esade

EsadeEcPol - Center
for Economic Policy

Open Internet Governance Institute

PAPER SERIES # 3

Data Science for Social Good: The Valencian Example during the COVID-19 pandemic

Nuria Oliver, PhD

Co-founder and Director of the ELLIS Alicante Unit Foundation,
aka The Institute of Humanity-centric AI

JUNE, 2022

About the author

Nuria Oliver, PhD
Co-founder and Director of the ELLIS
Alicante Unit Foundation, aka The Institute of
Humanity-centric AI

Summary

Data Science is a discipline of tremendous value in the public sector for at least two essential reasons:

- → it enables the design of public policies based on data and evidence, as opposed to intuitions, obsolete information, or political interests.
- → it allows the necessary empirical evaluation of the impact of the deployed public policies, to determine their strengths and weaknesses.

Despite its potential, the systematic use of data to support public policy making is still rare in most parts of the developed and developing world. In this article, I provide a summary of an internationally recognized example of leveraging data to support policy-making during the COVID-19 pandemic in Spain. I present the work that we carried out between March of 2020 and March of 2022 in the "Data Science against COVID-19 Taskforce", a pioneering experience in the Valencian Region of Spain. This taskforce --which I led-- was composed of a multi-disciplinary team of 20+ scientists from several universities and research centers in the Valencian Region of Spain, working closely with the region's policymakers at the Presidency level.

We focused on work on four impact areas to the use of Data Science in the fight against the coronavirus pandemic:

(1) modeling aggregate human mobility; which allowed us to

- monitor the impact of containment measures on the real mobility of citizens;
- identify areas where the confinement measures had a greater or lesser impact;
- quantify the success of the #stayathome campaign, as well as the measures adopted to restrict night mobility and perimeter closures of the Valencian region of Spain
- model the impact that reduced mobility had on the progression of the coronavirus
- better predict the spread and determine the usefulness of selective lockdowns
- track the return to normality

(2) developing computational epidemiological models, devoted to to make predictions about the evolution of the pandemic not only under the current conditions, but also under different scenarios of confinement and control measures of the pandemic, pre-existing immunity, vaccination or contact tracing.

(3) developing predictive models at the sub-regional (health department) level to estimate relevant variables for guiding policy decisions.

(4) applying citizen science via a large-scale online survey called the COVID19Impactsurvey, a key tool to to shed light on issues of great public importance for which there is no systematically captured data; e.g. the role of close contacts, a first evaluation of psychological impact, or the difference on the ability to follow lockdowns depending on socio-demographic or socio-economic conditions.

This effort is an inspiring example of close collaboration between the scientific community, citizens, and a public administration to ensure that public policies are effectively based on evidence and expert knowledge.

Introduction

In this article, I summarize the work carried out in the Data Science against COVID-19 Taskforce. This taskforce consisted of a multidisciplinary team of researchers from the Valencian region in Spain, working very closely with the General Directorate of Analysis and Public Policies within the Presidency of the Valencian Government during the COVID-19 pandemic (between March of 2020 and April of 2022).

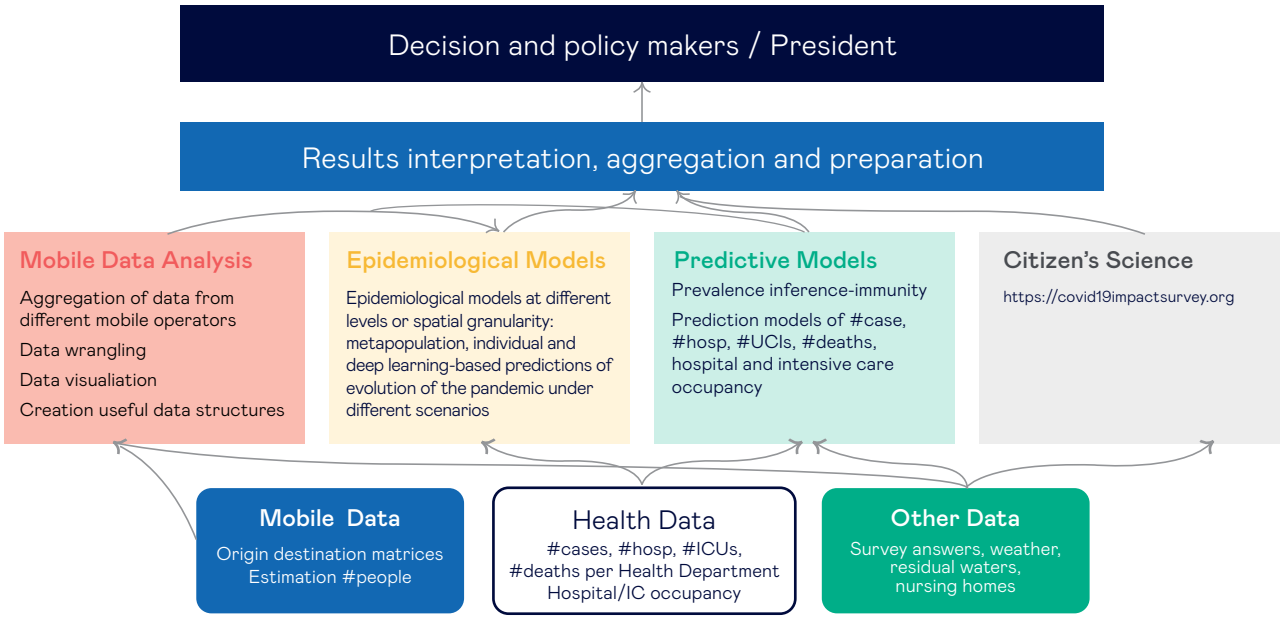
The taskforce was composed of the following scientists: Alberto Conejero, Miguel Rebollo, Manuel Portolés, Víctor de Elena, Miguel Angel García-March, Oscar Garibo and Eloy Piñol from the UPV; Francisco Escolano, Miguel Angel Lozano, Juan Carlos Trujillo and Miguel Angel Teruel from the University of Alicante; Francisco Falcó from the CEU Universidad Cardenal Herrera; Alejandro Rabasa, Aurora Mula, Xavier Barber, Kristina Polotskaya and Elisa Espín from the University Miguel Hernández; Joaquín Huerta, Marina Martínez, Emilio Sansano, Juan Camilo Gómez and Rubén Femenía from the Universitat Jaume I and Adolfo López from FISABIO.

The configuration, way of working and results of this team are unique not only at the national level, but also internationally, as reflected by the visibility obtained in both national [1,2,3] and international [4,5] media.

In addition to having had direct impact to support the policy-making during the COVID-19 pandemic and having made significant scientific contributions, we hope that the experience described in this article will inspire other governments and public administrations, both at a national and international level, to transition towards evidence-based public decision-making and evaluation.

Our work was divided into 4 areas, described below, and represented in Figure 1: (1) aggregate human mobility modeling; (2) computational epidemiological models; (3) predictive models; and (4) citizen science.

Figure 1. Areas of work of the Data Science against COVID-19 taskforce



1.
Aggregate human mobility analysis

The first line of work focused on the analysis of human mobility. Human mobility is key to the spread of infectious diseases. In the literature, there are examples of the use of aggregated and anonymized mobility from the mobile phone network to help fight diseases such as Ebola [6], Zika [7] and malaria [8].

When a human-to-human transmitted infectious disease is in a phase of community transmission, the containment of human mobility is one the most implemented non-pharmacological interventions (NPI) to limit the spread of the disease. In Europe, most countries limited the mobility of their population, to a greater or lesser degree, during the first weeks of the COVID-19 pandemic (from mid-March to May of 2020) and subsequently in response to significant increases of the disease incidence all throughout 2020 - 2022.

In Spain, a state of alarm was declared on March 14th, 2020, to implement these mobility containment measures that started on March 16th and ended when the "new normal" was returned on June 21st, 2020.

Our work of mobility analysis in the context of the first wave of the COVID-19 pandemic (March-June 2020) was a pioneering pilot in the Valencian Community, as announced by Vice President Nadia Calviño on March 23rd, 2020.

