

Trabajar con datos abiertos en tiempos de pandemia: uso de *covidDATA-19*

Working with open data in times of pandemics: using *covidDATA-19*

Antonia Ferrer-Sapena; José-Manuel Calabuig; Fernanda Peset; Isabel Sánchez-del-Toro

Cómo citar este artículo:

Ferrer-Sapena, Antonia; Calabuig, José-Manuel; Peset, Fernanda; Sánchez-del-Toro, Isabel (2020). "Trabajar con datos abiertos en tiempos de pandemia: uso de *covidDATA-19*". *Profesional de la información*, v. 29, n. 4, e290421.

<https://doi.org/10.3145/epi.2020.jul.21>

Artículo recibido el 18-04-2020
Aceptación definitiva: 28-05-2020



Antonia Ferrer-Sapena ✉
<https://orcid.org/0000-0001-6432-917X>

Universitat Politècnica de València
Instituto Universitario de Matemática Pura y Aplicada
Camí de Vera, s/n. 46022 València, España
anfesa@upv.es



José-Manuel Calabuig
<https://orcid.org/0000-0001-8398-8664>

Universitat Politècnica de València
Instituto Universitario de Matemática Pura y Aplicada
Camí de Vera, s/n. 46022 València, España
jmcalabu@mat.upv.es



Fernanda Peset
<https://orcid.org/0000-0003-3706-6532>

Universitat Politècnica de València
Instituto Universitario de Matemática Pura y Aplicada
Camí de Vera, s/n. 46022 València, España
mpesetm@upv.es



Isabel Sánchez-del-Toro
<https://orcid.org/0000-0003-2966-986X>

Universitat Politècnica de València
Camí de Vera, s/n. 46022 València, España
masande7@upvnet.upv.es

Resumen

Uno de los problemas al trabajar con datos es la falta de uniformidad en los criterios de su recogida, lo que hace que la reutilización sea difícil, cuando no imposible. Deben tenerse claros los protocolos de recolección y estandarización de los datos. Este hecho se ha podido constatar en la situación de pandemia global provocada por la Covid-19, en la que se han necesitado con urgencia datos uniformes para tomar decisiones rápidas y mantener informada a la población. En este trabajo se observan las dificultades para crear un portal que analiza la evolución de los datos de la pandemia en España. *CovidDATA-19* es una web de datos abiertos que presenta visualizaciones interactivas y permite descargar los datos filtrados según las elecciones del usuario. En su creación se ha puesto de manifiesto la escasez de fuentes oficiales fácilmente reutilizables y la fragilidad de la confianza en los datos en estos momentos de crisis.

Palabras clave

Covid-19; Coronavirus; Pandemias; Información de salud; Datos abiertos; Datos sanitarios; *Datadista*; *CovidDATA-19*; España.

Abstract

One of the problems when working with data is the lack of uniformity in the criteria for their collection, which makes their reuse difficult if not impossible. Data collection and standardization protocols must be clear. This problem has

Financiación

Este artículo está financiado por la *Cátedra de Transparencia y Gestión de Datos* de la *Universitat Politècnica de València* y la *Conselleria de Participación, Transparencia, Cooperación y Calidad Democrática* de la *Generalitat Valenciana*. Y parcialmente por el Plan I+D del *Mineco*, proyecto *Datause. Metodologías estables para evaluar y medir la calidad, interoperabilidad, blockchain y reutilización de los datos abiertos en el área agrícola* PID2019-105708RB-C21

been seen in the global pandemic caused by Covid-19, in which uniform data have been urgently needed to make rapid decisions and keep the population informed. The difficulties of creating a portal that analyzes the evolution of the data on the pandemic in Spain are discussed. *CovidDATA-19* is an open data website that features interactive visualizations and allows filtered data to be downloaded according to the user's choices. Its development has highlighted the scarcity of easily reusable official sources and the fragility of trust in such data during this crisis.

Keywords

Covid-19; Coronavirus; Pandemics; Health information; Open data; Health data; *Datadista*; *CovidDATA-19*; Spain.

