooltips are displayed by clicking on a geographical feature. A tooltip shows the values of the current indicator for the selected area for the whole period covered by the dataset. It can be repositioned on the map and resized. Figure 6 shows an example in

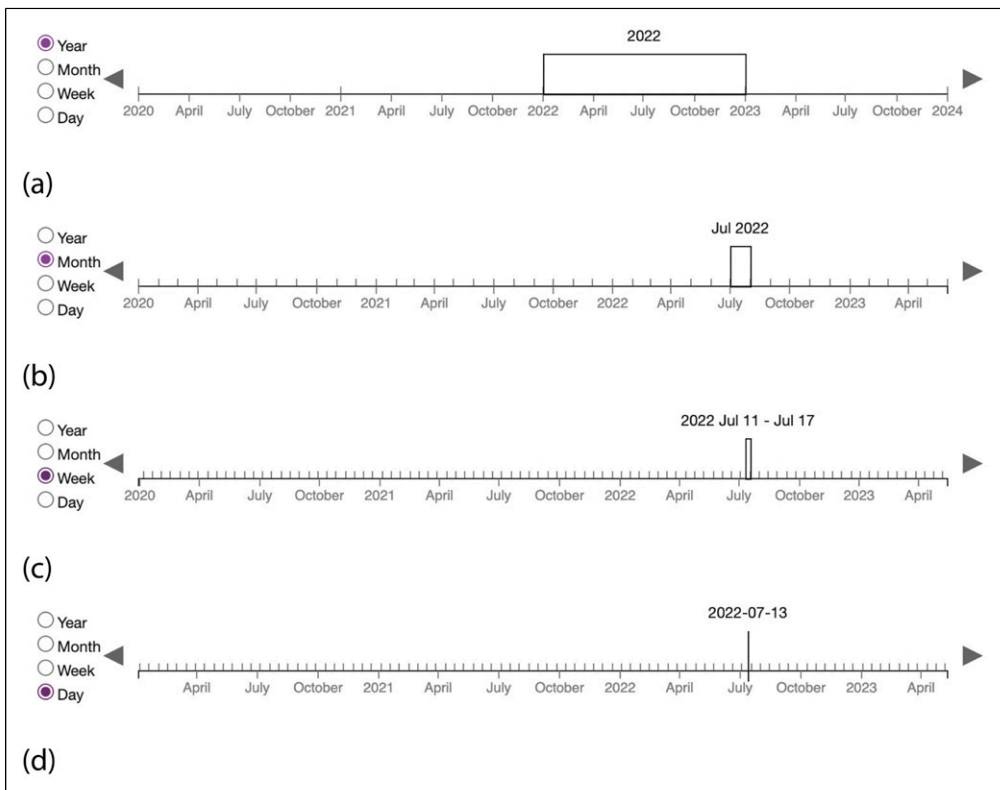


Figure 5. Available timelines according to the temporal granularity selected. (a) Annual timeline. (b) Monthly timeline. (c) Weekly timeline. (d) Daily timeline.

