Journal of Biomedical Informatics 117 (2021) 103760

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

COnVIDa: COVID-19 multidisciplinary data collection and dashboard

Enrique Tomás Martínez Beltrán [a,*], Mario Quiles Pérez [a], Javier Pastor-Galindo [a], Pantaleone Nespoli [a], Félix Jesús García Clemente [b], Félix Gómez Mármol [a]

[a] Department of Information and Communications Engineering, University of Murcia, Murcia 30100, Spain
[b] Department of Computer Engineering, University of Murcia, Murcia 30100, Spain

ARTICLE INFO

Keywords:
COVID-19 pandemic
SARS-CoV-2
Dashboard
Data gathering
Data visualization

ABSTRACT

Since the first reported case in Wuhan in late 2019, COVID-19 has rapidly spread worldwide, dramatically impacting the lives of millions of citizens. To deal with the severe crisis resulting from the pandemic, worldwide institutions have been forced to make decisions that profoundly affect the socio-economic realm. In this sense, researchers from diverse knowledge areas are investigating the behavior of the disease in a rush against time. In both cases, the lack of reliable data has been an obstacle to carry out such tasks with accuracy. To tackle this challenge, COnVIDa (https://convida.inf.um.es) has been designed and developed as a user-friendly tool that easily gathers rigorous multidisciplinary data related to the COVID-19 pandemic from different data sources. In particular, the pandemic expansion is analyzed with variables of health nature, but also social ones, mobility, etc. Besides, COnVIDa permits to smoothly join such data, compare and download them for further analysis. Due to the open-science nature of the project, COnVIDa is easily extensible to any other region of the planet. In this way, COnVIDa becomes a data facilitator for decision-making processes, as well as a catalyst for new scientific researches related to this pandemic.

1. Introduction

On December 31, 2019, the World Health Organization (WHO) China Country Office was informed of a number of patients with a pneumonia of unknown nature, detected in the city of Wuhan in the province of Hubei, China. Later on, such phenomenon was clarified as the outbreak of a novel coronavirus-caused respiratory disease (i.e., COVID-19). COVID-19 rapidly spread worldwide, being declared as a pandemic on March 11, 2020, by the WHO[1].

The severe health and social crisis resulting from the COVID-19 pandemic has forced institutions and authorities to make decisions that are transcendental for the lives of millions of citizens [1]. In fact, people from many different countries have suffered mobility restrictions, even lockdowns in the worst cases, which have harshly affected them from a socio-economic perspective [2,3]. In this regard, scientists from the most diverse research disciplines are struggling ever since to investigate the behavior of the disease against the clock. For instance, virologists and epidemiologists from medical sciences are committed to a continuous battle to find novel treatments to counteract the effects of the virus [4]. Likewise, economists make an effort to mitigate the negative economic impact [5] of the epidemic, while computational area researchers study the evolution of the pandemic to predict potential changes [6].

In both cases, the lack of rigorous and reliable data has undoubtedly been an unfortunate impediment to carry out these tasks with accuracy and precision [7]. The motivation of such data scarcity lies mainly in the novelty of the virus: every individual is currently facing a threat with no antecedents. In such an uncertain context, the causal inference from observational data [8] becomes an allied methodology to understand how and why this health crisis is unfolding. Additionally, the management and communication of official data made by the public authorities have often been questionable. Therefore, the acquisition and analysis of data from trusted sources represent a need more than ever, endowing humans with useful information against both the COVID-19 and the spread of disinformation [9].

Under these premises, COnVIDa[2] has come to fill this gap by offering

* Corresponding author.
E-mail addresses: enriquetomas.martinezb@um.es (E.T. Martínez Beltrán), mario.quilesp@um.es (M. Quiles Pérez), javierpg@um.es (J. Pastor-Galindo), pantaleone.nespoli@um.es (P. Nespoli), fgarcia@um.es (F.J. García Clemente), felixgm@um.es (F. Gómez Mármol).

[1] https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen.
[2] https://convida.inf.um.es.

https://doi.org/10.1016/j.jbi.2021.103760
Received 15 October 2020; Received in revised form 15 March 2021; Accepted 20 March 2021
Available online 30 March 2021

a user-friendly tool to easily collect relevant COVID-19 pandemic data in the context of Spain from different and trusted data sources. Additionally, COnVIDa permits to join such data, compare them on a geographic basis and, if desired, download them in commonly-used data analysis formats for subsequent analysis. Besides, COnVIDa has been designed and developed following the open science principles [10], including open data sources during the project lifecycle [11]. Given the open science nature of the project, the tool is easily extensible to any other region of the planet. Furthermore, COnVIDa is implemented as a robust and modular web dashboard, with different components working in synergy to offer an excellent user experience [12].

In this way, COnVIDa becomes a data facilitator for decision-making processes, as well as a catalyst for new scientific research related to this pandemic. In this sense, the tool can help users to study the underlying correlations between several parameters, such as the evolution of the pandemic, the containment measures adopted by the authorities, socio-economic and educational implications, and so forth [13]. More specifically, thanks to COnVIDa, it is possible to study the impact of lockdowns or preventive measures on the population from different viewpoints. In particular, the pandemic expansion is not only analyzed concerning variables of a health nature (e.g., increase in cases, deaths, etc.), but also in relation to parameters such as employment, consumption, tourism, and so on. Furthermore, it will also help to predict possible future outbreaks by observing, measuring, and cross-relating certain outstanding variables (e.g., the number of active cases with Internet searches related to COVID-19 symptoms). It is worth mentioning that any correlation detected must be analyzed and processed with rigor, not meaning in any case causality. In this sense, it is impossible to deduce a cause-effect relationship among different parameters by merely observing a correlation or association among them solely [14].

This tool effectively responds to the necessity of several individuals who need to work with reliable and multidisciplinary data from different sources, all linked to the COVID-19 pandemic in Spain. In our vision, those people do not necessarily possess the technical skills to automate the tedious task of data collection and representation.

At the time of writing this manuscript, COnVIDa incorporates, in addition to data of COVID-19 in Spain, data from INE (i.e., Spanish National Institute of Statistics), Mobility (provided by Google and Apple), MoMo (i.e., Daily Mortality Monitoring System from the Spanish Health Institute Carlos III), and AEMET (i.e., Spanish State Agency of Meteorology). Nonetheless, due to its modular architecture, it is easily extensible to integrate additional relevant data sources. It is, in fact, a very versatile tool, easily adaptable to geographic regions with different granularity (region, province, health area, municipality, etc.), and even to other countries.

The remainder of this paper is organized as follows. Section 2 depicts the principal characteristics of COnVIDa, which are compared with other similar tools. In Section 3, the COnVIDa Data Sources are presented and explained, motivating on their choice. Next, Section 4 presents the modular architecture of COnVIDa with a particular focus on the back-end modules. Then, Section 5 describes the COnVIDa dashboard, designed to be user-friendly and highly intuitive. Also, Section 6 describes the core element of the framework, which implements the logic to store and provide data from different data sources. In Section 7, the REST API developed within the COnVIDa framework is briefly explained. Additionally, Section 8 presents an in-depth discussion about COnVIDa, arguing the utility of the proposed platform and the lessons learned throughout its design and development. Finally, Section 9 concludes the paper, presenting some interesting future lines to further improve this useful tool.

2. COnVIDa's novelty

COnVIDa is a web-based platform that displays day-to-day updated data related to the impact and conditions of COVID-19 in Spain. The application, shown in Fig. 1, presents a web-based, user-friendly dashboard for interactive plotting. The website's main objective is to facilitate the visualization, interpretation, and comparison of high-dimensional data of different nature, considering the selected time range and locations.

2.1. Key features

Going into detail, COnVIDa is strongly characterized by the following key features:

- **F1. Multidisciplinary** COnVIDa is designed to collect and support different types of information. In the current version, we include (i) COVID-19 medical statistics to monitor the epidemiological evolution [15]; (ii) citizens' lifestyle and health measurements to reveal potential correlations [16], (iii) numbers in terms of mobility to analyze agglomerations or impact of lockdowns/reopenings [17], (iv) records on the excess of mortality to study actual deaths due to the pandemic [18], and (v) weather statistics to explore a possible environmental sensitivity [19]. These data sources are widely discussed later in Section 3.
- **F2. Region-sensitive** COnVIDa currently offers the data per Spanish region so that the end-users can use them as filters. Such fine-grained observation permits a more accurate monitoring than for the entire country. Additionally, it enables direct comparison between territories that may have a different pandemic, economic, social, or political situation. At the time of writing, COnVIDa supports Spanish communities and provinces as reference units, but the list could be extended.
- **F3. Date-sensitive** COnVIDa presents the data per day, so the end-user can select the range of days to analyze. By default, the suggested time window is from 2020-02-21 (the first day with official COVID-19 clinical records in Spain) to the present day. The current version of COnVIDa considers a day as the standard temporal unit, but a different granularity (week, month, etc.) could be included in the future.
- **F4. Flexible** COnVIDa adapts to the user's concerns by supporting any combination of features F1, F2, and F3. From scratch, the end-user may select i) the start and end dates, ii) the desired Spanish regions (or all), and iii) the data items to analyze for the selected time window and regions to finally build the associated plot. The tool also allows to choose between iv) line or bar plot and v) linear or logarithmic scale.
- **F5. Dual** COnVIDa uses two visualization panels to draw the selected input conditions. The first one maintains the focus on a temporal analysis while the other deals with a regional analysis. The *temporal visualization* shows the plot with the selected days on the X-axis and daily values on the Y-axis. The *regional visualization* draws boxplots with the selected regions on the X-axis to

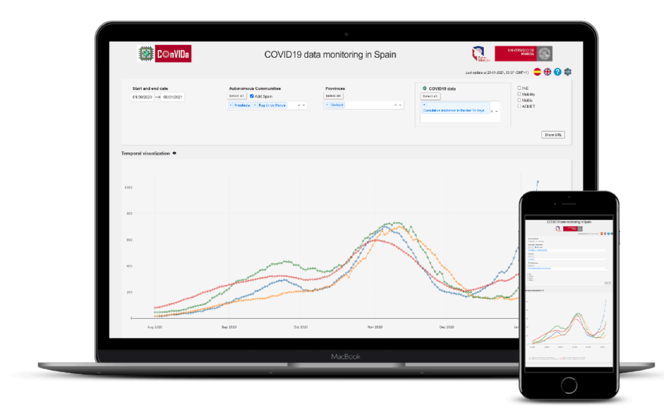


Fig. 1. COnVIDa screenshot in PC and smartphone.

represent descriptive statistics of the daily values for the selected date range. Additionally, a map is deployed to visualize the selected metrics geographically per region.