Starting during the first wave of SARS-CoV-2 infections, we analyzed aggregated and anonymized data extracted from the mobile phone network and shared by the National Institute of Statistics (INE), thanks to a collaboration agreement between the INE and the three largest mobile operators in Spain (Telefónica, Vodafone and Orange), corresponding to the period from March 16th, 2020 to June 30th, 2020 [9]. The project complies with the anonymization conditions of the General Data Protection Regulation, as described by the INE [10]. The mobility data shared by the INE was later made publicly available on its website[1].

We also analyzed mobility data shared by the Ministry of Transport, Mobility and Urban Agenda (MITMA)[2] and anonymized, aggregate mobility data shared by Facebook and Google.

1 https://www.ine.es/covid/covid_movilidad.htm

2 https://www.mitma.gob.es/ministerio/covid-19/evolucion-movilidad-big-data/opendata-movilidad

A quantitative analysis of large-scale human mobility enables to:

(a) monitor the impact of containment measures on the real mobility of citizens;

(b) identify areas where the confinement measures had a greater or lesser impact and monitor their behavior over time;

(c) quantify the success of the #stayathome campaign, as well as the measures adopted to restrict night mobility and perimeter closures of the Valencian region of Spain;

(d) model the impact that reduced mobility had on the progression of the coronavirus;

(e) identify communities based on population mobility to better predict the spread of the coronavirus and determine the usefulness of possible selective lockdowns, should new post-lockdown outbreaks occur.

(f) quantify the progressive return of pre-pandemic levels of mobility as the confinement measures were lifted.

The volume of information of the mobility data was one of the first challenges that we had to address. To process, visualize and analyze such large-scale data, we used different Big Data technologies (depicted in Figure 2) both in the backend as in the frontend (display).in Figure 2) both in the *backend* as in the *frontend* (display).

Figure 2. Technological solutions used in the mobility analysis work area



Enabling an intuitive visualization of such complex data is of paramount importance to support policy-making. Thus, we developed several visualizations using different tools according to the target audience to which we wanted to present the results. We developed different dashboards that make it easier for different profiles to understand information that, due to its complexity, is difficult to present in any other way, as illustrated in Figure 3.

Figure 3. Visualization dashboard of the mobility analysis



In the following sections we summarize the main conclusions of our analysis of the impact on mobility of different containment measures deployed during the pandemic.

Compliance with the "Stay at Home" campaign during the confinement of the first wave (March-May 2020)

Based on the INE data, we quantified the daily real mobility of citizens since the beginning of the pandemic and compared it with reference mobility data that was also provided by the INE for a "normal" or baseline day before the pandemic began (November 2019).

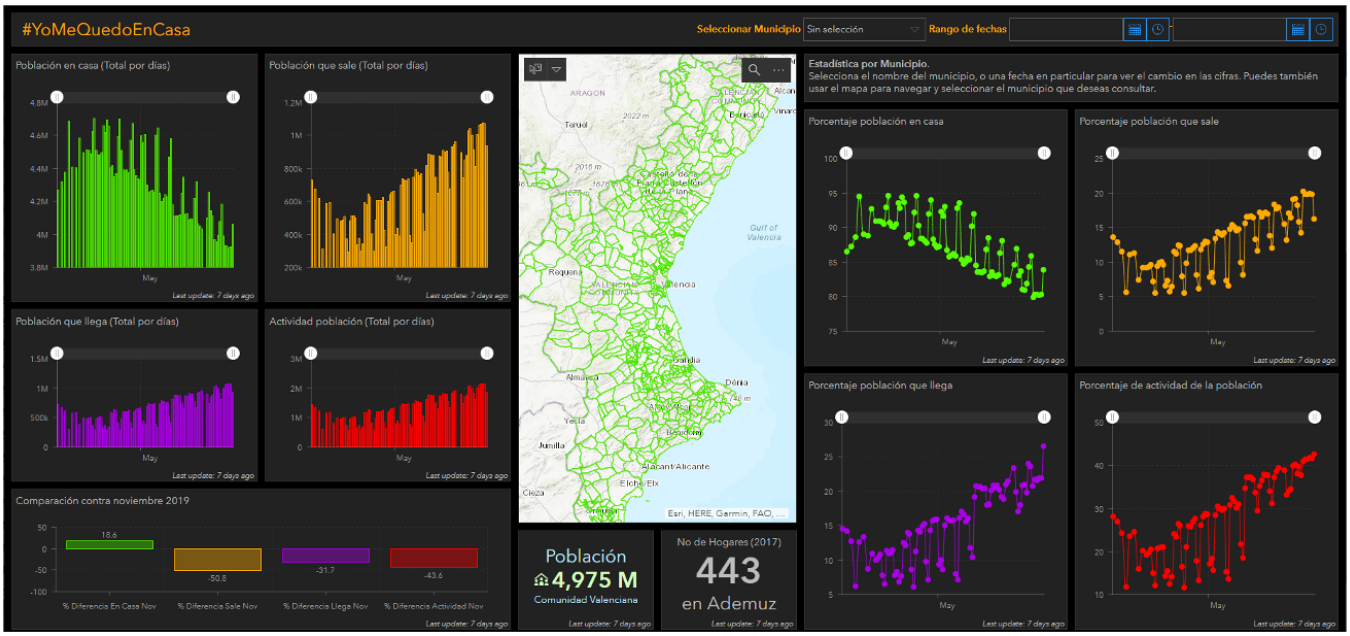
We found high levels of compliance of the “state at home” campaign. In the period between March 16th and April 27th, 2020: on average, 88% (working days) and 92% (weekends and holidays) of the population of the Valencian region remained in their area of residence, which illustrates a high compliance with the confinement measures.

Labor mobility during the confinement of the first wave (March-May 2020)

We also looked at levels of labor mobility compared to a pre-COVID-19 working period in November. On average, in the period from March 16th to April 27th, we observed a 59% reduction in the number of people who spent at least 2 hours outside their area of residence during working hours.

Figure 4 shows a visualization of the data and the mobility analysis that we carried out. A summary of this mobility work can also be found in two reports published by the Generalitat Valenciana in April [11] and May [12] of 2020.

Figure 4. Mobility data viewer (http://t.ly/9riC).



Mobility communities

In addition to the analysis of the flow of movements itself, the mobility data is the input to a graph analysis module. Graphs are structures that are present in numerous natural and artificial phenomena and have been used to model, for example, migratory flows or air traffic. We use such flows to identify patterns in people's mobility. This approach enables us to detect which areas are interrelated, regardless of the administrative boundaries that divide the territory. To do this, we carry out the automatic detection of the communities that emerge from the flows using the algorithm proposed by Newman [13].

A *mobility community* is composed of geographical areas with a high density of internal movements within them, and very few movements with the rest. From the analysis of communities based on the mobility data described above and their evolution during the confinement, we obtained the following conclusions:

→ We identified a clear structure of communities that is stable and does not exactly coincide with the administrative boundaries. From a mobility viewpoint, there is a lot of flow of population movement between the north of Castellón with the south of Tarragona, as well as the bordering towns between Murcia and Alicante. The same phenomenon is observed in the mobility between provinces. For this reason, it would be necessary to coordinate containment policies among the Autonomous Communities in Spain.