1. Introducción

La aparición de la pandemia global derivada del virus SARS-CoV-2 conocido como enfermedad por coronavirus o Covid-19 ha desatado una fiebre de datos que hasta ahora estaba limitada a algunos sectores. Por ejemplo, los gobiernos necesitan datos actualizados y fiables para la toma de decisiones, los periodistas los necesitan para informar a la población y los desarrolladores para crear nuevos servicios. Ahora los datos se requieren en una suerte de *datafilia* que ha incluido una urgencia inusitada debido a la tensión en los sistemas sanitarios y las medidas de emergencia sobre confinamiento e hibernación de la economía. Todo ha sido muy rápido: los primeros casos de China se refieren en diciembre de 2019, la OMS alertó la emergencia pública el 30 de enero de 2020 y el 11 de marzo la declaró pandemia. El acrónimo Covid-19 (*Corona Virus Disease*) se acuñó el 11 de febrero (*World Health Organization-WHO*, 2020) para desvincularlo de cualquier origen geográfico, animal u otros modos de coronavirus.

Se define como una enfermedad que provoca problemas severos en el sistema respiratorio, pero también se asocia con alteraciones inflamatorias o de los sistemas vascular, renal... (**Huang et al.**, 2020; **Madjid et al.**, 2020). La rápida transmisión agrava la situación mundial hasta los 4.801.202 de casos confirmados y 318.935 muertes el 21 mayo de 2020 según el *WHO Coronavirus Disease (Covid-19) Dashboard* (2020). La falta de conocimiento sobre su forma de transmisión, evolución de la enfermedad y la existencia de contagiados que no presentan síntomas dificulta el control sanitario de la epidemia (**Bai et al.**, 2020; **Rothe et al.**, 2020). Las medidas adoptadas difieren según países, incluyendo distanciamiento social, lavado de manos, higiene de superficies, aislamiento..., con graves secuelas socioeconómicas.

Para encontrar una solución a la Covid y sus consecuencias se están publicando de forma rápida y abierta múltiples estudios (**Torres-Salinas**, 2020; **Torres-Salinas et al.**, 2020), especialmente en el área sanitaria y de la aplicación de inteligencia artificial a los datos. Por el contrario, aspectos más sociales, centrados en la disponibilidad de datos fiables y abiertos, no están siendo suficientemente atendidos, según hemos constatado en las bases de datos *Scopus* y *Web of Science*. Dado que los datos condicionan tanto las políticas nacionales como las medidas sociales o el reparto de recursos económicos y humanos, se demuestra que son válidos los principios FAIR (*findable, accessible, interoperable, reusable*) que se aplican en el campo científico: la necesidad de encontrarlos, acceder a ellos de forma interoperable y reutilizable (**Wilkinson et al.**, 2016). Así, consideramos que los datos de la Administración sobre infectados, fallecidos, ingresos UCI y recuperados, complementados con datos demográficos y topográficos o número de camas, de mascarillas..., deberían cumplirlos. Pero lo que encontramos, sin entrar en detalle en la recopilación de **Cobarsí-Morales** (2020), son servicios con datos dispersos, poco homogéneos en su forma de acceso, y con un nivel de interoperabilidad y capacidad de ser reutilizados escasa, exceptuando algunos casos.