F6. Overlayable COnVIDa does not restrict selecting different data sources, even if they are of a different nature (F1). In this sense, the end-user can superimpose and compare data on COVID-19, excess mortality, lifestyle, meteorology or mobility, on the same visualization panels altogether. Axis sharing is a powerful way to visually and quantitatively compare multidisciplinary data.

F7. Downloadable COnVIDa allows the manual download of the filtered data in standardized CSV or XLS files to facilitate further analysis by end-users, data analysts, or scientists. This framework also implements an application programming interface (API) endpoint for programmers to retrieve the data in JSON format. The downloaded data can be used to predict pandemic dynamics such as COVID-19 infections and hospitalizations, or apply any machine learning analysis, as evidenced by [20]. It is also possible to generate a table in HTML format (which can be useful for journalists). Moreover, the generated graphs can also be downloaded in PNG format.

F8. Easy-to-use COnVIDa implements a simple interface that intuitively offers the filtering and plotting functionalities. The design has been improved through several releases considering both beta-testers and end-users recommendations and impressions. A summary table with the most remarkable statistics (counts, means, standard deviations, min, max, and percentiles) of the represented data has also been included under the regional visualization panel.

F9. Open and extensible COnVIDa is technically designed to be extended with new indicators and to promote collaborative work. In this sense, to facilitate the integration of new data sources, the main functionalities are decoupled and reusable, the data sources harvesting is programmed in a modular fashion, and the data representation is standardized regardless of the source. Moreover, all of the tool's source code is openly available[3] to the experts interested in contributing or reusing it.

2.2. Comparison with similar tools

The properties mentioned above make COnVIDa an innovative and alternative tool among the existing ones. In this line, Table 1 analyzes other similar platforms with pandemic monitoring purposes providing (some of) the features mentioned above. Despite the many tools that emerged at the beginning of the pandemic, we only considered those websites with interactive dashboards that are active and maintained at the time of writing. To this extent, we have selected five among the most searched COVID-19 dashboards based on Google indexing, together with the unique other multidisciplinary platform found on the web. Although some tools focus on countries and others explore regions of a particular country, their functionalities help contextualize the key features of COnVIDa.

Firstly, all the analyzed dashboards manage time series per geographical region, highlighting the importance of daily monitoring throughout territories to follow the spread of the pandemic. Additionally, the platforms tend to be very easy to use, with user-friendly interfaces and dual visualizations for temporal tendencies and regional statistics.

Secondly, it is less common for the websites to have options for downloading the raw data. Three of them allow to download just the requested data specifically, and another tool facilitates the whole dataset. On the other hand, it is hard to find open projects with publicly available source code and detailed documentation. In this sense, only one engine of the analyzed tools was published and remarkably documented.

In this way, only the DatAC framework and COnVIDa offer a multidisciplinary view of the pandemic. The former is designed for comparing COVID-19 data with climatology, air quality, and demographic information. In contrast, our platform permits the combined plotting of COVID-19 data together with mobility, excess mortality, social and weather statistics in the same shared graphic. Compared with the rest of the tools, both platforms are highly flexible, offering enormous freedom for filtering specific scenarios and configuring variables, regions, and dates at will.

However, DatAC only offers the crossing of two metrics at the same time. In this sense, COnVIDa is more powerful in terms of cross-comparisons between metrics, being able to superpose the desired number of different data series on the same overlayable panel. Besides, we opted for a project based on open science, and the source code of both the internal engine and the website is published and documented in detail for anyone willing to replicate it or contribute to further extend it.

Finally, the internal engine of our solution is easily extensible for developers to program new data sources. Thus, everyone can contribute to our platform or test their theories beyond our proposed data sources offered as standard. We have not observed such feature in any other analyzed application, except for OWiD Explorer, that maintains a highly active and well-documented GitHub repository.

3. COnVIDa data sources

As mentioned above, one of the strengths of the COnVIDa framework is the multidisciplinary nature of the data. These relevant data come from different official data sources to provide reliable information about the pandemic effects and avoid the spread of disinformation.

In addition, each selected data source is composed of one or more data items. A data item is a low-grain resource that codifies a specific piece of information and belongs to one of the aforementioned data sources.

Next, each of the COnVIDa data sources is deeply described and motivated. For an easy reference, see Table 2. The current version of COnVIDa includes five data sources related to the COVID-19 pandemic in Spain.

3.1. COVID-19 data source

The first data source is offered by *esCOVID19data*[4], which offers daily data at community and province levels about the COVID-19 evolution and its most essential aspects, such as cases confirmed by PCR (i.e., Polymerase Chain Reaction), PCR tests, vaccines, or deaths, among others. The official repository refreshes the data every day.

This data source is the spinal column of COnVIDa among all the supported data sources. It is designed for users who want to acquire information about the COVID-19 pandemic evolution effortlessly.

By leveraging the community and province-level data acquired daily, other data items not directly available in the data source are automatically calculated by COnVIDa. For instance, the country-level data series of Spain, or aggregated values such as the *cumulative lethality* or the *percentage of population vaccinated*, are derived from Autonomous Communities data. In other cases, data items are used to calculate averages per 100,000 inhabitants, such as *cumulative cases per 100,000 inhabitants* or *cumulative deaths per 100,000 inhabitants*.

3.2. INE data source

Another data source is provided by the INE[5] (Spanish National

[3] https://github.com/CyberDataLab/COnVIDa.

[4] https://github.com/montera34/escovid19data.

[5] https://www.ine.es.

Table 1
Comparison with other active COVID-19 dashboards.

Tool	Multi-disciplinary	Region-sensitive	Date-sensitive	Flexible	Dual	Overlayable	Downloadable	Easy	Open-extensible
OWiD Explorer[a]	✗	✓	✓	✓	✓	✗	✓	✓	✓
WHO Dashboard[b]	✗	✓	✓	✗	✓	✗	≈	✓	✗
JHU Dashboard[c]	✗	✓	✓	✗	✓	✗	✗	≈	✗
COVID-19 in UK[d]	✗	✓	✓	✗	✓	✗	✓	≈	✗
Zoho Dashboard[e]	✗	✓	✓	✗	✓	✗	✗	✓	✗
DatAC[f]	✓	✓	✓	✓	✓	≈	✓	✓	✗
COnVIDa	✓	✓	✓	✓	✓	✓	✓	✓	✓

Legend: ✓ Yes – ✗ No – ≈ Partially.
[a] https://ourworldindata.org/coronavirus-data.
[b] https://covid19.who.int.
[c] https://coronavirus.jhu.edu/map.html.
[d] https://coronavirus.data.gov.uk.
[e] https://www.zoho.com/covid/spain.
[f] https://covid19.genyo.es.

Institute of Statistics). It particularly offers information about different aspects of the Spanish reality, such as physical activity, tobacco consumption, etc. These parameters, which mainly describe Spanish communities, are absolute statistics not based on time series. By their nature, these statistics are not updated frequently on their original platform.

This data source may be useful for analyzing how social or health factors potentially affect the spread of the virus. The data from INE are calculated for every 1,000 inhabitants and are intended for users dedicated to research.

3.3. Mobility data source

In the same way, information about the mobility of the citizens is collected from two relevant sources. On the one hand, information about mobility gathered from Google[6], and on the other hand, the information provided by Apple[7]. The temporal granularity of these data is daily, while its regional granularity is constrained to Spanish communities. This information helps to understand the spread of the virus throughout those regions, which may affect the normal mobility of citizens across distinct areas (e.g., parks, residential, etc.). In this case, mobility values are updated by the official data sources every day.

3.4. MoMo data source

To evaluate the real mortality caused by the COVID-19, we should consider daily deaths regardless of cause and compare them with the death trend from previous years. In Spain, these official numbers can be obtained from the Spanish Health Institute Carlos III[8]. Therefore, COnVIDa acquires time series about the total number of actual deaths per day from this data source (taken from civil registrations, cemeteries, and undertakers registries), or expected deaths per day (based on previous statistics), both in terms of Spanish communities. Besides, relevant statistics are calculated on those data, e.g., the 1st and 99th percentiles of the series. Aligned with the temporal granularity, the official data is refreshed daily in the external source.