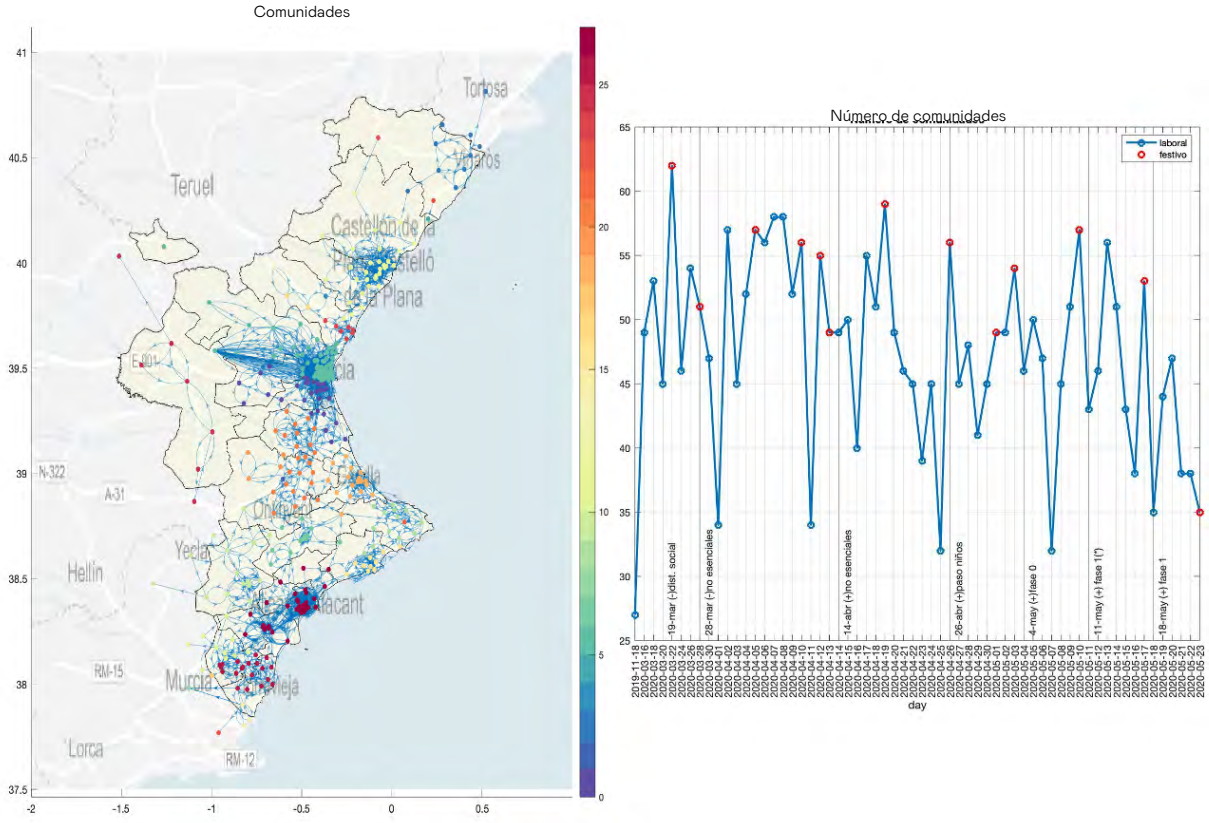
→ The number of communities increased as mobility restrictions expanded, from 27 communities in the reference week of November of 2019 to between 50 and 60 during the lockdown of the first wave of infections (see Figure 3, right). We then observed a gradual decrease in the number of communities, with peaks in holidays comparable to the period of maximum confinement. Throughout the period,

the number of communities increased by at least 50% when compared to the reference week. Having many small, separate communities helped contain the spread of SARS-CoV-2. This same reduction in size was also observed during holidays and weekends.

→ The division into communities is consistent with the division of health zones in the Autonomous Region. Thus, the data supports making decisions based on the health zones.

This analysis of communities based on mobility flows is important to support decision-making regarding selective confinements of certain geographical areas, depending on their epidemiological situation. The more self-contained the mobility of a geographical area, the lower the epidemiological impact of a confinement of that area since most of its mobility is internal, not with origin / destination in other geographical areas.

Figure 5. Division into mobility communities (left) and evolution of their size over time (right)



Return of mobility

As the confinement measures were lifted, the population's levels of mobility increased. During 2021, we analyzed aggregated mobility data shared by Facebook and Google to monitor such an increase in the levels of mobility. We observed a significant increase –to levels even higher than those before the pandemic-- in the activities related to outdoor spaces, recreation and parks, which probably reflects the desire in the population to be outdoors after months of confinement.

2.
Computational epidemiological models

In this area of work, we developed three types of computational epidemiological models: a meta-population SEIR-based model, an individual agent-based model and a model based on deep neural networks.

Epidemiological models allow us to make predictions about the evolution of the pandemic not only under the current conditions, but also under different scenarios of confinement and control measures of the pandemic, pre-existing immunity, vaccination or contact tracing.

Meta-population SEIR model

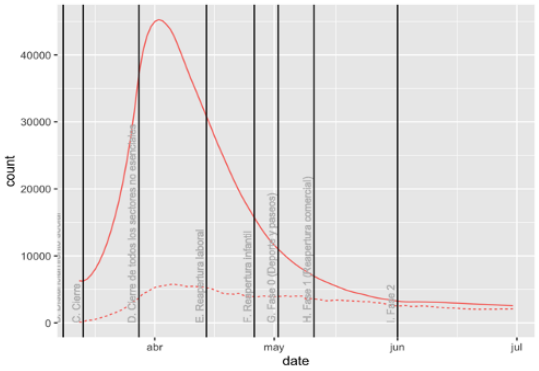
The meta-population model that we developed is based on a classical SEIR model [14], where each letter of the acronym represents a different state that members of the population might be in: S represents the number of individuals susceptible to contracting the disease, E the number of individuals in the population exposed to contracting the disease, I denotes the number of infected individuals who can transmit the disease, and finally R represents the number of individuals who are removed from the system, due to recovery or death. The parameters of the model represent the rates of transfer of individuals between the four different sub-populations in which the population has been divided, so that we can determine how many individuals make up each of the classes at any instant in time.

This deterministic model introduced by Joan L. Aron and Ira B. Schwartz [14] in 1984 is presented as a scientific paradigm of the dynamic behavior of the different groups into which we divide the population. We developed a stochastic extension of this deterministic model with the aim of introducing uncertainty into the original model.

Figure 6 shows the estimates made by the SEIR model when compared to the real data of cases reported in the Valencian Community. Note that a high percentage of COVID-19 cases were asymptomatic or had mild symptoms. This means that, especially in the initial stage of the pandemic when the availability of tests was limited, there was a large number of unreported cases. The model allows us to estimate the underlying number of total infections, and compare it with the number of reported cases, observing how the rate of reported cases has progressively increased from around 10% at the beginning of the pandemic to almost all cases in the month of July (see Figure 6, a) of 2020, once there were enough COVID-19 tests available.

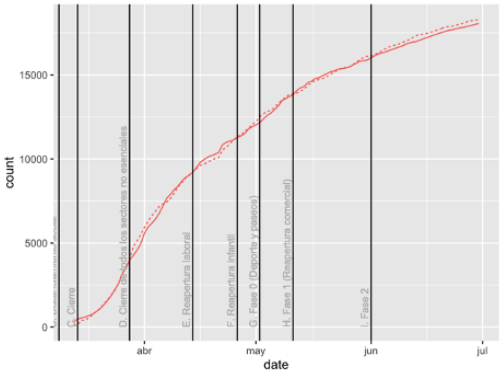
Figure 6.

a



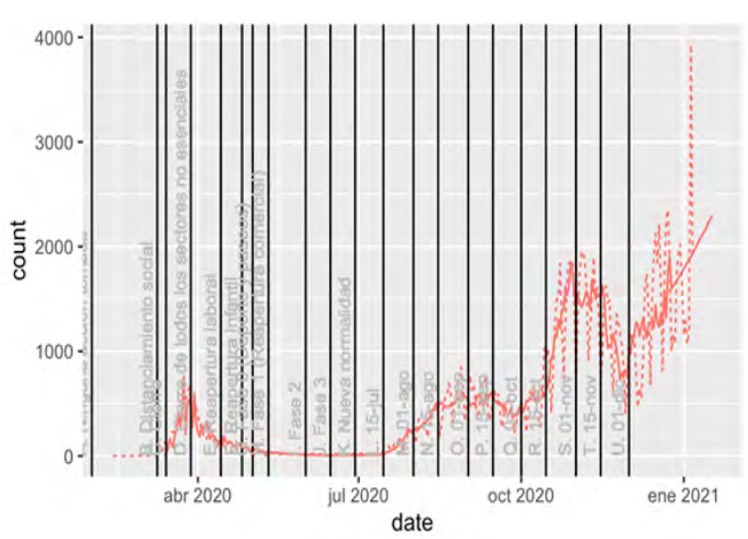
Comparison between the underlying active cases predicted by the model, with the active cases reported during the first wave (March-May 2020). Note the high percentage of underlying cases not reported (due to lack of tests, being asymptomatic, etc ...)

b



Cumulative total cases estimated by the model that would have been reported, and cumulative total cases reported in the actual data series.

c



Predictions of the SEIR model (solid line) and real data observed (dashed line) in the Valencian Community.