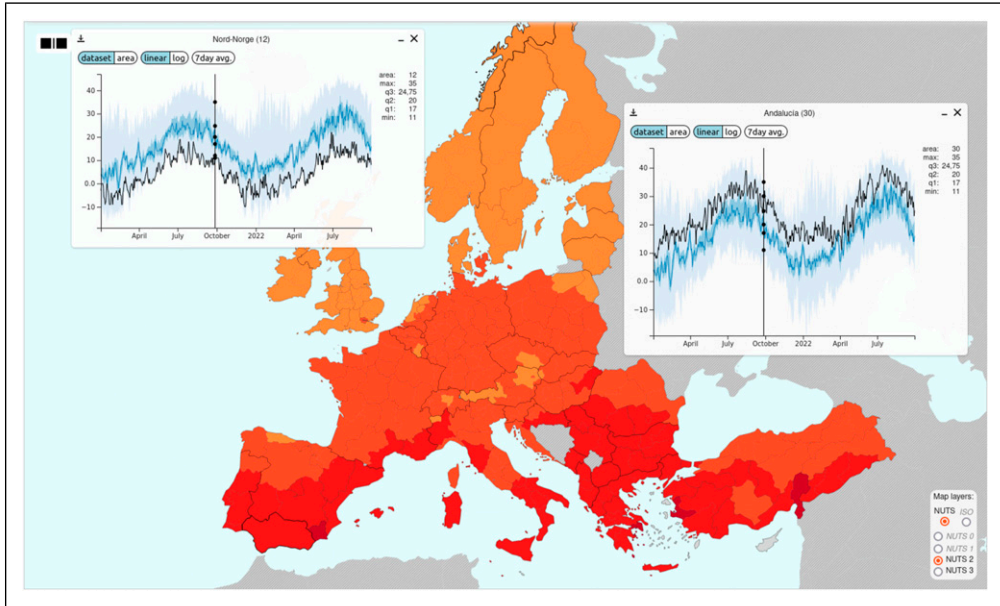


Figure 6. Example of a map with two open tooltips. The indicator is the maximum temperature for the day of September 27, 2021. The spatial granularity here is NUTS2.

which two tooltips are open: one for Andalusia, a region in Spain, and the second for Nord-Norge, a region in Norway. The indicator is the maximum temperature. The simultaneous display of several tooltips corresponds to requirement R3.

Figure 7 shows the key components of a tooltip. The values for the selected geographical entity over time are displayed as a black line (Figure 7(a)). The evolution of the median value, computed from the values for all geographical entities, is represented by a blue line (Figure 7(b)). The lighter blue zone (Figure 7(c)) represents all of the values for all geographical entities, with the delimitations corresponding to the minimum and maximum values. The dark blue zone (Figure 7(d)) around the line of the median value corresponds to the values between the first and third quartile for all geographical entities. All this gives context to show how the entity is situated in relation to others. The user can select different options (Figure 7(e)) to adapt the graph displayed. The first option, *dataset/area*, allows the user to choose the range of data displayed on the *y*-axis, either that of the entity that coincides with the tooltip or that of the entire dataset. With *linear/log*, the user can switch between a linear and a logarithmic scale. The last option, *7-days avg.*, smooths the curve by calculating the average value over a sliding window of 7 days.

When hovering over the graph, a vertical line that corresponds to a date on the *x*-axis is displayed (Figure 7(g)). It allows the user to browse the values of the selected indicator at the geographical entity level in detail. Starting from the top down, the points correspond to the maximum value, the value for the entity, the third quartile, the median, the first quartile and the minimum value for that date. The exact values are shown in Figure 7(h). The top part corresponds to the date selected in Figure 3(e), which is the date for which the values are displayed on the map. The bottom part shows the values for the date hovered over in the tooltip. The vertical line of Figure 7(f) is very similar to Figure 7(g) but corresponds to the date on the map. It is updated according to the modifications made

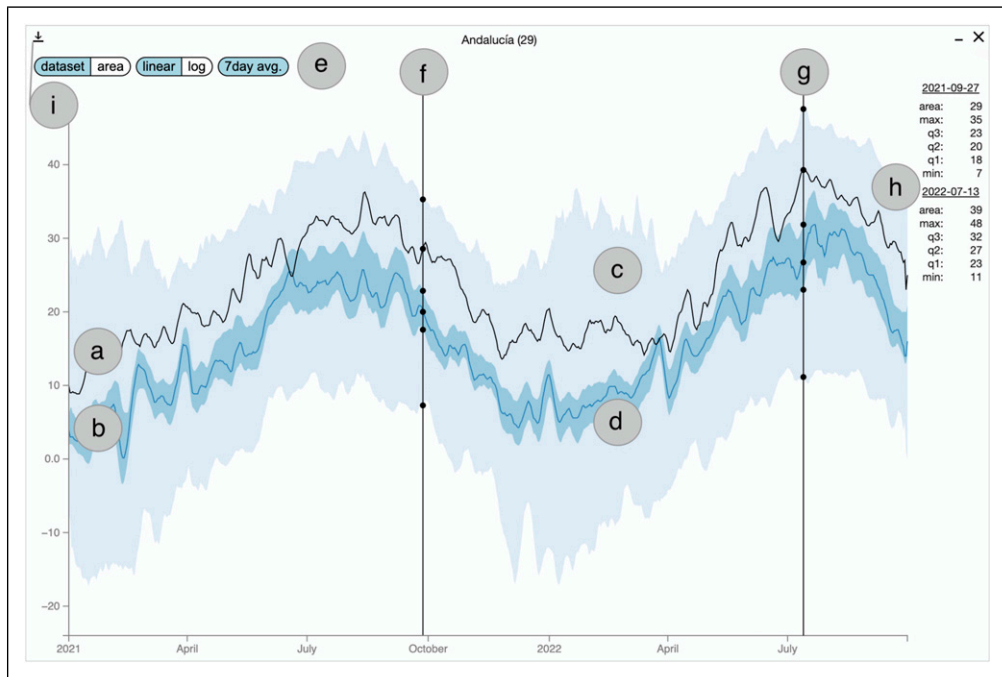


Figure 7. Tooltip components.

in [Figure 3\(f\)](#). If users wish to perform a more in-depth analysis of the selected data, they can easily download it in CSV format by clicking on the download icon at the top left ([Figure 7\(i\)](#)).