3.5. AEMET data source

To determine how the climate and weather conditions may eventually affect the pandemic, meteorological data are obtained from the AEMET[9] ("*Agencia Estatal de Meteorología*", the Spanish State Agency of Meteorology). The retrieved metrics corresponding to this data source (e.g., minimum/maximum temperature, rainfall, solar radiation, etc.) are organized by day and Spanish community. It is worth mentioning that the agency publishes, every day, the weather statistics corresponding to fives day ago.

4. Architecture overview

From a technical perspective, COnVIDa is a synergy of various elements implemented strategically in a web server to provide and display data securely, promptly, and reliably, as shown in Fig. 2. The resulting architecture is composed of the following modules:

1. *Web Application Firewall (WAF)*. Security mechanism that monitors, filters, and blocks suspicious HTTP traffic to and from COnVIDa. As revealing the configuration (i.e., the filtering rules) of this element would compromise the security of the service, the WAF implementation is not published in the repository.
2. *Dashboard*. Visual panel with which the end-users interact on the website to visualize the data in plots and summary tables. The information is extracted locally from a cache where all available data is updated. The dashboard is presented in detail in Section 5.
3. *COnVIDa library*.[10] Python implementation of the logic of the service, widely explained in Section 6. On the one hand, COnVIDa-server processes end-user requests and extracts the associated data from the local data cache, which is updated regularly. On the other hand, COnVIDa-lib implements the atomic functions and classes for externally requesting information from the official data sources (for example, in the update procedure).
4. *COnVIDa API REST*. Definition of endpoints for downloading data in JSON format programmatically, especially handy for analysts, developers, and programmers. This module also returns the information that is kept in the local cache. The functioning of the API REST is in-depth discussed in Section 7.

It is worth mentioning that these modules are not deployed physically on dedicated servers. We only maintain one server that runs each module in a Docker[11] container. This container-based distributed design benefits from the advantages of virtualization [21], such as scalability, dynamic management, and agile deployment. Therefore, the modules are interconnected in the same internal virtualized network.

[6] https://www.google.com/covid19/mobility.
[7] https://www.covid19.apple.com/mobility.
[8] https://www.isciii.es.
[9] http://www.aemet.es.
[10] https://github.com/CyberDataLab/COnVIDa-lib.
[11] https://www.docker.com.

Table 2
COnVIDa data sources.

Data source	Description	Data items	Temporality	Regionality
COVID-19	Information about the COVID-19 pandemic, from *esCOVID19data*	Cumulative incidence in the last 14 days, new daily cases, death cases, percentage of vaccinated population, cumulative lethality, cumulative COVID19 cases, deaths in the last 7 days, cumulative hospitalized cases, cumulative ICU cases, daily cases of antibodies, cumulative cases recovered, cumulative vaccines provided, cumulative vaccines supplied, percentage of new vaccines supplied, among others.	Daily	Communities, Provinces
INE	Information about different aspects of the Spanish reality, from Spanish National Institute of Statistics	Physical activity, body mass index (BMI), tobacco consumption, household by family type, households by occupation density and people over 65 years old who live alone	Non-temporal	Communities
Mobility	Information about citizens' mobility, from Google and Apple	Mobility in different spaces such as grocery and pharmacy, parks, residential, retail and recreation, transit stations, workplace and vehicles (driving)	Daily	Communities
MoMo	Information about the mortality monitoring system, from the Spanish Health Institute Carlos III	Daily observed deaths, lower and upper bounds of such series, the daily expected deaths, and the 1st and 99th percentiles of such series	Daily	Communities
AEMET	Information about meteorological data stemming, from AEMET (Spanish State Agency of Meteorology)	Rainfall, maximum/minimum pressure, maximum gust, isolation, maximum/minimum temperature, mean temperature, wind speed, altitude and gust direction	Daily	Communities

5. Dashboard for data visualization

As previously mentioned, COnVIDa implements a web-based dashboard for the visualization of the collected data. The proposed solution, developed leveraging Plotly Dash[12] capabilities, works as the front-end of the tool, where users can interact to explore diverse data sources. Particularly, this web interface enables one to intuitively select data of interest to analyze and compare them with other relevant data and the possibility of downloading the retrieved information in different commonly-used data analysis formats for further investigations.

Besides, the primary purpose of the dashboard is to provide a fast, flexible, and intuitive way of checking data, in addition to contrasting data for scientific researches about the SARS-CoV-2 pandemic. To achieve this ambitious objective, the dashboard has been divided into different modules, each one performing a specific task (i.e., data selection, graphing, summarizing). On the one hand, different variables can be selected to create connections between them. On the other hand, the data are represented in graphs and summary tables that the user can export or display in different formats.

Furthermore, the design of the dashboard is geared towards ease of use. Indeed, it is optimized to generate automatic graphs and offer clear explanations of the web interface elements. The principal objective of such a design is to make any user perceive a sense of order within the page, thus facilitating its wide use and acceptance.

5.1. Dashboard views

The dashboard itself is in turn composed of the following main panels:

M1. Data selection panel It is in charge of selecting the specific data to be represented, as illustrated in Fig. 3. To start the display, a range of dates should be chosen. Then, users should select the specific region(s) for which the platform has to retrieve information. Later, users should mark which data item(s) within the provided data sources will be plotted. To facilitate the visualization of this panel, the COVID-19 data source and relative data items are shown by default, although users can select and display any additional data source at will. From this point, the last configuration of the user is saved for his/her next visit to the platform.

M2. Temporary data panel In this panel, the selected data items are represented on a chart indexed by a standard time unit (up to date, only days are supported). As indicated in Table 3, although the INE data source cannot be considered for temporal representation, the rest of the data items offer daily metrics that can be crossed together. For instance, a user could select one data item from each temporal data source, two different regions, and a time window from 21/02/2020 until 8/03/2020. As a result, the X-axis will show all the days between those two dates. In contrast, the Y-axis will show the corresponding values for these data items in the regions. Therefore, this panel allows obtaining a graphical representation of the information (as reported in Fig. 4).

M3 Regional data panel In this panel, the selected data items are represented to categorize each region with multiple statistical values per variable. While the values of COVID-19, MoMo, AEMET, and Mobility are temporary, the INE data are geographical and therefore need to be transformed to adjust to the temporary representation. On the left, each selected region is represented on the lower axis with a set of boxplots (one chart per data item) that enable the direct comparison with basic statistical values (Fig. 5a)). In a complementary manner, a geographical map is deployed on the right to visualize the desired aggregated value on the selected regions, as shown in Fig. 5a. Additionally, a summary table is also built characterizing each selected data item (see Fig. 5c).

5.2. Integrated view

Firstly, it is worth remarking that all data sources are programmed and integrated into COnVIDa through the same methodology, which

[12] https://plotly.com/dash.

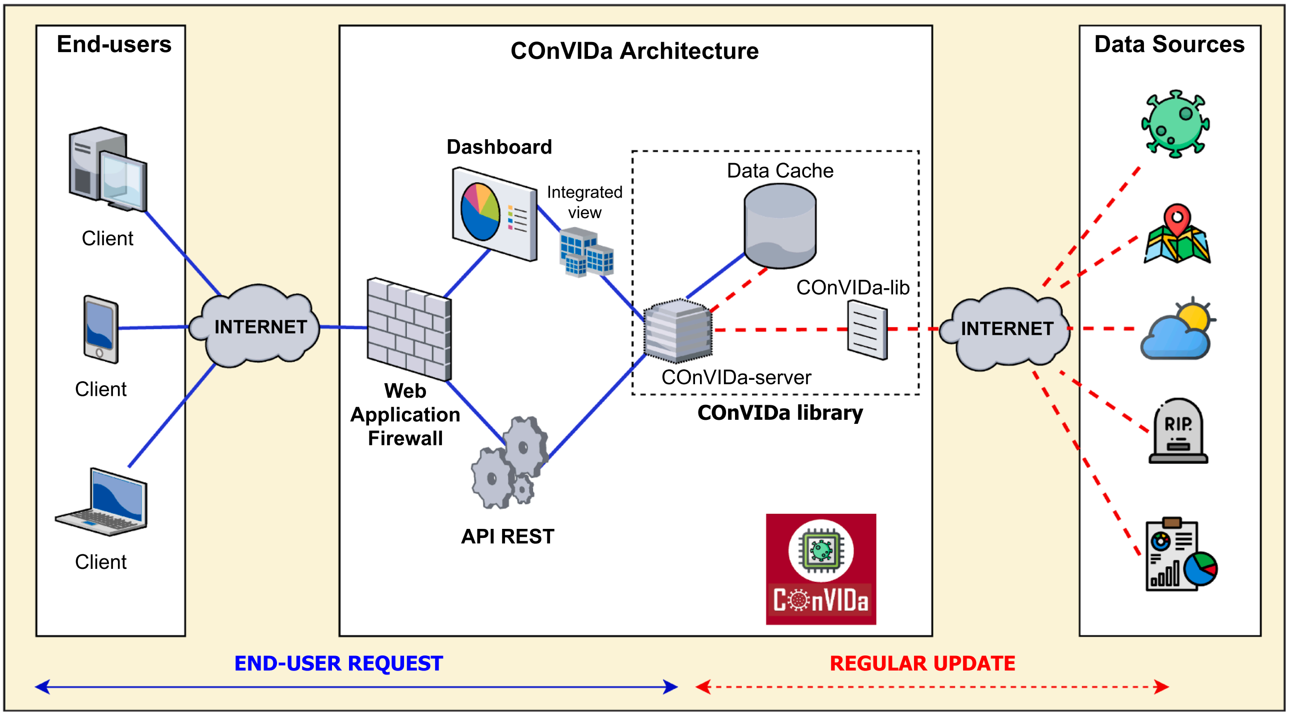


Fig. 2. COnVIDa architecture overview.

makes integration between data items much more direct. In particular, to guarantee a homogeneous data modeling, data sources must respect the interface with the tool through minimum configuration requisites and implementation guidelines. These details are in-depth explained in Section 6.3.