Individual agent-based model and the impact of contact tracing

One of the most effective measures to reduce the rate of contagion is contact tracing. The technique is based on an early detection of as many positive cases as possible with the aim of isolating them and thus cutting the chain of transmission. The positive detected cases are isolated while those who have been in close contact with them are identified. Ideally, close contacts at risk of having been infected perform the

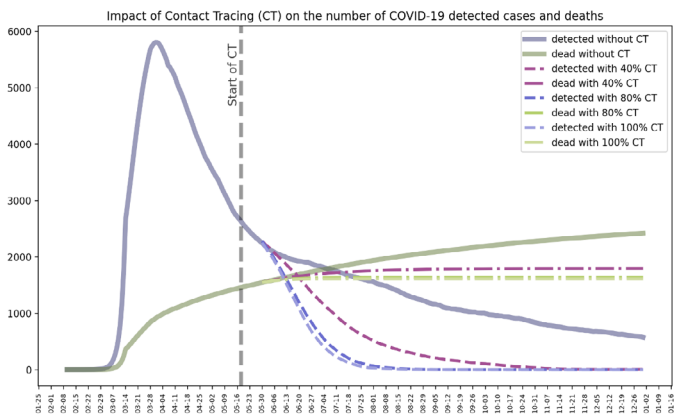
necessary tests to determine if they have been infected and isolate themselves until the results of the tests are known. All close contacts that test positive for a coronavirus infection, are asked to isolate themselves and their close contacts are identified in an attempt to break the chain of transmission as fast as possible.

In our study, we modeled the interaction between the existing capacity to perform effective contact tracing and the resulting epidemiological curve based on that capacity, with the aim of estimating different scenarios in the evolution of the pandemic. We simulated a series of scenarios that contemplate different contexts in terms of the effectiveness of contact tracing from mid-May 2020, when the confinement measures began to decline giving way to a gradual de-escalation.

The data used for the simulations were publicly available or were shared with us by the Valencian Government. The estimates were made using the REINA model [16], which we adapted to the characteristics of the Valencian Community. It is an individual, agent-based epidemiological model that models population dynamics through social interactions between individuals. One of the distinctive features of this type of model is that agents are not homogeneous entities but are created with certain characteristics that represent the behavior of different populations. The simulation is then run through a series of rules and interactions between the agents that try to reflect the characteristics of the population under study. Thus, the spread of the epidemic is not the same for all agents but depends on their characteristics and their interactions with other agents. For example, an elderly agent who becomes infected will be more likely to need medical services than a younger agent. In general, agent-based models require a considerable amount of demographic, sociological, and behavioral information in order to develop realistic interactions. Likewise, they are usually computationally expensive since they require carrying out simulations of the individual behavior of millions of agents (in our case, ~5 million inhabitants of the Valencian region of Spain).

Possible scenarios regarding contact tracing range from the absence of contact tracing to a 100% trace, which would imply that all the contacts of each positive case are detected, and completely isolated for two weeks. As expected, (see Figure 14) this last scenario, although impossible to carry out in the real world, entails a rapid reduction in the number of detected cases. The results of our simulation show that a 40% efficiency in contact tracing considerably reduces the number of cases and consequently the number of hospitalizations and deaths.

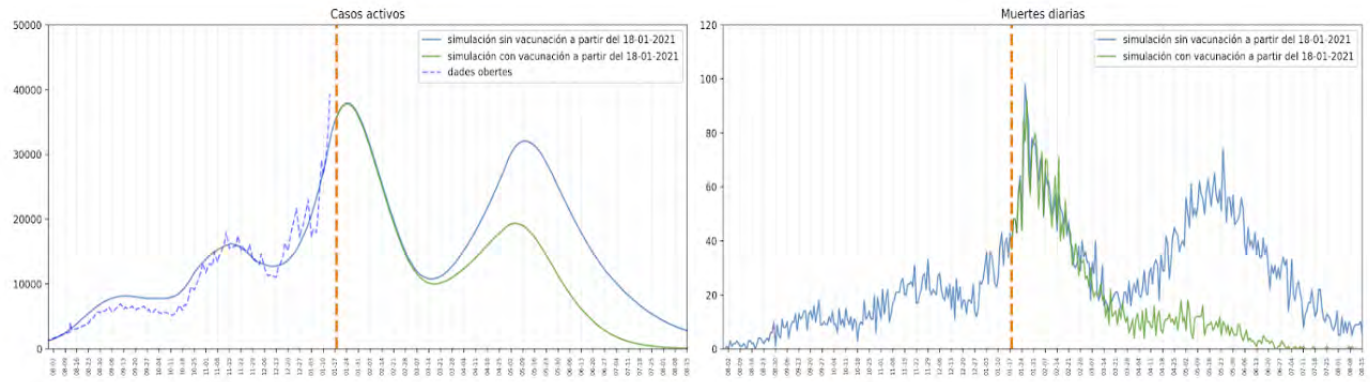
Figure 7.
Agent-based model predictions according to different contact tracing scenarios.
The solid line represents the number of cases detected (blue) and deaths (green) in case of not implementing any type of contact tracing. The dashed lines represent the different simulated contact tracing scenarios (40%, 80% and 100%)



With the advent of the SARS-CoV-2 vaccines, we expanded the REINA model in the Spring of 2021 to include different vaccination scenarios and their impact on the expected number of COVID-19 cases, hospitalizations and deaths.

Figure 8 depicts the predictions provided by this model as of January of 2021. Note how the model correctly predicts the fourth wave of infections that took place in the Summer of 2021, with significantly milder impact in terms of deaths (right-hand graph, green curve) than if there had not been any vaccination.