Han proliferado múltiples paneles web que utilizan datos crudos de muy distinto origen para monitorizar la evolución y ayudar a tomar decisiones a las autoridades que utilizan datos crudos de muy distinto origen. Los tres datos principales son infectados, fallecidos y curados por distribución geográfica o de género. Las visualizaciones pueden incluir mapas, curvas de contagiados, o la proporción de éstos y fallecidos por población (*The New York Times*, 2020). Uno de los primeros fue el panel de la *Johns Hopkins University-CSSE*, que apareció el 22 de enero (**Dong; Du; Gardner**, 2020) y que ofrece sus series temporales en *GitHub*, al igual que el de *The New York Times*. Por sentido común, los mejores datos son los oficiales, como los del panel de *WHO* mencionado o el del *Instituto de Salud Carlos III* (2020) para España. Ambos permiten descargar series diarias desde sus páginas, pero automatizar el proceso no es fácil. En líneas generales, apreciamos proyectos derivados de la investigación, de los medios de información, de fuentes oficiales, o de secciones en proyectos globales como *Worldometer* (2020). Para valorar qué datos crudos reutilizar para llevar a cabo un producto se debe tomar en cuenta no sólo este origen, la autoridad emisora de la información, sino también la apertura o facilidad para reutilizar los datos ya que, por ejemplo, *GitHub* permite atacarlos automáticamente pero no otros mecanismos como hemos mencionado.

Este trabajo tiene cuatro secciones. Tras esta primera sección introductoria, en la segunda se presenta un análisis sobre cómo ha evolucionado la recogida de datos sobre Covid-19 que el *Ministerio de Sanidad* reclama a las CCAA. Se muestra cómo, mediante el panel interactivo *covidDATA-19*, ha sido posible analizar gran parte de los problemas surgidos en esta recogida. Es el caso por ejemplo de las diferentes formas de proporcionar los datos las CCAA, lo que ha provocado que se hayan tenido que revisar las series temporales de datos en varias ocasiones, siendo necesario eliminar temporalmente a una CCAA de la serie. En la tercera sección se presenta el panel *covidDATA-19*, elaborado a partir de la fuente de datos *Datadista*, y construido mediante el programa estadístico *R*. Es un instrumento que, a partir de tablas de datos y distintas visualizaciones, aporta una visión detallada de aspectos seleccionados a partir de los datos fiables obtenidos de las CCAA con los que se puedan ensayar algoritmos de predicción de la pandemia. Para ello se ofrece un resumen de datos de evolución tanto diarios como acumulados sobre aspectos esenciales: casos confirmados, fallecidos, casos UCI y altas.

También presenta mapas comparativos con las CCAA con mayores valores en el día seleccionado. Por último, incluye otros datos de visualizaciones interactivas de España desagregados por sexo y edad, un ranking por CCAA que presenta los datos de casos confirmados, casos UCI, fallecidos y altas; y otros como la distribución de mascarillas. En la cuarta y última sección se exponen las lecciones aprendidas en la elaboración y uso del portal.

2. Los datos en España

Veamos a continuación la situación de los datos que se utilizan para sistematizar la información que los gobiernos de las comunidades autónomas reportan al *Ministerio de Sanidad* para hacer el seguimiento del coronavirus en España. A partir de la declaración del estado de alarma el 14 de marzo (España, 2020) el *Ministerio* asumió la coordinación de la lucha contra la pandemia en España, donde las competencias en sanidad están descentralizadas en 17 comunidades y 2 ciudades autónomas (CCAA en adelante). Como fuente oficial distribuidora de los datos reclama a los gobiernos autonómicos esa información de manera muy precisa. Las CCAA deben reportar a la *Red Nacional de Vigilancia Epidemiológica (Renave)* los datos de los centros sanitarios sobre:

- casos confirmados (acumulado)
- casos de personal sanitario
- hospitalizaciones
- ingresos en UCI
- recuperados
- fallecidos.

A partir de ellos el *Ministerio* realiza los informes sobre la situación de la Covid-19. A 18 abril de 2020 hay 22 informes publicados por *Renave* sobre la situación de la Covid-19 en España.

La web del *Ministerio de Sanidad* presenta informes casi diarios con información pormenorizada junto con las anotaciones correctoras, lo que pone de manifiesto las dificultades a la hora de sistematizar la información. En la primera nota (26 marzo) aclaró que

“Los casos confirmados no provienen de la suma de pacientes hospitalizados, curados y fallecidos, ya que no son excluyentes. Pacientes fallecidos y curados pueden haber precisado hospitalización y por tanto computar en ambos grupos. Los pacientes que han precisado UCI también computan en los pacientes que han requerido hospitalización”.