Additionally, COnVIDa defines reference units of time and regions. Currently, we consider the 'day' as the standard time unit (due to its importance in monitoring the pandemic), and both 'communities' or 'provinces' as valid regional units (for its relevance in socially and politically managing the pandemic). These formats should be respected by any data source integrated into COnVIDa for homogeneity. However, despite the above, some points should be considered when using COnVIDa and mixing different types of data sources, as summarized in Table 3.

With respect to the temporal integration across data sources, any temporal data source can be selected, cross-compared and downloaded, per time units, on the temporary panel. However, geographical data sources, such as the INE, cannot be visualized here due to the lack of time series.

Concerning the regional integration, the regional panel does support the whole list of data items. Thus, all the supported metrics are comparable per region unit. Additionally, COVID-19 data can also be compared at the province level. Geographical data sources, such as the INE, are directly drawn by its timeless nature. However, temporal data sources, such as COVID-19, Mobility, MoMo, and AEMET, should be transformed to unique values. With that in mind, COnVIDa gets descriptive statistics from time series such as the mean, median, 25th percentile, 75th percentile, minimum, and maximum. Those values can now be indexed by region, visualized in boxplots and maps, and compared with geographical metrics.

Thus, every time that a temporal visualization is generated to study temporal progressions, another regional perspective, composed of comparative boxplots and a map, is available to characterize regions quickly.

5.3. Representation choice

Regardless of the selected data items being temporal or regional, the graphical representation of data can be changed. First, it can be shown in a line or bar chart, adapting to the user's purpose. Then, it allows changing the scale (linear or logarithmic, being the last one useful when comparing data series of significantly different scales). Any selection made is applied directly to the data, immediately and automatically.

5.4. Download feature

Furthermore, COnVIDa includes the functionality to download the data in a generic format (CSV, XLS, JSON, or HTML) (see Fig. 6). To this extent, it has to be stated that it is possible to download both the raw data and the summary table of the selected data.

5.5. Sharing capability

Finally, another essential functionality of the platform is the possibility of sharing results through a unique link. Thus, COnVIDa allows users to generate a URL from the data they have selected and copy it to the clipboard. In this way, when another user accesses the page through such a URL, he/she will see the same data and graphics selected previously. Even if the user has accessed from another URL, the functionality would be the same as if she had accessed from the original URL, with the

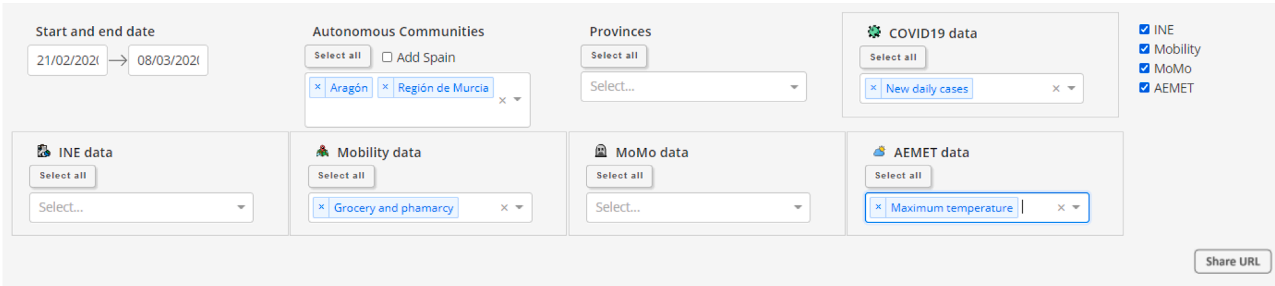


Fig. 3. Data selection for display and analysis.

Table 3
Integrated view of the data sources.

Data source	Temporal integration	Regional
COVID-19	✓	`Statistics`
INE	✗	✓
Mobility	✓	`Statistics`
MoMo	✓	`Statistics`
AEMET	✓	`Statistics`
Cross-comparison	*Per days*	*Per communities*[a]

[a] COVID-19 data items can be also compared per provinces.

possibility of modifying data, changing chart types, or downloading files.

6. COnVIDa library

COnVIDa library is the core Python-based implementation of the end-user requests management and external data sources collection. Particularly, the library offer functions that, given a range of dates, a list of data items, and a list of regions, return the tabular data in a standard format from a local cache or external repositories. The main elements of COnVIDa library are the following:

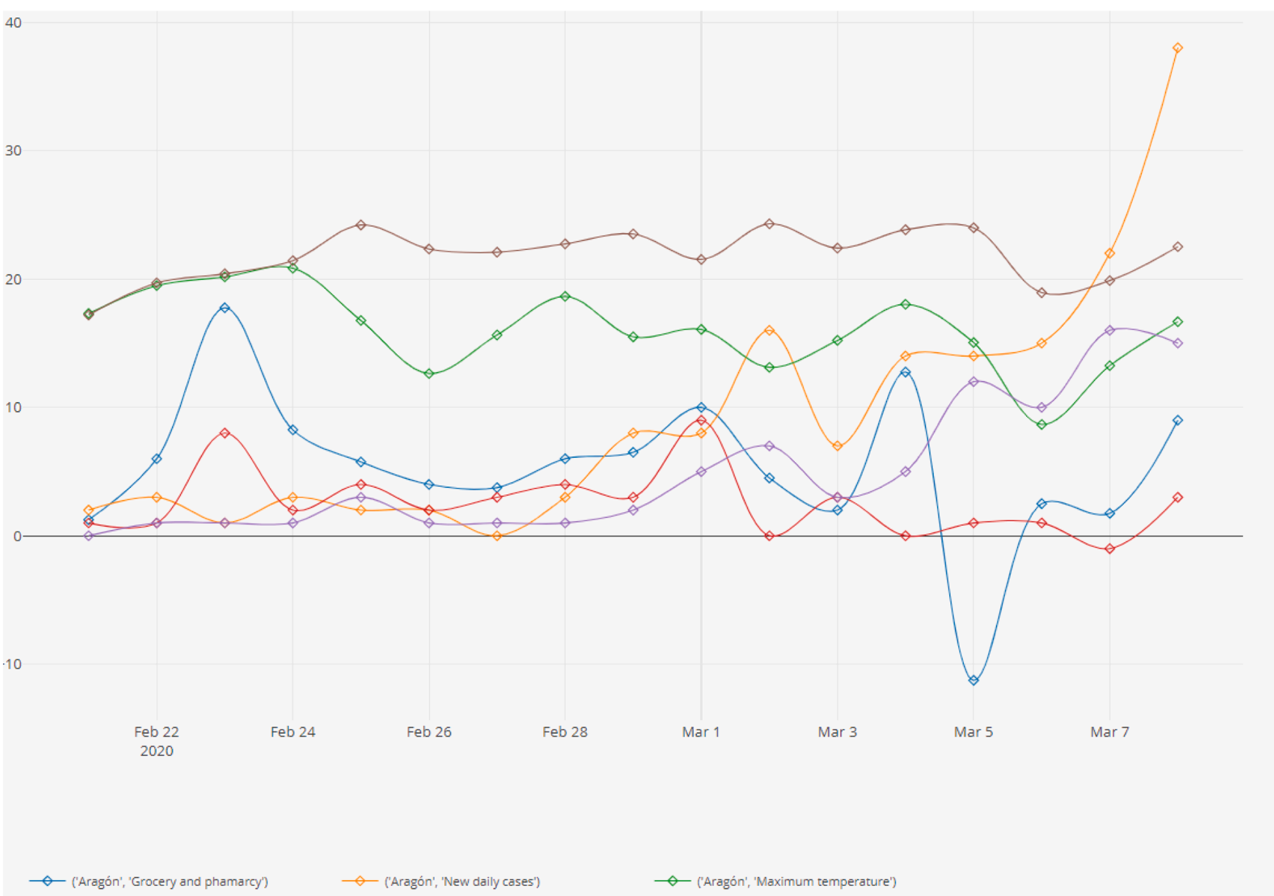
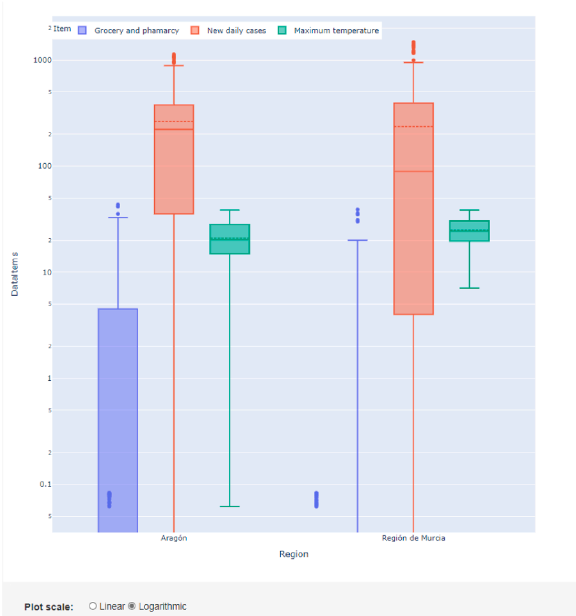
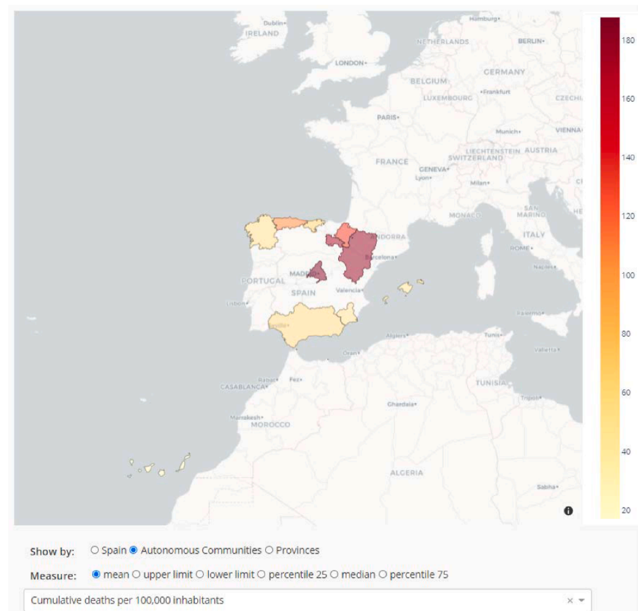


Fig. 4. Display of temporal visualization (linear scale).