Figure 8. Agent-based model predictions including vaccination

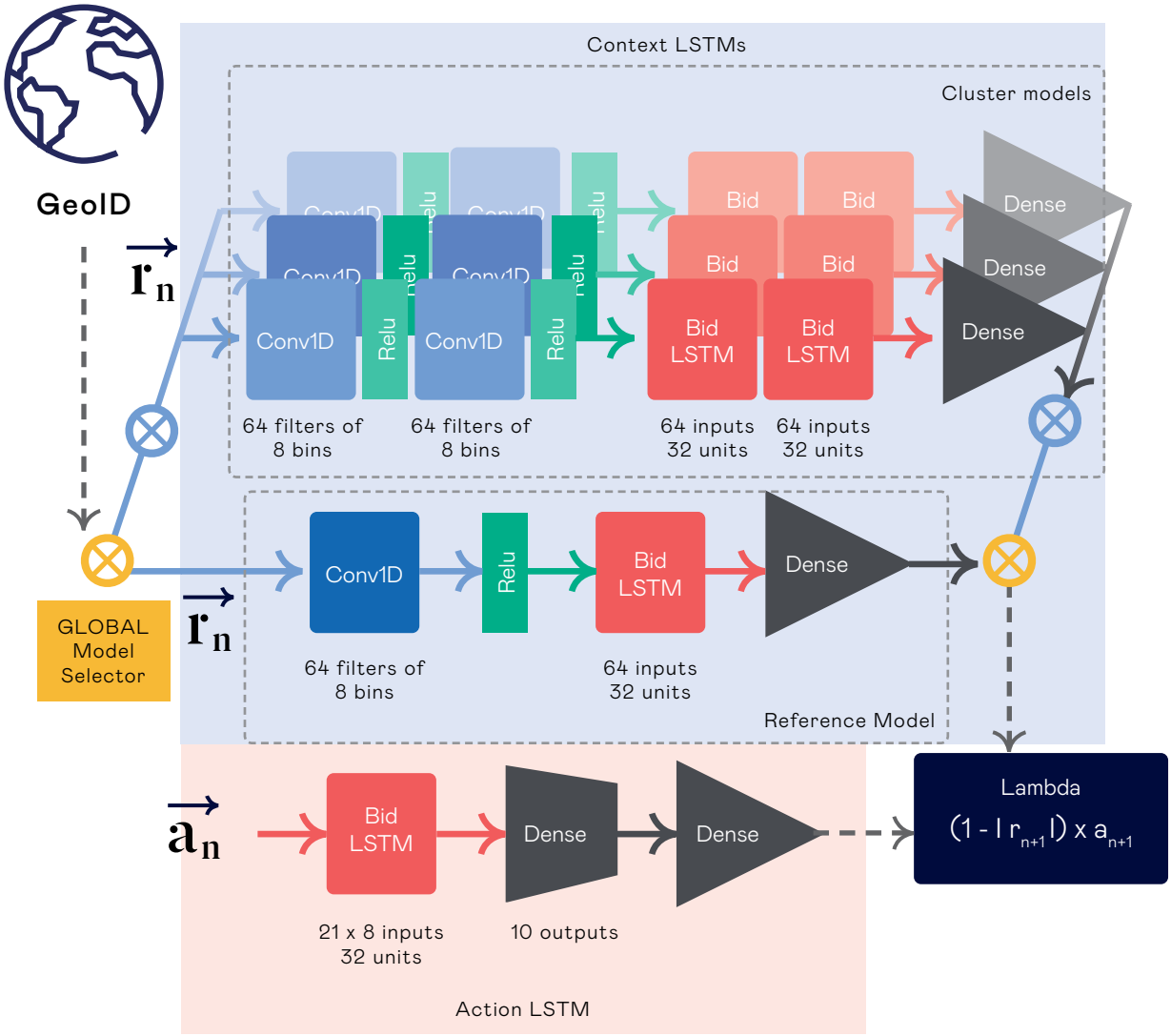


Model based on deep neural networks

The third methodological approach that we developed to model the progression of the pandemic is based on deep neural networks. The main objective is to have a predictive model of the number of cases that considers the non-pharmaceutical interventions (NPI) applied in each country/region. This model was developed in the context of the XPRIZE Pandemic Response Challenge world competition, where we had to create computational epidemiological models for 236 regions/countries of the world.

Figure 9 shows the architecture of this model. It consists of 2 LSTMs trained in parallel, whose outputs are combined into a Lambda module. The first LSTM is called context LSTM and models in number of cases over time. The second LSTM is called action LSTM and models the time series of non-pharmacological interventions (NPIs) implemented at each moment to contain the pandemic. We developed a bank of 9 context LSTMs applied to different countries/regions of the world.

Figure 9. Epidemiological model based on deep neural networks



The model described above was part of the solution that we provided to the XPRIZE competition. Our team, ValencialA4COVID, was declared the world winner of the XPRIZE challenge, being the first time that a Spanish team wins an XPRIZE competition. The technical details of our model are described in this publication [25], winner of the award for best scientific article in Data Science at ECML-PKDD 2021.

This model was extensively used to make daily predictions of COVID-19 cases in the Valencian Region of Spain since the end of December 2020. Figure 10 illustrates the predictions made by the model before the third and sixth waves of COVID-19 cases in January of 2021 and 2022, respectively. Note how accurate the predictions are when compared to the real data (yellow dotted line in the figures).

Figure 10. Predictions of COVID-19 cases by the deep neural network-based epidemiological model for the Valencian region of Spain

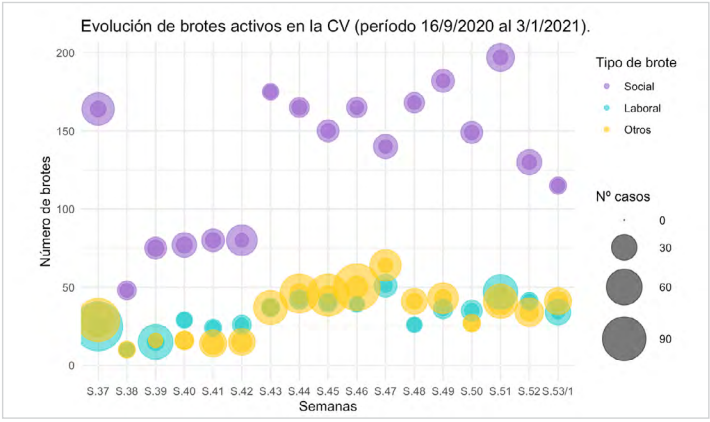


Model of evolution of epidemic outbreaks

As of September 2020, thanks to increased data availability, we were able to study the dynamics of the identified outbreaks of SARS-CoV-2 and analyzed their characteristics in terms of their origin. This information is of interest to be able to assess a *posteriori* if the NPI's implemented locally impact the containment of the disease. After a first analysis of the outbreaks, we observed that they follow a pattern based on free distributions of scale, which is to be expected given that these models are behind human mobility [18]. After this analysis, we established as a control measure the maximum size of 90% of the outbreaks of each of the following types: Social, Labor, and Other types.

Figure 11 illustrates the temporal evolution of the size and number of outbreaks of three different types (social, labor and others) in the Valencian region of Spain between September and December of 2020. As seen in the Figure, the most common type of outbreak in the Valencian region of Spain was outbreaks of social nature. We observe a decrease in the number and size of outbreaks of social origin, starting at the end of 2020, which could be indicative of an increase in community transmission and a decrease in the number of cases linked to outbreaks given that the third wave of infections started at that point in time.

Figure 11.
Weekly evolution of COVID-19 outbreaks.
The outer circle indicates the size of the largest outbreak of that type and the inner circle indicates the size of the 90th percentile of outbreaks of that type.



This analysis allows us to determine if the temporal evolution of the open outbreaks corresponds to the observed progression of the disease or if a significant portion of new cases originate mostly from community transmission and therefore are not classified as outbreaks.

3.
Predictive models

Predictive model by health department of the number of positive cases

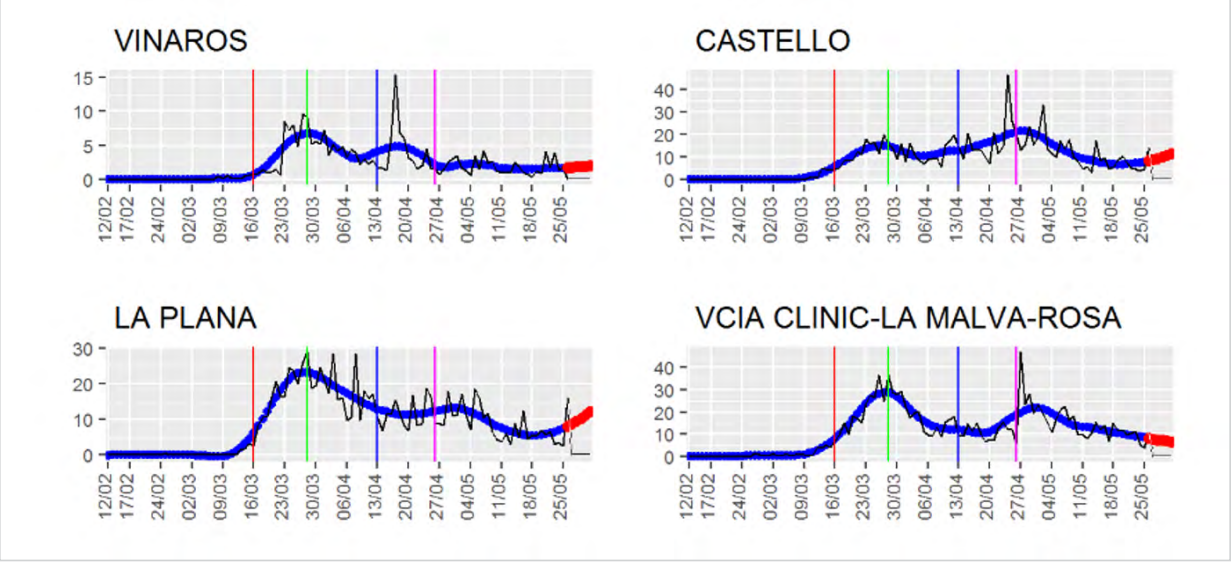
In this area of work, we developed predictive models to estimate relevant variables on a daily basis and in each of the 24 health departments of the Valencian Region of Spain, namely: the number of COVID-19 cases, the number of hospitalizations, active cases, people admitted to intensive care and deceased with a time horizon of 5 days.