Functionalities of the 2-maps view

The 2-map view facilitates the comparison of different time periods, geographical regions, and indicators. These three elements correspond to the dimensions mentioned for requirement R1. The 2-map view consists of a juxtaposition of two maps next to each other as shown in [Figure 8](#). These two maps can be navigated in exactly the same way as the single map.

Synchronization. We use a padlock system to implement synchronization and thus facilitate comparison ([Figure 8\(a\)–\(c\)](#)). This padlock system allows any combination of the 3 dimensions of requirement R1 to be used for comparison.

The user can synchronize the navigation of the two maps. By locking the padlock at the top ([Figure 8\(a\)](#)), the slightest change to one of the maps (zoom or pan) is automatically applied to the second map. Conversely, if the padlock is unlocked, the two maps are completely independent of each other. The principle is exactly the same for the synchronization of the indicator and the dataset ([Figure 8\(a\)](#)) and the choice of date or period ([Figure 8\(c\)](#)).

In the example of [Figure 8](#), the padlocks for map navigation ([Figure 8\(a\)](#)) and synchronization of the temporal dimension ([Figure 8\(c\)](#)) are locked. The geographical area for both maps is Europe and the selected period is January 2023. The third padlock, associated with the choice of indicator, is unlocked ([Figure 8\(b\)](#)). This means that each map can represent the values of a different indicator. This is indeed the case: the map on the left shows the values for “Total confirmed cases” while the

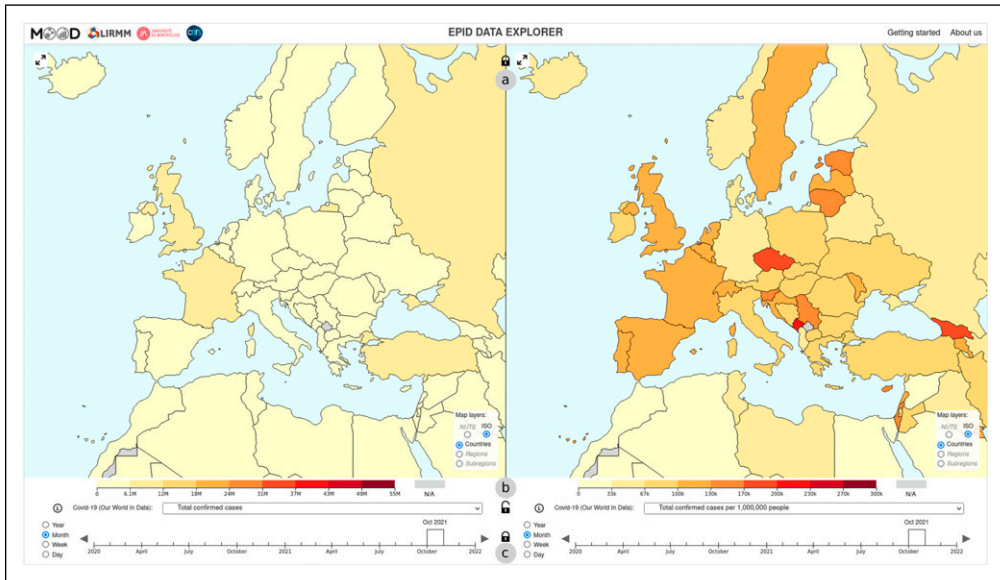


Figure 8. Additional components in the 2-map view.

map on the right shows “Total confirmed cases per 1,000,000 people”. In this example, we can therefore compare these two indicators in Europe for the period January 2023, which corresponds to requirement R1.

System architecture

During the application design phase, we identified a close relationship between requirements R4 and R5: users want to access both public and private datasets while ensuring the privacy of their own data. The general principle of the architecture is illustrated in Figure 9.