(a) Regional visualization (boxplots chart)



(b) Regional visualization (geographical map)

Summary table

Region	Item	count	mean	std	min	25%	50%	75%	max
Andalucía	Cumulative cases per 100,00…	77	2841.49	464.69	1763.39	2579.34	2931.16	3135.33	3692.61
Andalucía	Cumulative deaths per 100,0…	77	50.88	10.6	30.56	42.18	53.02	60.25	65.56
Aragón	Cumulative cases per 100,00…	77	5679.94	518.79	4479.6	5380.09	5715.8	6004.29	6716.18
Aragón	Cumulative deaths per 100,0…	77	181.26	23.18	130.07	166.6	186.24	200.11	212.24
Canarias	Cumulative cases per 100,00…	77	1089.15	186.53	817.46	929.09	1054.94	1235.45	1461.84
Canarias	Cumulative deaths per 100,0…	77	17.11	2.52	13.47	14.91	16.35	19.27	21.78
Cantabria	Cumulative cases per 100,00…	77	2922.98	457.7	1923.15	2639.58	3011.47	3242.08	3679.37
Cantabria	Cumulative deaths per 100,0…	77	58.56	8.99	43.37	50.77	59.72	66.94	71.25
Ceuta	Cumulative cases per 100,00…	78	2862.64	330.78	1942.74	2766.08	2908.22	3000.22	3471.46
Ceuta	Cumulative deaths per 100,0…	78	62.83	11.85	29.49	57.8	67.24	70.77	76.67

(c) Summary table of selected data

Fig. 5. Display of regional data.

- *COnVIDa-server*. Server side of COnVIDa service, which implements the logic to process the dashboard/API requests and query the Data Cache to ultimately return the results. Moreover, it is programmed to update the Data Cache daily. This element is in-depth described in Section 6.1.
- *Data Cache*. Structure that stores the data locally for the dashboard and REST API. This structure permits the dispatch of dashboard/API requests locally on our server, without the need of externally downloading the requested data every time. It is explained in Section 6.2.
- *COnVIDa-lib*. Core implementation of the crawling functionality. It offers functions to retrieve the data supported by the tool. Each request using this library entails the downloading of data. This submodule can be augmented with new data sources without altering the functioning of the rest of the elements, as suggested in Section 6.3.

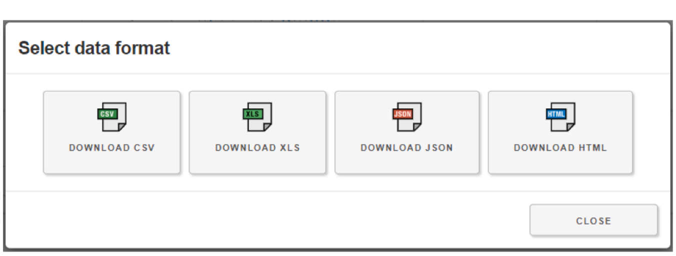


Fig. 6. Types of data export.

6.1. *COnVIDa-server*

COnVIDa-server[13], whose class diagram is presented in Fig. 7, is in charge of dispatching the queries of the users by requesting the Data Cache. With this aim, the function *get_data_items()* receives as parameters the list of data item names, the list of region names, the start date, and the end date selected in the dashboard or requested through the REST API, and returns the requested data from the Data Cache.

Additionally, this submodule is in charge of creating the Data Cache, which should be updated with some frequency. To this end, COnVIDa-server also offers the *daily_update* function, which is responsible for checking if any data source should be updated or not and refresh it accordingly. COnVIDa-server uses COnVIDa-lib (precisely, its *get_data_items()* procedure) to collect new data from their respective official sources.

This submodule does not have to be modified if new data sources are programmed and included in the system.

6.2. *Data Cache*

The Data Cache is a persistent structure that contains the information of all the data sources and the corresponding data items within the range in which these data series exist to date. In particular, it is an HDF5 binary data file (.h5) containing two main Pandas DataFrames[14] for both

[13] https://github.com/CyberDataLab/COnVIDa-lib/tree/master/server.

[14] https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.html.

TEMPORAL and GEOGRAPHICAL data.

When setting up COnVIDaserver for the first time, a Data Cache should be generated using a data generation notebook devoted explicitly to this purpose. The name of the Data Cache follows the format `cache_YYYY-MM-DD.h5`, which specifically indicates the last update. In addition, it should be placed within the data folder.

Fig. 2 shows how the dashboard and the REST API communicate with the Data Cache through the COnVIDa-server when managing end-user requests. On the other hand, COnVIDa-server uses COnVIDa-lib to refresh such Data Cache through an automatic regular process.

The regular update of COnVIDa ensures that data is always up to date. With that objective, it daily checks the update frequency and the timestamp of the last update of each data source and, if required, it accordingly collects the necessary data series. It is worth noting that the temporal granularity of the time series should not necessarily match with the refresh time in the availability of the data in the original repositories.

6.3. COnVIDa-lib

COnVIDa-lib[15] is responsible for collecting the data from external repositories. As mentioned above, the regular update of the data is performed through the functions of this library.

An appealing feature of this package is that it can be used in isolation without deploying the rest of the web server with the associated architecture. In this sense, the classes and functions implemented can be used by any programmer to launch gathering processes in their applications.

Moreover, COnVIDa-lib constitutes an object-oriented package ready to be extended. Considering the principal elements and terminology, implementing a new data source (and associated data items) should be simple. Fig. 8 shows the relationships to be considered when implementing a new data source in COnVIDa-lib.

The unique indispensable requirement for a new data source to be directly integrated into COnVIDa is to be compliant with the supported regions and time units. The following points detail the implementation process to develop a new data source:

1. First of all, some elements should be defined regarding a new Data Source. The following attributes act as a common interface for all data sources to enable a direct integration into the platform and to ensure compatibility between them:
   - Name of the Data Source
   - Data Type of the Data Source (TEMPORAL, GEOGRAPHICAL)
   - Temporal granularity (DAILY)
   - Regional granularity (COMMUNITY, PROVINCE)
   - Representations of the regions within the Data Source (iso 3166 2, ine code, …)
   - Data Format of the resource (JSON, CSV)
   - Update frequency (in days)
   - Information of each Data Item of the Data Source
     - Name (literally used by the Data Source)
     - Display Name (used to change the third-party nomenclature to the desired custom one)
     - Description (meaning of the Data Item)
     - Data unit (metric of the Data Item values: kg, persons, etc.)
2. Configure the aforementioned principal elements of the new Data Source:
   - The name, data type, temporal granularity, regional granularity, region representation, data format, and update frequency should be included in the configuration file of the data sources. With this aim, append a new entry in the JSON object with the data source name as a key, and a dictionary with the corresponding information regarding DATA TYPE, TEMPORAL GRANULARITY, REGIONAL GRANULARITY, REGION REPRESENTATION, DATA FORMAT, and UPDATE FREQUENCY as values. If needed, specific configuration elements of the Data Source can also be included here (for example, AEMET data source defines its API KEY necessary for it to work).
   - For each region, the representation used by the Data Source should be appended accordingly in the configuration file of the regions (in case it does not exist yet). Note that the key of the new entries to be added for each region should match with the REGION REPRESENTATION attribute (defined in the configuration file of the data sources).
   - The information of the Data Items offered by the new Data Source should be included in a new configuration file `NewDataSource-config.json` in a specific configuration folder. As in the other configuration files residing in that folder (which may guide you in this procedure), each Data Item should constitute an entry. Each entry is defined by the Data Item name (literally used by the Data Source) as the key, and the properties display name, description, and data unit as the values. The latter should include, in turn, translation in both Spanish and English (or any other language you may define). If needed, specific properties of the new Data Items can also be included here (for example, the Mobility data source includes an attribute to distinguish the resource where each Data Item comes from).
3. Create the Data Source class in a new empty Python file `NewDataSourceName.py` and place it within the folder of the data sources together with the rest.
4. Develop a new Data Source class as follows. Notice that the data sources folder contains other implemented Data Sources that can be of help.
   - Extend the parent data source class.
   - Declare to None the class attributes (In the first execution of the class, these class attributes will load the values from the configuration files.)
   - Define and fulfill the functions of the class. Specifically, the function which processes partial data should apply the necessary transformations to return data compliant with standard temporal and regional granularity.
   - If extra configuration elements were needed for the Data Source, they should be read.