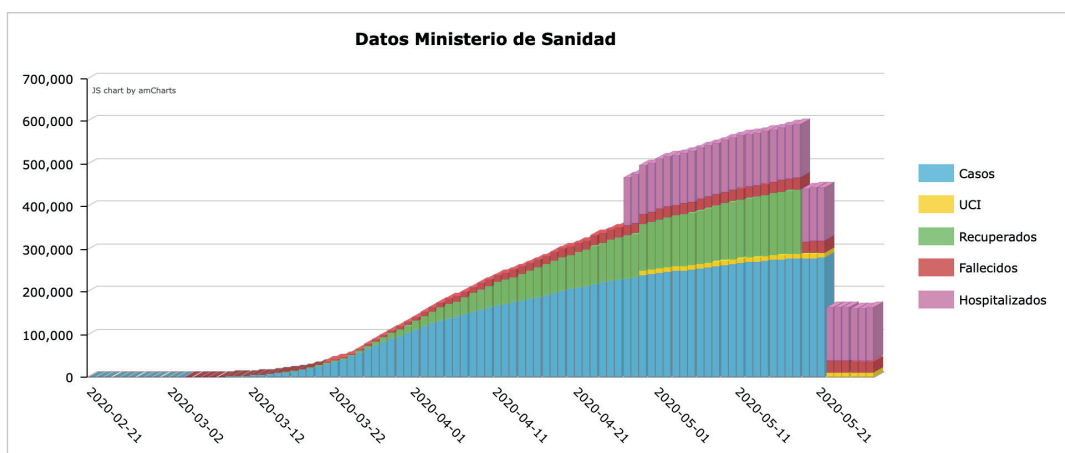


Figura 1. Datos de altas, casos, fallecimientos, ingresados en UCI y hospitalizados en España. Elaborado a partir de *Datadista*. <https://indicame.upv.es/coviddata-19>

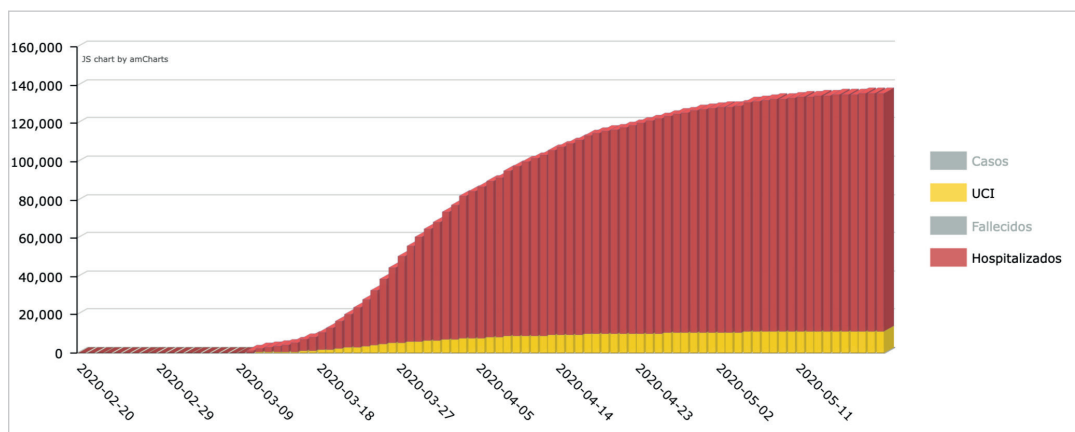


Figura 2. Corte de datos totales de UCI y hospitalizaciones el 8 de abril. Elaborado a partir de *Datadista*. <https://indicame.upv.es/coviddata-19>

El resto de las notas aclaratorias del *Ministerio de Sanidad* incluyen un criterio diferenciador que algunas CCAA han incluido en su recogida de datos y que ha cambiado el modelo de entrega. Es el caso de la prevalencia en casos de hospitalizados y en UCI en el período del 2 al 7 de abril. Hay que indicar que la prevalencia no son los casos acumulados a día de notificación sino sólo los producidos en el mismo día.