The Epid Data Explorer application is available online with public data. However, to fulfill requirements R4 and R5, a local version of EDE must be installed. First, users need to download the EDE application at: <https://gite.lirmm.fr/advanse/EDE/epid-data-explorer> and install it on their own computer. To further facilitate the process, a new interface called *EDE Datasources Manager* is provided. This interface not only allows users to install and run EDE locally but also enables easy addition of new datasets. When accessing public data, users only need to provide the address of the data source. To enable users to work with their own data, it is necessary to setup a new *EDE Datastore*. For this purpose, an additional interface called *EDE Datastore Manager* is provided. This interface allows users to easily integrate local data in CSV format and perform various data processing tasks such as geolocated aggregation, data formatting, and selecting relevant data. Once the necessary processing is applied, the interface makes the data available, and provides an address that can be used by *EDE Datasources Manager*. It is worth noting that these steps need only to be performed once per dataset. On subsequent analysis, the user just starts their local version of EDE (and any *EDE Datastore* that runs on their local machine) to be able to go back to exploring the data. By using a local data server (*EDE Datastore*), requirement R5 regarding private data access is met. Furthermore, if a user wishes to make their data public for the community, they can simply specify the address provided by the *EDE Datastore Manager*. To assist users in installing these interfaces,

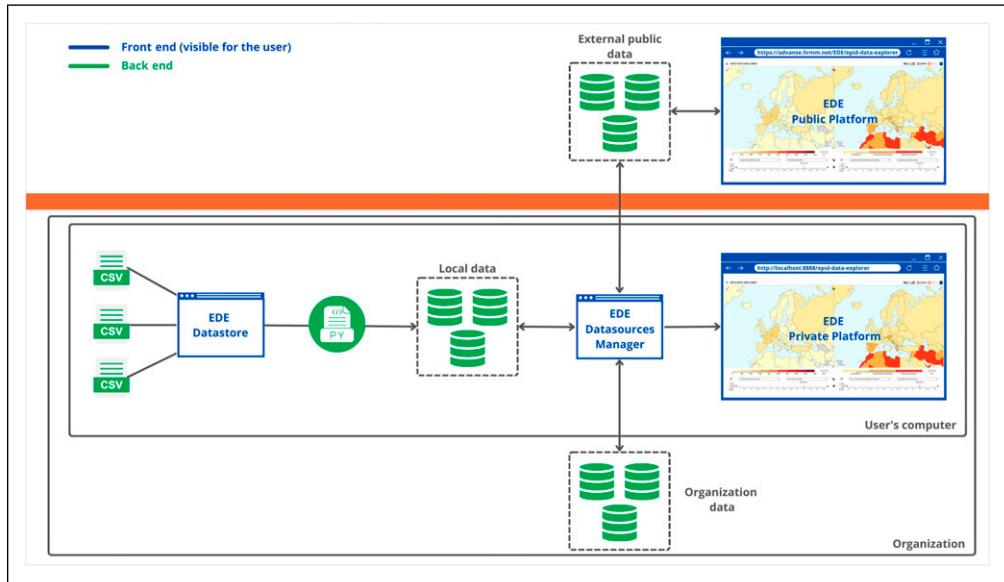


Figure 9. Epid Data Explorer system architecture.

multiple versions are available for different systems (Linux, Windows, and macOS). Furthermore, detailed information is available on the GitLab links provided. Through these two interfaces, we fulfill requirements R4 and R5.

The current design uses the *EDE Datastore* to preprocess and store the data, which allows the user to provide their data “as-is”, and then to decide who can access it. Another secure solution could be enabling the users to load their raw data in their browser, on any instance of EDE, and run our preprocessing pipeline within the browser. This would remove the need to setup a local *EDE Datastore*, but it would have the drawback of requiring the preprocessing pipeline to be run each time. In future versions of EDE, we aim to explore a third option: allowing the users to load already pre-processed CSV files in their browser.

Use cases

In this section, we present two use cases to demonstrate the relevance of EDE’s visualization functionalities.

Restrictions and mobility

During the COVID-19 pandemic, various countries worldwide implemented restrictions to combat the spread of the virus. These measures significantly affected people’s everyday lives, particularly in terms of mobility. For our initial use case, we will focus on the implementation of these restrictions in Europe. EDE allows us to observe the impact thereof, as shown in [Figure 10](#).