7. COnVIDa REST API

REST (i.e., Representational State Transfer) is an application of a web-based communication architecture model. REST applications are often used for the development of web or mobile-based services. The REST application uses HTTP as the standard protocol for communicating data from the COnVIDa library to third-party applications.

COnVIDa REST API implements two different routes for obtaining data, as reported in Table 4. The first one is used to recover information from temporal data where, as previously mentioned, it is necessary to indicate a range of dates. Instead, the second is employed to recover regional data and, more precisely, extract information from the INE data source.

To better understand the functioning of this module, an example is presented next. COnVIDa REST API is requested for "Cumulative incidence in the last 14 days" data in two regions: Murcia and Badajoz. To provide such data, a POST request is made to the URL https://convida.inf.um.es/api/temporal. The parameter *data* is entered with the value "`Cumulative incidence in the last 14 days`", ***regions*** with the values "`Murcia`" and "`Badajoz`", ***start_date*** with the date "`2021-01-01`", and ***end_date*** with "`2021-01-08`". Once the request has been made, the API responds with an accepted 200 and, in turn, provides requested data for each region and day of the chosen range (see Fig. 9).

The request made by the user goes directly to the REST service without going through the dashboard (see Fig. 2). This offers a faster

[15] https://github.com/CyberDataLab/COnVIDa-lib/tree/master/lib.

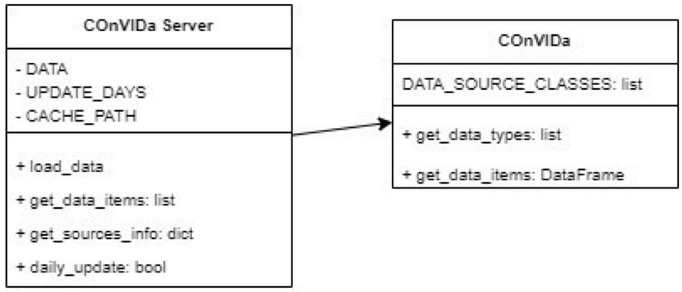


Fig. 7. COnVIDa-server class diagram.

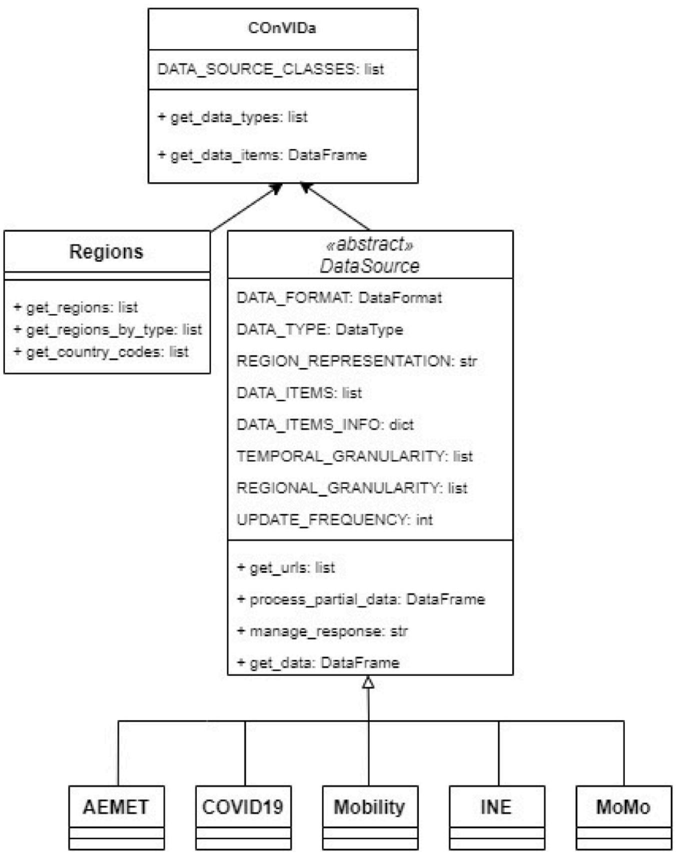


Fig. 8. COnVIDa-lib class diagram.

response since no graphs or tables have to be generated. Moreover, this service fosters the possibility of efficiently implementing third-party applications leveraging the modularity of COnVIDa library.

8. Discussion

In the timely battle against the spread of COVID-19, COnVIDa has been designed and developed to gather, manage, and visualize the pandemic data. Throughout the previous sections, COnVIDa has been deeply analyzed, focusing on its most relevant features and modular architecture. In this Section, an in-depth discussion is presented, explicitly arguing the utility of the proposed platform and the lessons learned through its design and development.

8.1. User profiles and use cases

As previously mentioned, COnVIDa is designed as an open and easy-to-use platform to reach several user categories interested in getting informed about the COVID-19 data and its implications through reliable data sources. To this extent, we foresaw three main user profiles that would benefit from the interaction with the proposed platform at different scales, namely, common citizens, public authorities and private managers, and researchers.

Firstly, any **citizen** interested in knowing the evolution of the pandemic and its relationship with other variables may benefit from the COnVIDa usage. By acquiring reliable information from trusted sources, they are not only getting informed but also battling against the spread of disinformation and fake news about the COVID-19 evolution [9]. As COnVIDa features powerful graphical capabilities and simplicity to draw the pandemic data in precise plots, we believe that these particular users do not need to possess the technical knowledge to interact with the developed platform. For instance, let us assume that a user is concerned about the evolution over time of a certain COVID-19 variable (e.g., the cumulative incidence in the last 14 days) in two specific Spanish Regions (e.g., Region of Murcia and Aragón). In this case, COnVIDa illustrates the trend of the selected variable within the specified time frame in the temporal visualization panel. Simultaneously, several statistical parameters of the data item are reported in the regional visualization window using boxplot graphs in combination with a supportive geographic representation.

Then, **public authorities** and **private managers**, who have to make far-reaching COVID-19 affected decisions in the coming months that influence millions of citizens, may explore COnVIDa functionalities. To this extent, the proposed platform allows them to contextualize the impact of previous decisions in different areas (e.g., health, socioeconomic, etc.). Based on objective data collected from official and trusted sources, they possess the right means to enforce better decisions. This

Table 4
COnVIDa REST API.

Path	Method	Parameters
`/api`	GET	• General information about using COnVIDa REST API.
`/api/temporal`	POST	• data: List of data items (e.g. `[Daily Cases]`). • regions: List of regions (e.g. `[Murcia, Andalucía]`). • start_date: First day in the range, yyyy-mm-dd format (e.g. 2020–01-01). • end_date: Last day in the range (e.g. 2020–05-13).
`/api/regional`	POST	• data: List of data items (e.g. `[Physical activity]`). • regions: List of regions (e.g. `[Murcia, Andalucía]`).

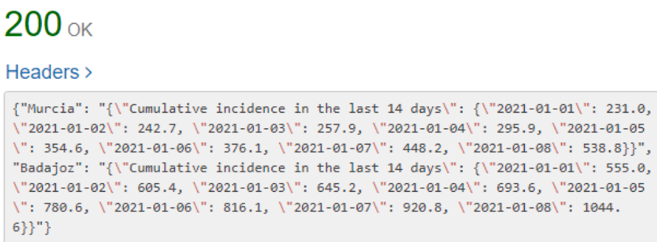


Fig. 9. COnVIDa REST API response.

comparison and contrast of variables is possible thanks to the selection and filtering options of COnVIDa. In this case, some scientific skills are required to formulate specific hypotheses or to conduct rigorous interpretations of data trends. For the sake of simplicity, let us give an example of usage for this profile. Suppose that a mayor is experiencing an acute increase of COVID-19 diagnostic cases within his governed city. In such circumstances, the mayor is willing to compare the tendency of the COVID-19 spreading with the citizens' mobility to decide potential circulation restrictions [22]. Thus, by clicking on the "Mobility" button on the web page, COnVIDa shows the possible mobility data to depict in combination with the selected COVID-19 data for a particular geographic area in the temporal visualization panel. Besides, the regional visualization window depicts the boxplots of the chosen variables within the time frame together with a helpful map.

Last but not least, those **researchers** that are interested in studying the COVID-19 pandemic expansion in Spain from different perspectives, as well as **journalists** who wish to report on its effects, may leverage COnVIDa to reach their goals. In this sense, COnVIDa facilitates the collection of data related to the pandemic from different sources for subsequent downloading, analyzing and comparing them graphically in a user-friendly way. Specifically, the REST API or the manual downloading button of COnVIDa are available for those willing to carry out sophisticated analysis from raw data outside of the platform. To this extent, few will oppose that mainly experts in the field will be able to mine those data series through causal inference or epidemiological techniques in the search for potential hypotheses validations. Also, in this case, let us clarify the concept with an example. Consider a researcher interested in finding potential correlations between COVID-related and meteorological data (e.g., temperature, rainfall, pressure, etc.) in several Spanish areas [23]. In such a case, the researcher can select the regions of preference and subsequently choose the desired COVID-19 and AEMET items (after enabling the latter data source). For the other examples, COnVIDa portrays the trend of the selected variables in the temporal visualization panel and their statistics in the regional visualization window together with the geographical representation. Nevertheless, the expert user can directly download the raw data to perform subsequent specialized analysis in the most suitable format (e.g., CSV). In addition, since the platform offers the possibility of requesting the data of interest from the REST API, the skilled user can code an *ad hoc* program to import the data straight within his workspace.