To make COVID-19 case predictions, we used a smoothing model based on a non-parametric regression method adjusted to the neighborhood in a time window of 5 days to detect trends and cycles. Short-term predictive models are ARIMA-type models, with different structures for each health department and GAM models using historical observations from each health department as predictors.

We applied these models daily to the time series of number of COVID-19 cases, number of hospitalizations both in hospital and in intensive care units, and deceased.

Figure 12 illustrates the predictions for four health departments in the Valencian region of Spain as of May 20th, 2020. The solid black line represents the number of COVID-19 cases, the blue dots correspond to the data smoothed to the past, the red dots are the predictions provided by our method. The vertical lines indicate the different phases of the de-escalation: red line, date of the start of the state of alarm (16/03/2020); green line, 10 days after the state of alarm (26/03/2020); blue line, day of the partial lifting of the confinement (13/04/2020); purple line indicates the beginning of de-escalation of the confinement measures.

Figure 12. Predictions of COVID-19 cases by health department
(illustrative example for 4 health departments).



Computational epidemiological models that predict the number of COVID-19 cases are undoubtedly of paramount importance. However, predictive models of other variables -- especially of the number of hospitalizations and of intensive care units-- and with other geographic scales --such as at the health department level-- are of vital importance to support public policy-making. These models allow health authorities to prepare resources to prevent a system collapse. The methodology that we developed to build this type of models is different from the traditional computational epidemiological models, such as the SEIR model previously described.

We therefore developed three families of models to predict the number of **hospitalizations** and **intensive care** units' usage in each of the 24 health departments of the Valencian Region of Spain.

Spatio-temporal model by health department of the number of daily hospitalizations

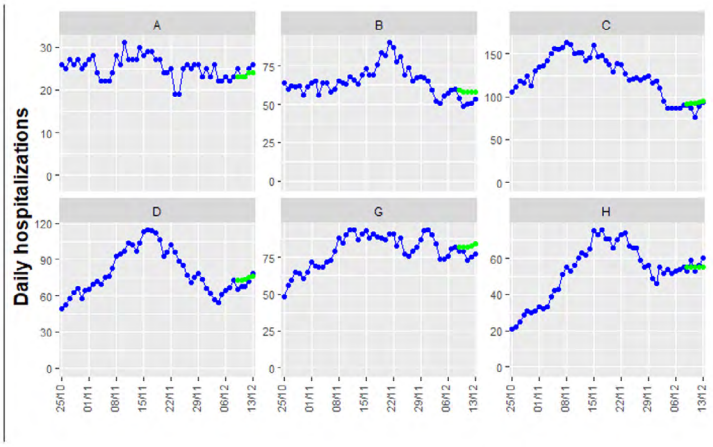
To predict the number of hospitalizations, linear models under an assumption of normality are not appropriate: generalized linear models with negative Poisson, binomial or binomial distributions are more suitable given the nature and characteristics of the underlying phenomenon.

As we have described above, human mobility gives rise to a spatio-temporal dynamic in the spread of infectious disease that should also be considered in the modeling. For this reason, spatio-temporal models are necessary.

In our case, each spatial area is the geographical area served by a health department. Each health department D has K neighbors which are the adjacent health departments. In our model, we assume that the cases and hospitalizations observed in D depend not only on what happens in D but also on what happens in their K neighbors given the mobility between health departments. A common challenge when modeling this type of data is autocorrelation, meaning that observations in geographically close areas and in temporarily close time periods tend to have more similar values than observations in more separate areas and time periods.

The model that we developed takes this factor into consideration to provide predictions of the number of hospitalizations that will occur in the next 5 days in each health department.

Figure 13.
Predictions of daily hospitalizations by health department (real data in blue; predictions at 5 days in green)
Illustrative example for 6 health departments



Prediction model of the percentage of ICU occupancy at 7 days by health department

Since mid-March of 2020, as the number of COVID-19 cases increased, so did the number of hospitalized patients, leading to unprecedented levels of hospital occupancy near saturation levels. Therefore, having access to reliable, accurate predictions of intensive care occupancy was of paramount importance to optimize the management of the health departments.

To predict future occupancies of ICUs, we developed a model based on deep neural networks.

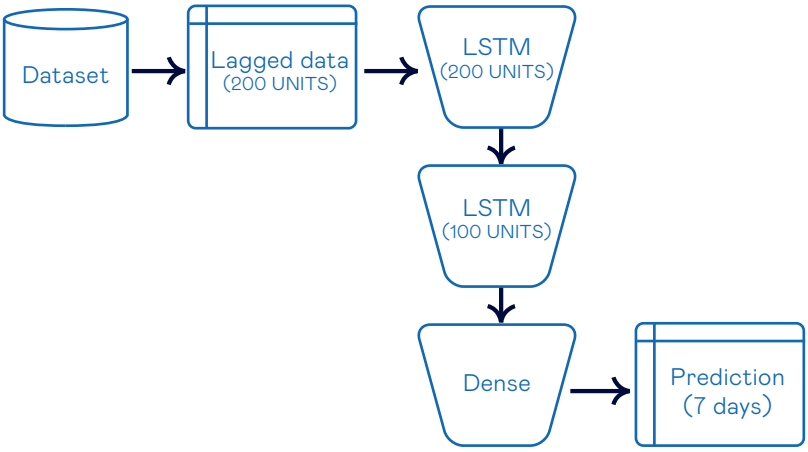
The objective is to predict, 7 days in advance, the percentage of occupancy of ICU beds in each health department of the Valencian region of Spain. The information used to make such a prediction, in addition to the history of ICU occupancy, is the number of COVID-19 cases and hospitalizations due to coronavirus in each health department.

Deep neural network-based model

We used recurrent neural networks (LSTMs) to predict the percentage of occupancy in the ICUs. These networks integrate feedback loops and are very well suited to model time series.

Figure 14 shows the architecture of the LSTMs-based model. As can be seen, it is a stacked LSTM model with two layers, of 200 and 100 LST neurons respectively. The 5-day prediction is given by the 3rd layer, consisting of a single output neuron. In total, this model will have 287-701 parameters that are trained with data from the past.

Figure 14.
LSTM-based model architecture for ICU occupancy prediction



We trained the model with all the available historical data on ICU occupancy and the number of COVID-19 cases of coronavirus for each health department in the Valencian Region of Spain. We used as a loss function the Mean Square Error (MSE) and the Adam optimizer. We applied this model every day to predict the percentage of ICU occupancy in the next 5 days. This model produced competitive predictions, which were used every day to support planning in the departments of health.

4.
Citizen Science

Given the exponential growth in the number of confirmed cases of SARS-CoV-2 in Spain in March of 2020, being able to make a rapid assessment of the situation and perception of citizens was of vital importance in the context of the design, prioritization, and decision-making of public policies. Therefore, in March of 2020 we designed and launched a large *online* citizen survey called *Covid19impactsurvey* [19], which, with more than 470,000 responses from Spain and more than 220,000 from other countries in the world, has become one the largest citizen surveys on COVID-19 to date.

The survey is anonymous and was originally structured in 26 questions that include demographic information about the participants (country, age, sex, postal code), their situation in the home (type of household, number of co-inhabitants in the household and their ages), their social behavior (individual protection measures, perception of security of different activities, estimated number of close contacts, if they have had close contact with someone infected with coronavirus, etc ...), their perception of the measures taken to alleviate the pandemic, their work and economic situation, the emotional impact that the pandemic is having on their lives and their state of health. In June of 2021, we added 6 questions related to social isolation by means of the Lubben 6-item social isolation scale.

In our experience, this survey is a valuable tool to shed light on issues of great public importance for which there is no systematically captured data. Given the success and value of the analysis of the responses of the first days after the launch of the same (from March 28th to April 4th, 2020), we decided to keep the survey online and analyze the data in weekly waves. The questions, analysis methodology and main conclusions of the analysis of the responses of the first wave are described in a scientific article published in JMIR [20].

In this section we summarize, briefly, the most important results.