Two distinct time periods are depicted: the week of March 2 to 8, 2020, and the week of March 23 to 29, 2020. In [Figure 10\(a\)](#), which corresponds to the first period, Italy is the sole European country to have implemented a lockdown (Denmark had only recommended restrictions), as illustrated by

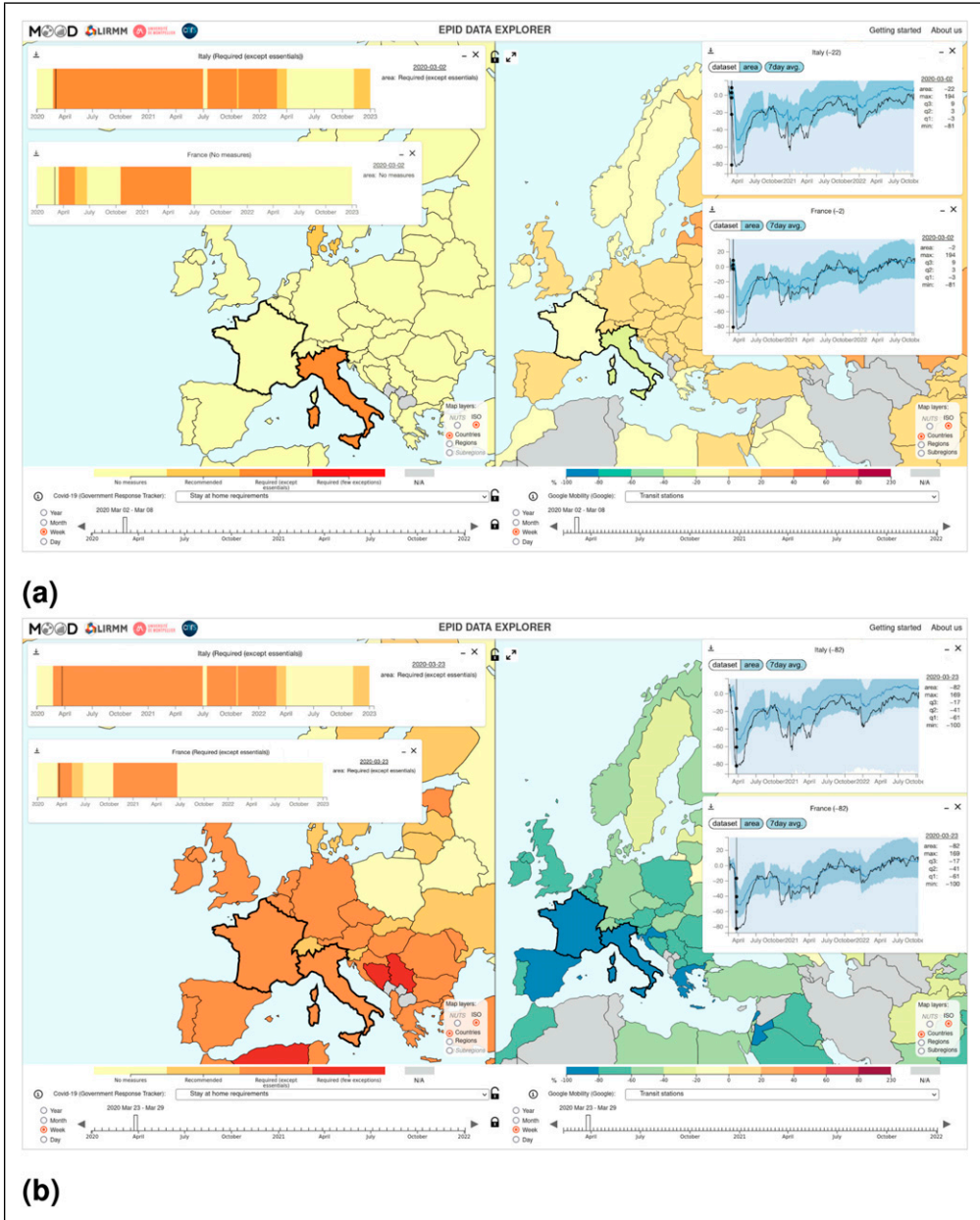


Figure 10. Visualizations of stay-at-home restrictions (maps on the left) and evolution of mobility in transit stations (maps on the right). (a) Visualization for the week of March 2 to 8, 2020. (b) Visualization for the week of March 23 to 29, 2020.

the map on the left. With regard to public transport usage (map on the right), there was a slight decrease in some countries, such as France, Germany, and the UK, while others experienced a slight increase (Spain, Portugal, Ireland, Norway, etc.). Notably, Italy had already witnessed a decline in public transport usage at that moment owing to the implemented restrictions. During the week of