8.2. Usage statistics

COnVIDa was launched in May 2020 and has received much attention in Spanish mass media[16] and specialized media[17,18] ever since. Moreover, the interest in COnVIDa has spread to other countries, such as

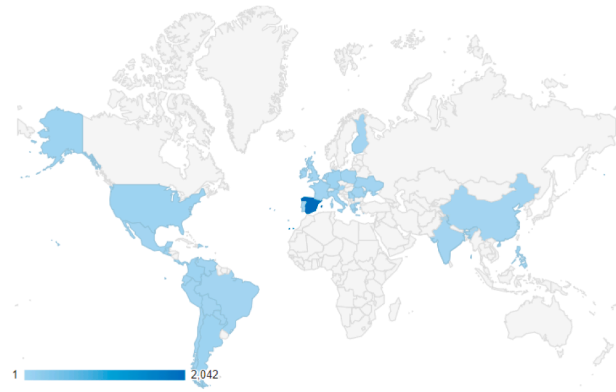


Fig. 10. COnVIDa usage map.

the USA, Colombia, Germany, Argentina, and others, as shown in Fig. 10.

Furthermore, COnVIDa has achieved more than 2,000 active users worldwide, with a particular emphasis on Spain. Thanks to its design, users with both mobile and computer accessed the COnVIDa website, according to Google Analytics statistics.

In turn, Fig. 11 shows the preferred query selectors for COnVIDa users. Based on the queries realized, in total 2800, we can see that positive cases, PCR cases, deaths, ICU patients, and hospitalized patients are the most popular selectors. Other selectors related to Mobility or Weather are also used but to a lesser extent.

In Fig. 12, it can be seen that the most consulted regions through the website are Murcia, Andalucía, Madrid, and Navarra, surpassed only by the consultation of data for the whole country of Spain. Similarly, a graph has been included on the most consulted provinces (Fig. 13), and it can be seen that Murcia and Madrid are still at the top of the list.

8.3. Lessons learned

During the pandemic period, COnVIDa has been designed and developed to support the community with reliable and diversified data from trusted sources as a primary goal. One could easily say that informing interested users about the evolution of several variables during such a delicate period is essential. Nonetheless, the tasks accomplished to achieve such an ambitious objective are not trivial.

Firstly, the recollection of data from several sources has been challenging during the COnVIDa life cycle. Indeed, diverse data sources provide data scarcely or incompletely, including those that are related to governments. Additionally, the presented data often need a preprocessing phase to be acquired and, consequently, managed by the proposed platform. In this sense, it is worth remarking on the importance of furnishing reliable, open, frequently-updated, and machine-readable information that can be collected in a quick and programmatic fashion. On top of that, this fact becomes even more crucial when it comes to handling COVID-19 data. In particular, very few will oppose that providing trustworthy health-nature data is fundamental to correctly inform interested citizens and to fight against the spread of disinformation and fake news within various communities.

Furthermore, the development of the modular COnVIDa architecture and functions has required extensive efforts. Once the challenge mentioned above of identifying valuable data sources has been accomplished, the platform needs to acquire such data with a certain frequency and manage them. To achieve this, COnVIDa architecture contains two core modules, namely, *COnVIDa-server* and *COnVIDa-lib*, as detailed in Section 6. In particular, COnVIDa-lib has been designed as a class hierarchy to form an object-oriented package ready to be extended. In fact, more expert users with programming skills can reuse the functions and classes implemented in the library within their workspaces. Moreover, the entire COnVIDa framework has been secured through a WAF (Web

[16] https://www.europapress.es/murcia/noticia-investigadores-umu-lanzan-convida-plataforma-recopila-todos-datos-covid-19-20200604101726.html.

[17] https://www.infosalus.com/asistencia/noticia-coronavirus-investigadores-lanzan-convida-plataforma-recopila-todos-datos-covid-19-20200604104639.html.

[18] https://www.agenciasinc.es/Noticias/Una-nueva-plataforma-web-recopila-todos-los-datos-oficiales-de-la-COVID-19.

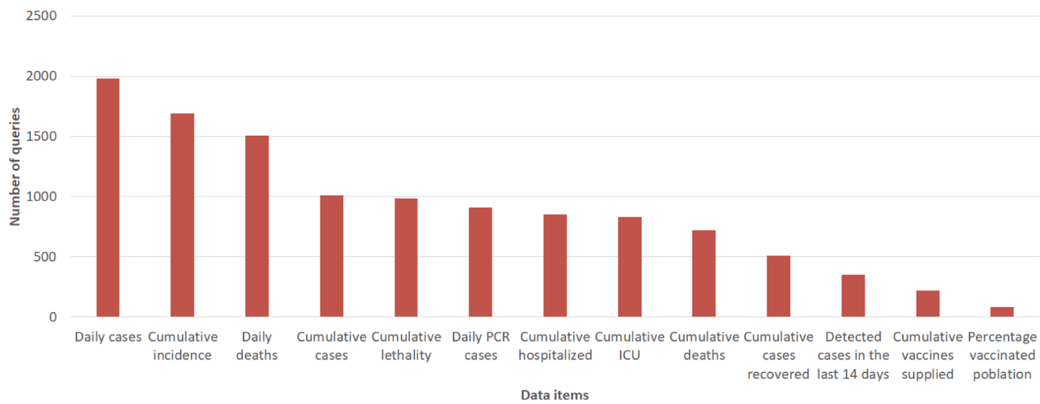


Fig. 11. Main query selectors for the last 2,800 queries.

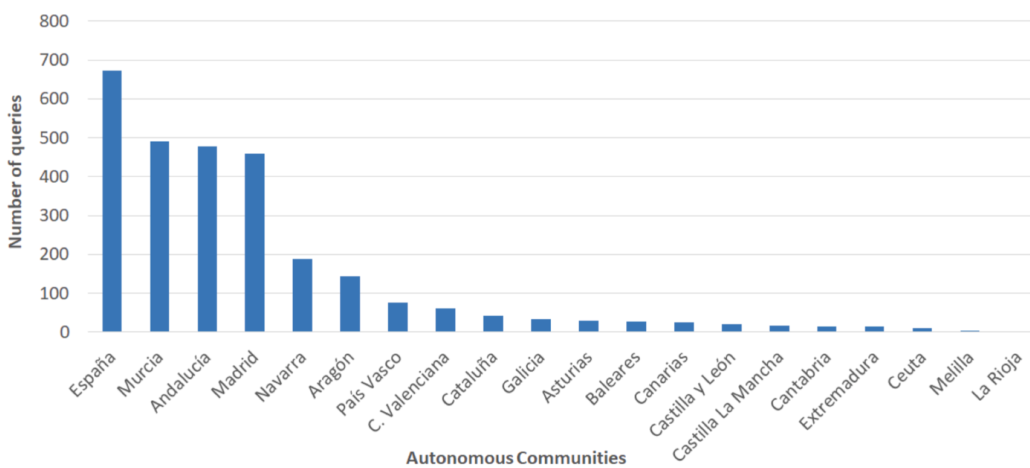


Fig. 12. Queried Autonomous Communities in the last 2,800 queries.

Application Firewall). Even if this choice directly implies an additional effort to the overall process, it is highly recommended due to the exposure of the platform on the Web, which was targeted by a vast quantity of malicious traffic after its first public release.

Besides, COnVIDa has been proposed to reach user categories with different expertise and intentions. Thus, it is designed to provide a great user experience regardless of user purposes and skills. To this extent, COnVIDa passes through periodical usage tests run by volunteer *beta testers*. Notably, the platform has been submitted to the evaluation of various users with different profiles (i.e., normal users, managers, researchers, journalists), aiming to receive meaningful feedback on improving the dashboard and its functionalities. Thanks to their work, COnVIDa has been enhanced iteratively and is ready to support diverse users.

9. Conclusions and future work

This paper has described COnVIDa, a web-based tool for compiling and displaying multidisciplinary data about the current COVID-19 pandemic. The user can select the range of dates, the regions, and

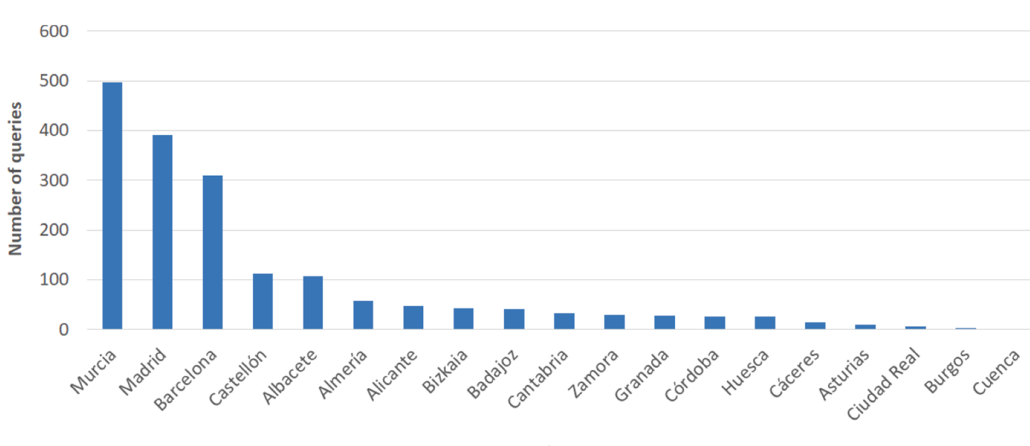


Fig. 13. Queried provinces in the last 2,800 queries.

data items to inspect, and the framework smoothly plots the associated graphics. At the time of writing, the platform can collect clinical/pandemic, social, mobility, excess mortality, and weather information from official Spanish sources. Moreover, thanks to the adoption of the open science philosophy, our solution is technically designed to be easily extended with additional data sources and geographical regions. In this sense, we have found very few projects whose functionality can be directly extended by third-party contributions.