First, we found that close contacts play a crucial role in coronavirus transmission. Among respondents who reported testing positive for SARS-CoV-2 in the first wave, 81% also reported having close contact with a coronavirus-infected person they knew. That is, the probable source of infection was known in most cases. This percentage is so high probably because at the end of March and the beginning of April of 2020 we were in a situation of confinement. Since the suspension of the state of alarm and the establishment of the “new normal”, this percentage decreased to 65%. However, it is still a high percentage. Household members are the most frequently cited close contact, followed by family and friends and co-workers.

Second, we identified statistically significant gender differences in the impact of the pandemic that place women in a situation of greater vulnerability or exposure compared to men. Likewise, women report the highest levels of anxiety, stress, sadness, and loneliness compared to men in the same age group.

Third, age is also an important factor. We identified statistically significant differences in social contact behavior between participants over the age of 60 and younger participants. Older people were nearly twice as likely to report staying at home without going out than younger participants. We also identified age differences with respect to attitudes towards confinement measures, with young people showing both greater support for the adoption of stricter measures to alleviate the pandemic and, at the same time, less ability to remain in confinement.

Overall, respondents typically demanded more government actions as the cumulative incidence of COVID-19 cases increased. We observed this behavior throughout the entire pandemic.

The economic impact of the pandemic is evident. The most impacted sectors according to the survey are hospitality, construction, domestic service, and commerce.

The psychological impact is also evident, particularly among young women (18-29 years old), who have been reporting the highest levels of stress (50%), anxiety (46%) and abusive use of technology (57%) throughout the pandemic. Young men (18-29 years) report the highest levels of arguments at home, alcohol abuse and drug abuse, sadness, and loneliness, with levels higher than those reported by people over 60. Also noteworthy are the high levels of abusive use of technology by children in families with children which surpassed 70% at some points during the pandemic.

Regarding personal protection measures, we observed a widespread (90% or higher) use of masks, hand disinfection and intention to get vaccinated. The least adopted measure, especially by young people, was indoor ventilation. Physical distancing measures (avoiding hugging and kissing and shaking hands; maintaining physical distancing; limiting close contacts) were significantly less adopted by young respondents (18-29 years old) when compared to older adults.

Individual sport, shopping in small shops and attendance at places by appointment such as hairdressers are the activities that participants considered safest with respect to the probability they entail of contracting coronavirus. Those considered less safe are traveling by plane, attending religious services, and using public transport. Roughly a third of respondents believed that it was possible to attend schools and institutes with low risk of coronavirus infection.

A worrying aspect throughout the pandemic is the high percentage of participants (about 50%) who report not being able to do an effective quarantine if necessary. We observed very significant differences in age, being young people significantly less likely to be able to do an effective quarantine when compared to older adults. The main reason is the sharing of the home, followed by the care of children or other people, psychological reasons (including the fear of stigmatization) and economic-labor reasons.

Finally, during the first wave of the survey, we developed a logistic regression model to estimate the prevalence of SARS-CoV-2 from three survey questions (symptoms, close contact with someone in the household positive for coronavirus, and sex). This model was important as there was a scarcity of PCRs at the time. Our model estimated in April of 2020 a prevalence of 5% nationwide, very aligned with the estimate provided by the seroprevalence study carried out by the Carlos III Institute.

Based on the responses of the citizen survey, we built models to predict two variables of special relevance in decision-making during confinement: the will reported by citizens to remain confined and the satisfaction with the measures adopted by the government. Such models are based on classification trees [21, 22], and pattern extraction algorithms [23]. They highlight which combinations of survey responses lead (and with what probability) to certain levels of willingness to remain confined and to the assessment of government measures. This analysis was published in Nature Scientific Reports [26].

The patterns extracted through trees and classification rules allowed us to propose a series of very specific communicative actions to the population, depending on their demographic, socio-economic and health situation. Some of these recommendations pointed towards the need for progressive de-escalation without relaxing awareness messages to young people; the convenience of offering alternative isolation solutions to intergenerational families with difficulties for confinement at home and the active promotion of teleworking to reduce mobility, as well as the importance of detecting asymptomatic patients in the working-age population, among others.

We also made a comparative analysis of the effectiveness of TTI (trace-test-isolate) pandemic control strategies in Spain and Italy from June 2020 to January 2021, to shed light on the factors that could have contributed to the emergence of the second wave in autumn 2020 [27].

Finally, worried about the social impact of the sustained implementation of the confinement measures, we analyzed the prevalence of social isolation in Spain in the second half of 2021 [28]. We obtained worrisome results, with an estimated prevalence of social isolation of almost 26% in the general population. We observed age and gender differences, with the largest prevalence of isolation among middle-aged individuals; a strong relationship between economic impact and social isolation; and differences in social isolation, depending on the number of COVID-19 protection measures and on the perception of coronavirus infection risk by the participants in the study.

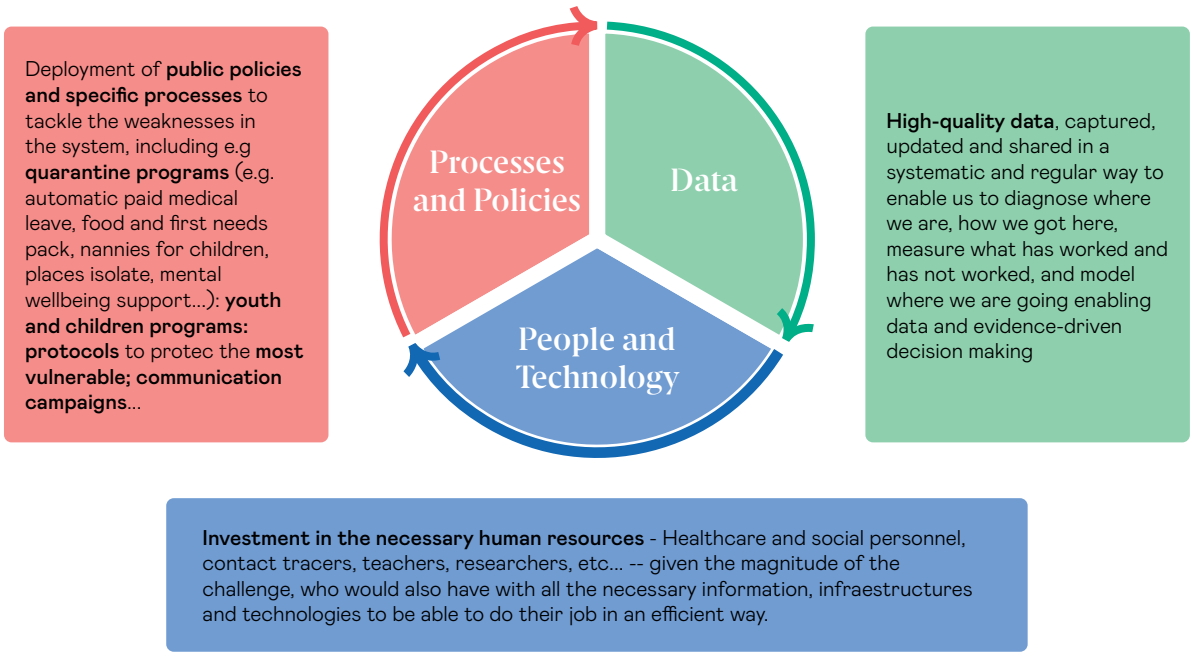
5. Conclusions

With our work during the two years of COVID-19 pandemic, we have learned a pandemic is not just a public health issue, but a societal challenge that requires holistic, multi-disciplinary solutions: there are opportunities to develop synergies between technological, health and social tools and processes to better respond to the complexities and challenges that a pandemic represents.

We have faced a pandemic that has lasted two years and which has shown us the importance of achieving sustainable solutions from a variety of perspectives, including the psychological, economic, environmental and social dimensions.

This holistic approach entails not only the development of a clear vision and in collaboration with the different agents but more importantly the urgent execution of it, including the following three key aspects that form a virtuous circle, illustrated in the Figure below.

Figure 15. Circle of data, people, technology, processes, and policies in a holistic approach to fight against the SARS-CoV-2 pandemic



1. The availability of specific indicators and quality data, captured, updated, and shared in a systematic and regular manner, that allow us to make a diagnosis of where we are, analyze the causes, determine what has worked and what has not worked, and model where we are going, enabling decision-making based on evidence and knowledge, including the decision on whether or not to use an app for contact tracing.