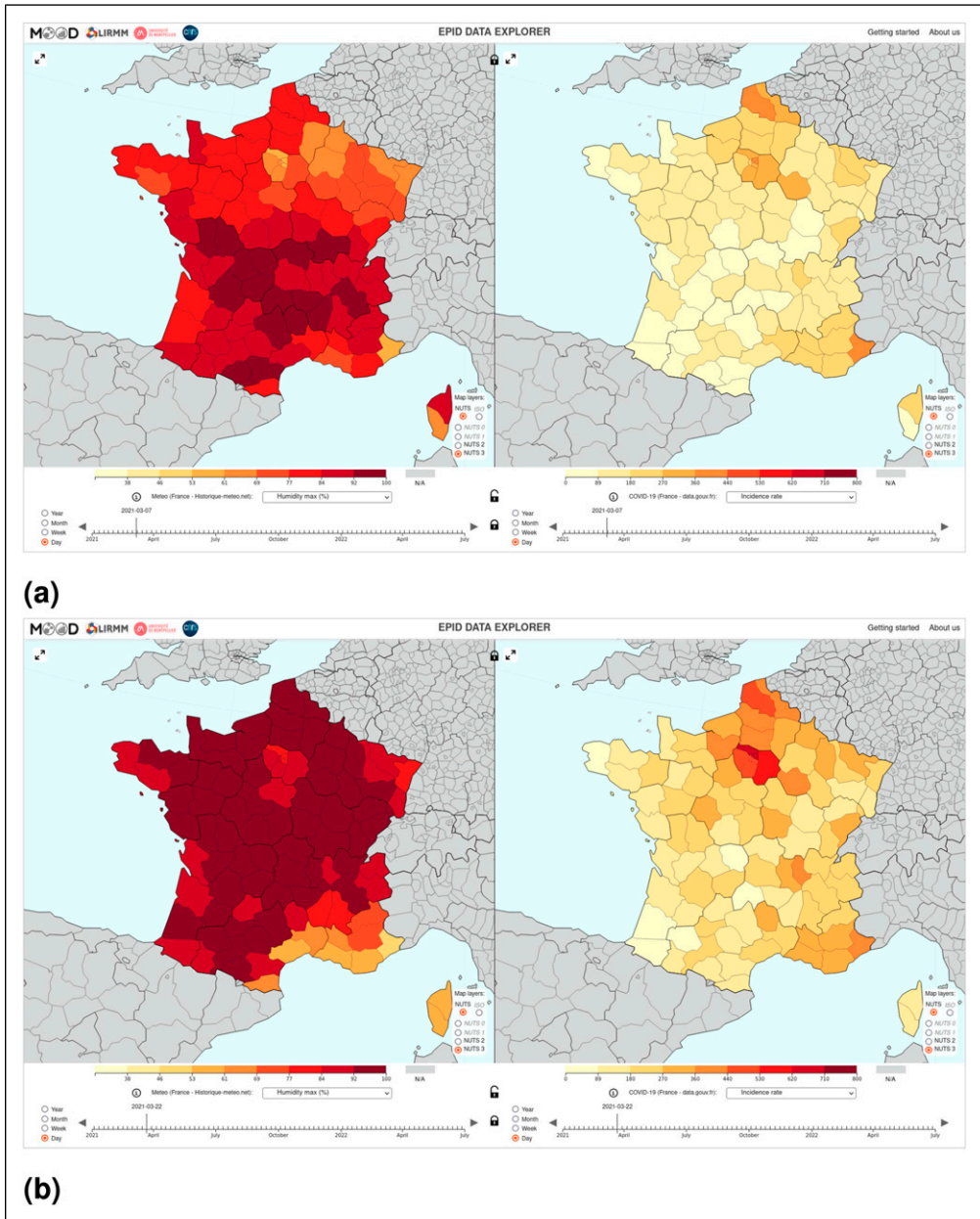


Figure 11. Visualization of the weather (humidity max) and COVID-19 indicators (incidence rate) during the third wave of the pandemic in France. (a) 2021-03-07. (b) 2021-03-22.