El problema más relevante es no haber determinado desde el inicio el método de recogida de información: qué es lo que abarca cada uno de los indicadores de la muestra, y cómo y cuándo debe ser recogido

Las primeras CCAA en presentar estos datos de prevalencia en hospitalizados y en UCI fueron Madrid, Castilla La Mancha y Comunitat Valenciana. El día 3 de abril se añadieron Castilla y León a los datos de prevalencia de hospitalizados y Galicia a los de UCI. El día 7 de abril Castilla y León sólo proporcionó los casos de prevalencia de UCI y el *Ministerio de Sanidad* siguió publicando los datos acumulados de hospitalizados y de UCI.

La primera consecuencia de estas incidencias concluyó el 8 de abril cuando el *Ministerio de Sanidad* anunció que dejaba de publicar el sumatorio de todas las personas que han requerido hospitalización o ingreso en UCI en España porque los datos no reflejaban el total de personas que habían sido hospitalizadas o ingresadas en UCI a lo largo del período de notificación (como hemos señalado, los datos que presentaron las CCAA mencionadas solo eran de prevalencia).

Hay que destacar la notificación del 15 de abril en la que el *Ministerio de Sanidad* aclaró que Andalucía había notificado un total de 169 positivos asintomáticos incluidos en el total de confirmados. La figura 3 resume en una línea temporal todos los cambios notificados hasta el 15 de abril.

La última incidencia tuvo lugar el 17 de abril de 2020. En la comparecencia, Fernando Simón, director del *Centro de Coordinación de Alertas y Emergencias Sanitarias* del *Ministerio de Sanidad*, anunció que el *Ministerio* tenía que corregir la serie histórica de datos. El desencadenante final fue una clara incoherencia en los datos de fallecimientos y altas de Catalunya, la cual se puede observar en la pantalla de resumen y bienvenida del portal *covidDATA-19* (figura 4).

En la primera gráfica de la derecha (acumulado de fallecidos en Catalunya) se observa como el número de fallecidos el día 16 de abril es mayor que el del 17 de abril. En la columna central de datos vemos cómo la variación diaria de fallecidos entre estos dos días ha sido negativa: -103 personas para ser exactos.

Sin entrar en mucho detalle, el motivo de este error es el siguiente: como ya hemos indicado, el *Ministerio de Sanidad* reclamó a las CCAA el número acumulado de fallecidos confirmados positivamente por una prueba. Esto dejaba sin contabilizar a fallecidos que no habían sido certificados por una prueba (como, por ejemplo, los de las residencias de ancianos). Para que las cifras fueran más reales Catalunya solicitó a las funerarias los certificados de defunción en los que el coronavirus pudiera ser la causa de la muerte. Esto provocó un gran aumento en la cifra de fallecimientos (pasando de las 3.756 personas contabilizadas el 14 de abril a las más de 7.000 el día 15 de abril). El *Departament de Salut* de Catalunya emitió el propio 15 de abril dos notas de prensa explicando dicha situación e informando de que se iban a ofrecer al *Ministerio* dos datos relacionados con los fallecidos: los que provenían de hospitalizados (confirmados con test o PCR) y los que provenían de funerarias.

<https://canalsalut.gencat.cat/ca/salut-a-z/c/coronavirus-2019-ncov/notes-premsa>