2. Investment in the human resources – health and social personnel, trackers, teachers, researchers ... – necessary for the magnitude of the challenge, having access to the information, infrastructures, and technologies necessary to be able to carry out their work effectively.

3. The deployment of public policies and specific processes to address weaknesses in the system, including programs to facilitate quarantine (for example, immediate paid medical leave, packs of food and basic necessities, caregivers for children or adults who need them, places where they can isolate themselves if it is not possible at home, guarantees not to lose their job, psychological support, etc.); communication campaigns to foster a culture in which people do not relate to others if they have the slightest suspicion of being infected with coronavirus; protocols to protect the most vulnerable groups; regulations to minimize the risk of contagion in places and activities prone to outbreaks (more examples, food processing plants, discos, family celebrations,...) and a set of specific actions for children, adolescents and young people, which we cannot forget, are what are suffering most intensely the emotional burden of the pandemic.

Let's work together, people and technology, civil society, companies, and administrations, in the fight against the virus. The work described in this article is an example in this direction. Unity, without a doubt, is what gives us strength.

Acknowledgements

The work described in this article has been partially financed by the Generalitat Valenciana (Decreto 202/2020 and Convenio Singular 2021 between the Innovation, Universities, Science and Digital Society Ministry and the ELLIS Unit Alicante Foundation), the BBVA Foundation (IA4COVID19 project) and the Supera COVID fund of Banco de Santander with the CRUE (CD4COVID19 project).

Bibliography

[1] https://www.muyinteresante.es/salud/articulo/una-foto-semanal-de-la-situacion-del-coronavirus-en-espana-811585915653

[2] https://www.lavanguardia.com/politica/20200424/48693659917/una-encuesta-recoge-que-42-de-espanoles-podria-estar-un-mes-mas-confinado.html

[3] https://www.rtve.es/alacarta/audios/por-tres-razones/tres-razones-inteligencia-artificial-para-combatir-pandemia-24-04-20/5564188/

[4] https://www.politico.eu/newsletter/ai-decoded/politico-ai-decoded-how-ai-is-helping-fight-a-pandemic-europes-coronavirus-app-insights-from-valencia/

[5] https://www.youtube.com/watch?feature=youtu.be&v=A06j9Nv8-CA

[6] A. Wesolowski A, C.O. Buckee, L. Bengtsson, E. Wetter, X. Lu, A.J. Tatem. "Commentary: containing the ebola outbreak - the potential and challenge of mobile network data". *PLoS Curr*. 2014;6: currents.outbreaks.0177e7fcf52217b8b634376e2f3efc5e. Published 2014 Sep 29. doi:10.1371/currents.outbreaks.0177e7fcf52217b8b634376e2f3efc5e

[7] C.M. Peak, A. Wesolowski, E. zu Erbach-Schoenberg, A. J Tatem, E. Wetter, X. Lu, D. Power, E. Weidman-Grunewald, S. Ramos, S. Moritz et al. Population mobility reductions associated with travel restrictions during the Ebola epidemic in Sierra Leone: use of mobile phone data. International Journal of Epidemiology, Volume 47, Issue 5, October 2018, Pages 1562–1570, https://doi.org/10.1093/ije/dyy095

[8] A. Wesolowski, N. Eagle, A.J. Tatem, D.LSmith, A.M. Noor, R.W Snow, C.O. Buckee (2012). "Quantifying the impact of human mobility on malaria". Science 338:267-270.

[9] https://www.ine.es/covid/covid_movilidad.htm

[10] https://www.ine.es/covid/exp_movilidad_covid_proyecto.pdf

[11] http://infocoronavirus.gva.es/documents/170024890/170025022/Informe+Movilidad+gva+Abril+2020.pdf/2d728f25-f202-4e7d-81ca-cc25eef9e7d4

[12] http://infocoronavirus.gva.es/documents/170024890/170025022 Informe+Movilidad+gva+-Mayo+2020.pdf/5b043319-eed9-4a66-8477-d214ffe11c39

[13] M. E. J. Newman. "Modularity and community structure in networks". PNAS 103(23) 8577-8582. 2006

[14] Joan L.Aron and Ira B.Schwartz. "Seasonality and period-doubling bifurcations in an epidemic model". Journal of Theoretical Biology, Volume 110, Issue 4, 21 October 1984, Pages 665-679. https://doi.org/10.1016/S0022-5193(84)80150-2

[15] F. Escolano, P. Suau, B. Bonev. "Information Theory in Computer Vision and Pattern Recognition", Springer 2009.

[16] J.T. Tuomisto, J. Yrjölä, M. Kolehmainen, J. Bonsdorff, J. Pekkanen, T. Tikkanen. "An agent-based epidemic model REINA for COVID-19 to identify destructive policies". medRxiv 2020.04.09.20047498; Doi: https://doi.org/10.1101/2020.04.09.20047498https://www.medrxiv.org/content/10.1101/2020.04.09.20047498v2

[17] Miikkulainen, R., FranconOr. Meyerson, E., Qiu, X., Canzani, E., & Hodjat, B. (2020). From Prediction to Prescription: AI-Based Optimization of Non-Pharmaceutical Interventions for the COVID-19 Pandemic. arXiv preprint arXiv:2005.13766.

[18] A.L. Barabasi, The origin of bursts and heavy tails in human dynamics. Nature, 435:7039, 207-211 (2005).

[19] https://covid19impactsurvey.org

[20] N. Oliver, X. Barber, K. Roomp, K. Roomp. "Assessing the Impact of the COVID-19 Pandemic in Spain: Large-Scale, Online, Self-Reported Population Survey". J Med Internet Res 2020;22(9):e21319. DOI: 10.2196/21319

[21] J.R. Quinlan. "Induction of decision trees". Mach Learn 1, 81–106 (1986). https://doi.org/10.1007/BF00116251

[22] L. Breiman, J. Friedman, C.J. Stone, , R.A. Olshen. "Classification and Regression Trees". Ed. Taylor & Francis. The Wadsworth and Brooks-Cole statistics-probability series (1984).

[23] M. Almiñana, L.F. Escudero, A. Pérez-Martín, A. Rabasa, L. Santamaría, "A classification rule reduction algorithm based on significance domains". TOP, 22. 397-418 (2012)

[24] F. Escolano, M.A. Lozano, E.R. Hancock. "The Entropy of Graph Embeddings: A proxy of Potential Mobility in Covid19 Outbreaks. S+SSPR2020: IAPR Joint International Workshops on Statistical Techniques in Pattern Recognition (SPPR 2020) and Syntactic and Structural Pattern Recognition (SSPR 2020).

[25] M.A. Lozano, O. Garibo, E. Piñol, M. Rebollo et al, Open Data Science to fight COVID-19: Winning the 500k XPRIZE Pandemic Response Challenge, Proc. European Conf. on Machine Learning and Data Mining (ECML-PKDD), 2021 **Best Paper Award (Data Science track)**

[26] M. Martinez-Garcia, To. Rabasa, X. Barber, K. Polotskaya, K. Roomp, N. Oliver, Key factors affecting people's unwillingness to be confined during the COVID-19 pandemic in Spain: a large-scale population study, Nature Scientific Reports, 2021

[27] M. De Nadai, K. Roomp, B. Lepri, N. Oliver, The Impact of Control and Mitigation Strategies during the Second Wave of COVID-19 Infections in Spain and Italy, medRxiv, 2021

[28] M. Martinez-Garcia, E. Sansano-Sansano, A. Castillo-Hornero, R. Femenia, K. Roomp, N. Oliver, Social isolation during the COVID-19 pandemic in Spain: a population study, MedRxiv, 2022

esade

EsadeEcPol - Center
for Economic Policy

Open Internet Governance Institute

The OIGI is EsadeEcPol's effort to shape debates on Internet, data & digital governance both in Spain and across the European Union, while simultaneously contributing to a better understanding of how best use new data and AI-related tools to support and improve policymaking.

We intend to contribute in a balanced and evidence-based manner, departing from the delimitation of weighed dilemmas to focus on offering viable solutions. Our ultimate goal is to help building a system of global and open internet governance, fostering the best possible digital environment among the many future worlds that open before us.

Supported by

Meta