March 23 to 29, numerous countries implemented lockdown measures, leading to a widespread and unprecedented impact on mobility. This phenomenon can be easily observed in [Figure 10\(b\)](#). By visualizing the two maps side by side, we can assess the effect of the implemented restrictions on mobility across European countries. Additionally, we can compare the evolution of the same

indicator in multiple countries over the same period. In [Figure 10](#), the tooltips for Italy and France are open on all the maps, enabling a convenient comparison of the implementation of lockdown measures and trends in public transport usage between these two countries. On the left-hand maps, we can observe that Italy implemented measures earlier and maintained them for an extended period, whereas France exhibited two distinct periods: from March 2020 to the summer, and from the end of 2020 onwards. Furthermore, a similar trend in public transport usage can be observed in both countries, but Italy tends substantially lower and more frequently below the median value.

This first use case illustrates the effectiveness of the platform in highlighting phenomena and facilitating comparisons.

COVID-19 and weather

Numerous research studies have extensively examined the correlations between weather conditions and COVID-19. For example, Ganslmeier et al.³⁰ conducted an analysis using a comprehensive dataset. In a separate study, Majumder et al.³¹ performed a meta-analysis. The results highlighted a notable correlation between temperature, humidity, and wind speed and both the death rate and incidence of COVID-19. In a similar study, McClymont et al.³² conducted an analysis of 23 articles after a rigorous selection process. The researchers found that temperature and humidity were frequently reported as a significant factor in many studies.

EDE can be used to analyze data and identify potential correlations that may require further in-depth analysis. For example, if the user wants to examine correlations between weather and COVID-19 in France, they can retrieve data from different sources.^{27,33} With the EDE Datastore interface, this data can be effortlessly integrated. To do this, the original data undergoes pre-processing, during which it is transformed according to geocoding standards such as NUTS. This transformation utilizes data processing resources available from the EDE Data Store. The integrated data can then be seamlessly incorporated into the platform. The data in this example is at the department level (NUTS3), and users can select the aggregation function for each indicator to move to higher levels such as region or country. Finally, users have the flexibility to choose the specific indicators they wish to track. [Figure 11](#) showcases how EDE enables the comparison of weather data and COVID-19 in France during the third wave of the pandemic to identify potential correlations. It allows an analysis of variables like temperature, humidity, and wind speed in relation to the incidence rate or any other chosen indicator across different geographic regions and over time. The left-hand map in [Figure 11](#) displays the maximum humidity recorded, while the right-hand map shows the incidence rate.

This second use case illustrates the usefulness of importing data to explore and observe possible correlations. These observations can then lead to a more in-depth and optimized data analysis by allowing the appropriate approach to be selected.

Conclusion

Epid Data Explorer is a flexible data visualization platform that lets users explore and compare spatio-temporal datasets. The system's architecture has been designed to be able to import and compare any private dataset with public datasets in a secure and simplified manner. As for the visualization tool, it has been designed to enable the detailed exploration of data, as well as the comparison of different indicators. The visualization is interactive: the user can browse the maps, select the date or period of interest, as well as easily change indicators or even datasets. All the visualization functions are available simultaneously in a single window. EDE can be an excellent

tool to support epidemiologists, researchers and healthcare professionals in monitoring infectious diseases or even generating new hypotheses based on their findings when exploring new data.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by EU Grant 874850 MOOD and is catalogued as MOOD 116. The contents of this publication are the sole responsibility of the authors and not necessarily reflect the views of the European Commission.