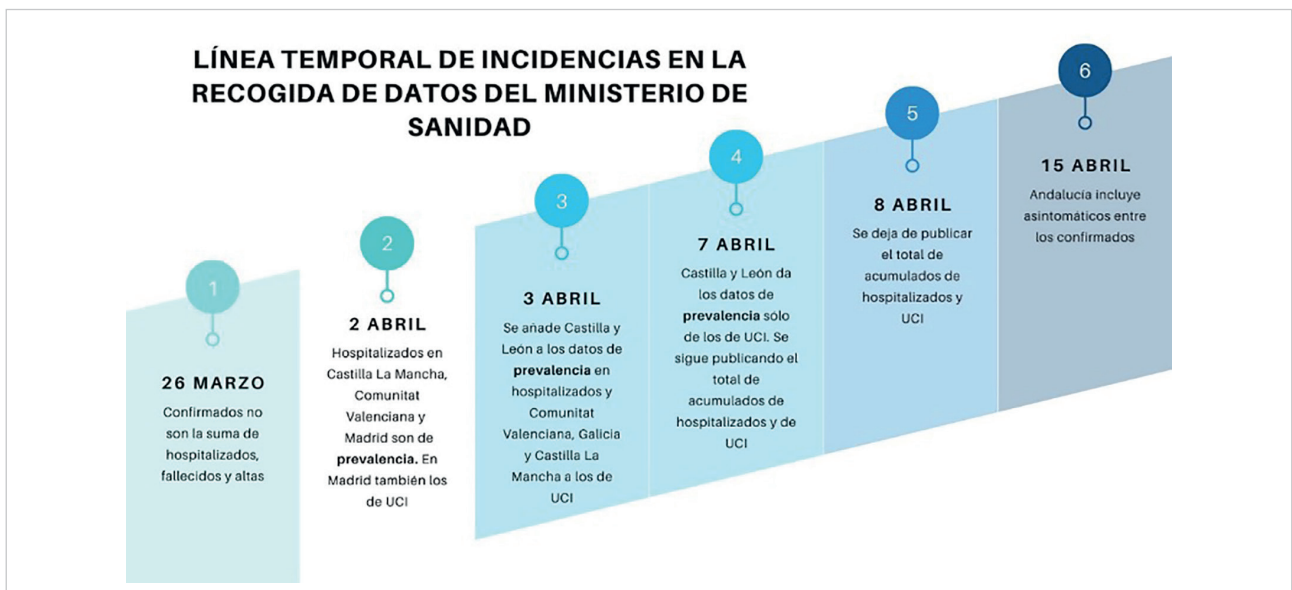


Figura 3. Línea del tiempo elaborada a partir de los datos de los informes del *Centro de coordinación de alertas y emergencias sanitarias*. Actualizaciones n. 56, 63, 64, 68, 69 y 76. Enfermedad por coronavirus (Covid-19).

<https://www.mscbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov-China/situacionActual.htm>