Having analyzed other similar platforms, we can conclude that COnVIDa is the most powerful tool in terms of flexibility when filtering specific scenarios and cross-comparing any set of variables, hence enabling scientists or decision-makers to consult cross-cutting relationships and, ultimately, expose correlations in specific situations. Mass media have broadly published about such novel dashboard, and the usage statistics confirm that it is proving particularly useful in exploring places where the virus is most prevalent.

As for future directions, authors are preparing different releases that aim to increase the functionalities and scope of the website. In this sense, the main milestones to come ahead are: i) to allow further types of granularity when filtering places, such as countries, health areas, or even municipalities, or dates, such as for a week or month level monitoring; ii) to add data items related to the use of public transport, economy, employment, and Google search trends; iii) to support user authentication, manage users' queries records, logs and statistics, and even permit those users the proposal of new data series or specific correlations; and iv) to introduce basic analytical capabilities by incorporating options for applying machine learning algorithms, making causal inference from observational data, running automatic correlation studies, and suggesting prediction models.

CRediT authorship contribution statement

Enrique Tomás Martínez Beltrán: Software, Validation, Writing - original draft. **Mario Quiles Pérez:** Software, Writing - original draft. **Javier Pastor-Galindo:** Methodology, Software, Validation, Investigation, Writing - original draft. **Pantaleone Nespoli:** Methodology, Investigation, Writing - original draft. **Félix Jesús García Clemente:** Validation, Resources, Writing - review & editing. **Félix Gómez Mármol:** Conceptualization, Methodology, Software, Validation, Resources, Writing - review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was partially funded by grants from the Spanish Government with codes FPU18/00304 and RYC-2015-18210, co-funded by the European Social Fund, and by a predoctoral grant from the University of Murcia.

References

[1] Q.A. Ahmed, Z.A. Memish, The cancellation of mass gatherings (MGs)? decision making in the time of covid-19, Travel Med. Infect. Dis. 34 (2020) 101631, https://doi.org/10.1016/j.tmaid.2020.101631.
[2] P. Deb, D. Furceri, J.D. Ostry, N. Tawk, The Economic Effects of Covid-19 Containment Measures, CEPR Discussion Paper No. DP15087.
[3] S. Yezli, A. Khan, Covid-19 social distancing in the kingdom of saudi arabia: Bold measures in the face of political, economic, social and religious challenges, Travel Med. Infect. Dis. 37 (2020) 101692, https://doi.org/10.1016/j.tmaid.2020.101692.
[4] P. Luo, Y. Liu, L. Qiu, X. Liu, D. Liu, J. Li, Tocilizumab treatment in covid-19: A single center experience, J. Med. Virol. 92 (7) (2020) 814–818, https://doi.org/10.1002/jmv.25801.
[5] M. Nicola, Z. Alsafi, C. Sohrabi, A. Kerwan, A. Al-Jabir, C. Iosifidis, M. Agha, R. Agha, The socio-economic implications of the coronavirus pandemic (covid-19): A review, Int. J. Surg. 78 (2020) 185–193, https://doi.org/10.1016/j.ijsu.2020.04.018.
[6] C. Anastassopoulou, L. Russo, A. Tsakris, C. Siettos, Data-based analysis, modelling and forecasting of the COVID-19 outbreak, PLOS ONE 15 (3) (2020) e0230405, https://doi.org/10.1371/journal.pone.0230405.
[7] J.H. Stock, Data gaps and the policy response to the novel coronavirus, Working Paper 26902, National Bureau of Economic Research (March 2020). doi:10.3386/w26902.
[8] T.A. Glass, S.N. Goodman, M.A. Hernán, J.M. Samet, Causal inference in public health, Annu. Rev. Public Health 34 (1) (2013) 61–75, https://doi.org/10.1146/annurev-publhealth-031811-124606.
[9] E. Chen, K. Lerman, E. Ferrara, Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set, JMIR Public Health Surveill 6 (2) (2020), https://doi.org/10.2196/19273.
[10] M.J. Hollaway, G. Dean, G.S. Blair, M. Brown, P.A. Henrys, J. Watkins, Tackling the challenges of 21st-century open science and beyond: A data science lab approach, Patterns 1 (7) (2020) 100103, https://doi.org/10.1016/j.patter.2020.100103.
[11] J. Pastor-Galindo, P. Nespoli, F. Gómez Mármol, G. Martínez Pérez, The not yet exploited goldmine of OSINT: Opportunities, open challenges and future trends, IEEE Access 8 (2020) 10282–10304, https://doi.org/10.1109/ACCESS.2020.2965257.
[12] L.D.L. Torre, L.T. Neustock, G.K. Herring, J. Chacon, F.J.G. Clemente, L. Hesselink, Automatic generation and easy deployment of digitized laboratories, IEEE Trans. Ind. Informat. 16 (12) (2020) 7328–7337, https://doi.org/10.1109/TII.2020.2977113.
[13] V.D. Daniel Shu Wei Ting, Lawrence Carin, T.Y. Wong, Digital technology and COVID-19, Nature Med. 26 (2020) 459–461, https://doi.org/10.1038/s41591-020-0824-5.
[14] J.M. Rohrer, Thinking clearly about correlations and causation: Graphical causal models for observational data, Adv. Methods Practices Psychol. Sci. 1 (1) (2018) 27–42, https://doi.org/10.1177/2515245917745629.
[15] O. Morgan, How decision makers can use quantitative approaches to guide outbreak responses, Philosoph. Trans. Roy. Soc. B: Biolog. Sci. 374 (1776), https://doi.org/10.1098/rstb.2018.0365.
[16] N. Stefan, A.L. Birkenfeld, M.B. Schulze, D.S. Ludwig, Obesity and impaired metabolic health in patients with COVID-19, Nature Rev. Endocrinol. 16 (7) (2020) 341–342, https://doi.org/10.1038/s41574-020-0364-6.
[17] C.O. Buckee, S. Balsari, J. Chan, M. Crosas, F. Dominici, U. Gasser, Y.H. Grad, B. Grenfell, M.E. Halloran, M.U. Kraemer, M. Lipsitch, C.J.E. Metcalf, L.A. Meyers, T.A. Perkins, M. Santillana, S.V. Scarpino, C. Viboud, A. Wesolowski, A. Schroeder, Aggregated mobility data could help fight COVID-19, Science 368 (6487) (2020) 145–146, https://doi.org/10.1126/science.abb8021.
[18] L.S. Vestergaard, J. Nielsen, L. Richter, D. Schmid, N. Bustos, T. Braeye, G. Denissov, T. Veideman, O. Luomala, T. Möttönen, A. Fouillet, C. Caserio-Schönemann, M. An der Heiden, H. Uphoff, T. Lytras, K. Gkolfinopoulou, A. Paldy, L. Domegan, J. O'Donnell, F. De' Donato, F. Noccioli, P. Hoffmann, T. Velez, K. England, L. van Asten, R.A. White, R. Tønnessen, S.P. da Silva, A.P. Rodrigues, A. Larrauri, C. Delgado-Sanz, A. Farah, I. Galanis, C. Junker, D. Perisa, M. Sinnathamby, N. Andrews, M. O'Doherty, D.F. Marquess, S. Kennedy, S.J. Olsen, R. Pebody, ECDC Public Health Emergency Team for COVID-19, T.G. Krause, K. Mølbak, Excess all-cause mortality during the COVID-19 pandemic in Europe - preliminary pooled estimates from the EuroMOMO network, March to April 2020, Euro surveillance: bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin 25 (26). doi:10.2807/1560-7917.ES.2020.25.26.2001214.
[19] C.J. Carlson, A.C. Gomez, S. Bansal, S.J. Ryan, Misconceptions about weather and seasonality must not misguide COVID-19 response, Nature Commun. 11 (1) (2020) 2–5, https://doi.org/10.1038/s41467-020-18150-z.
[20] M. Allam, S. Cai, S. Ganesh, M. Venkatesan, S. Doodhwala, Z. Song, T. Hu, A. Kumar, J. Heit, C.-. Study Group, A.F. Coskun, Covid-19 diagnostics, tools, and prevention, Diagnostics 10(6) (2020). doi:10.3390/diagnostics10060409.
[21] C. Pahl, Containerization and the paas cloud, IEEE Cloud Comput. 2 (3) (2015) 24–31, https://doi.org/10.1109/MCC.2015.51.
[22] H.S. Badr, H. Du, M. Marshall, E. Dong, M.M. Squire, L.M. Gardner, Association between mobility patterns and covid-19 transmission in the usa: a mathematical modelling study, Lancet. Infect. Dis. 20 (11) (2020) 1247–1254, https://doi.org/10.1016/S1473-3099(20)30553-3.
[23] B.F. Zaitchik, N. Sweijd, J. Shumake-Guillemot, A. Morse, C. Gordon, A. Marty, J. Trtanj, J. Luterbacher, J. Botai, S. Behera, Y. Lu, J. Olwoch, K. Takahashi, J. D. Stowell, X. Rodó, A framework for research linking weather, climate and COVID-19, Nature Commu. 11 (1) (2020) 5730, https://doi.org/10.1038/s41467-020-19546-7.