ORCID iDs

Laetitia Viau  <https://orcid.org/0000-0003-0698-0997>

Vincent Raveneau  <https://orcid.org/0000-0002-0548-2526>

Nancy Rodriguez  <https://orcid.org/0000-0002-5496-1603>

Arnaud Sallaberry  <https://orcid.org/0000-0001-7068-176X>

References

1. Munzner T. A nested model for visualization design and validation. *IEEE Trans Vis Comput Graph* 2009; 15(6): 921–928.
2. Sedlmair M, Meyer MD and Munzner T. Design study methodology: reflections from the trenches and the stacks. *IEEE Trans Vis Comput Graph* 2012; 18(12): 2431–2440.
3. Gleicher M, Albers D, Walker R, et al. Visual comparison for information visualization. *Inf Visual* 2011; 10(4): 289–309.
4. Broeck WVD, Gioannini C, Gonçalves B, et al. The gleamviz computational tool, a publicly available software to explore realistic epidemic spreading scenarios at the global scale. *BMC Infect Dis* 2011; 11(1): 37.
5. Antoniou S-E, Dórea F, Dupuy C, et al. Syndromic surveillance: developing an early warning system for rabies. *EFSA J.* 2022; 19(12): 7785E.
6. Goel R, Valentin S, Delaforge A, et al. Epidnews: extracting, exploring and annotating news for monitoring animal diseases. *Journal of Computer Languages* 2020; 56: 100936.
7. Fadloun S, Sallaberry A, Mercier A, et al. Epidvis: a visual web querying tool for animal epidemiology surveillance. *Inf Visual* 2020; 19(1): 48–64.
8. Kuo YH, Martínez-López B and Ma KL. Investigating animal infectious diseases with visual analytics. In: *16th Pacific Visualization Symposium (PacificVis)*. Piscataway: IEEE, 2023, pp. 71–81.
9. Freifeld CC, Mandl KD, Reis BY, et al. Healthmap: global infectious disease monitoring through automated classification and visualization of internet media reports. *J Am Med Inf Assoc* 2008; 15(2): 150–157.
10. Meng Z, Okhmatovskaia A, Polleri M, et al. Biocaster in 2021: automatic disease outbreaks detection from global news media. *Bioinformatics* 2022; 38(18): 4446–4448.
11. IHME. Institute for Health Metrics and Evaluation. *Epi visualization*. Seattle, WA: IHME, University of Washington, 2020. <https://vizhub.healthdata.org/epi>

12. Rosling H and Zhang Z. Health advocacy with gapminder animated statistics. *J Epidemiol Glob Health* 2011; 1(1): 11–14.
13. Claes F, Kuznetsov D, Liechti R, et al. The empres-i genetic module: a novel tool linking epidemiological outbreak information and genetic characteristics of influenza viruses. *Database* 2014; 2014: bau008–bau013.
14. Dong E, Du H and Gardner L. An interactive web-based dashboard to track covid-19 in real time. *Lancet Infect Dis* 2020; 20(6): 533–534.
15. Mathieu E, Ritchie H, Rodés-Guirao L, et al. Coronavirus pandemic (covid-19). <https://ourworldindata.org/coronavirus>
16. Chen AT, Altschuler K, Zhan SH, et al. COVID-19 CG enables SARS-CoV-2 mutation and lineage tracking by locations and dates of interest. *Elife* 2021; 10: e63409.
17. Chen C, Nadeau S, Yared M, et al. CoV-Spectrum: analysis of globally shared SARS-CoV-2 data to identify and characterize new variants. *Bioinformatics* 2022; 38(6): 1735–1737.
18. Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 2018; 34(23): 4121–4123.
19. Bernasconi A, Gulino A, Alfonsi T, et al. VirusViz: comparative analysis and effective visualization of viral nucleotide and amino acid variants. *Nucleic Acids Res* 2021; 49(15): e90.
20. Bernasconi A and Grandi S. A conceptual model for geo-online exploratory data visualization: the case of the COVID-19 pandemic. *Information* 2021; 12(2): 69.
21. Munzner T. *Visualization Analysis and design. A.K. Peters visualization series, A K peters*. Boca Raton: CRC Press, 2014.
22. ECDC. European Centre for disease prevention and Control. <https://www.ecdc.europa.eu/en>
23. OWD. Our World in data. <https://ourworldindata.org>
24. Google. Community mobility reports - covid-19. <https://www.google.com/covid19/mobility/>
25. Oxford. COVID-19 government response tracker. <https://www.bsg.ox.ac.uk/research/covid-19-government-response-tracker>
26. Obépine, wastewater-based epidemiology. <https://www.reseau-obepine.fr>
27. Météo. Historique - Météo. <https://www.historique-meteo.net>
28. ISO. ISO 3166 - country codes. <https://www.iso.org/>
29. NUTS. Nuts - nomenclature of territorial units for statistics. <https://ec.europa.eu/eurostat/web/nuts/>
30. Ganslmeier M, Furceri D and Ostry JD. The impact of weather on covid-19 pandemic. *Sci Rep* 2021; 11(1): 22027.
31. Majumder P and Ray PP. A systematic review and meta-analysis on correlation of weather with covid-19. *Sci Rep* 2021; 11(1): 10746–10810.
32. McClymont H and Hu W. Weather variability and covid-19 transmission: a review of recent research. *Int J Environ Res Publ Health* 2021; 18(2): 396.
33. Datagouv. Synthèse des indicateurs de suivi de l'épidémie COVID-19. <https://www.data.gouv.fr/fr/datasets/>