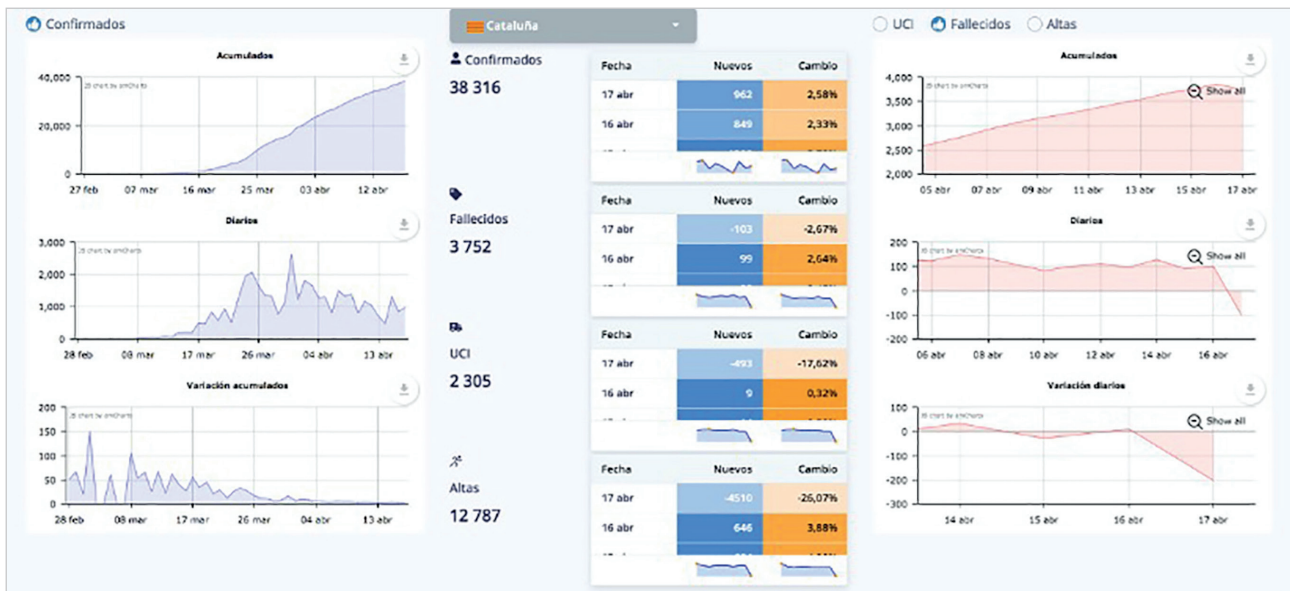


Figura 4. Panel de control principal de covidDATA-19. Elaborado a partir de Datadista.
<https://indicame.upv.es/coviddata-19>

Además de todos estos datos, las CCAA completaban una encuesta individualizada para cada uno de los casos según accedían a la información. La encuesta incluía información clínico-epidemiológica consensuada y aprobada por la *Ponencia de Alertas y Planes de Preparación y Respuesta* y la *Renave*, y la notificaban mediante la plataforma informática SiViES (*Sistema para la Vigilancia en España*), que gestiona el *Centro Nacional de Epidemiología-Instituto de Salud Carlos III*. Se puede tener acceso a la información que se solicitaba en dichas encuestas consultando los *Informes Covid-19* disponibles en formato pdf en:

<https://www.isciii.es/QueHacemos/Servicios/VigilanciaSaludPublicaRENAVE/EnfermedadesTransmisibles/Paginas/InformesCOVID-19.aspx>

También es posible obtener datos sobre Covid-19 en algunos portales autonómicos como Aragón, Asturias, Canarias, Castilla-León y Catalunya. En la mayoría se recuperan desde los portales de datos abiertos, pero algunos han abierto secciones especiales en sus sitios web (ver Anexo).

Como vemos, reutilizar datos desde las fuentes oficiales, sobre todo al inicio de la pandemia, y mantenerlos actualizados de forma coherente, no es fácil. Algunos de los obstáculos que se presentan son:

- el *Ministerio* presenta datos en pdf, con la consiguiente necesidad de convertirlos a formatos reutilizables con otras aplicaciones;
- existen diferencias a la hora de la recogida de datos, acentuadas desde el día 17 de abril ante las críticas del recuento de fallecidos;
- desde el inicio algunas comunidades no presentaban la misma información, ni los datos seguían la misma estructura. Por ello, no describen la misma realidad y la evolución, ni son fácilmente comparables;
- la cantidad de anotaciones mencionadas obligan a revisar y recalcular datos.

Esto también lo observó un grupo de profesionales que se unieron para tratar los datos de manera conjunta. A este proyecto colaborativo denominado *Providencialdata19: Capturando datos por provincias en España* se puede acceder a través del enlace

https://code.montera34.com:4443/numeroteca/covid19/-/blob/master/README_providencialdata19.md

En el mismo se señalan incoherencias en una hoja de cálculo de *Google Drive* del tipo:

- Un cambio metodológico en la Comunidad Canaria el 06/04/2020 donde los datos consolidados del día 05/04/2020 a las 20:00h son casos registrados en función del lugar de declaración y no de residencia. Este detalle que se señala en las Islas Canarias no aparece reflejado en ningún sistema de recogida de información de otras comunidades autónomas. Señalan pues el desfase existente entre los días 05/04/2020 y 06/04/2020 en el número de casos nuevos, llegando a tener un valor negativo en algunas islas. Esto se debe a que los valores cambiaron y son en función del lugar de declaración y no de residencia.
- En Castilla y León, figuran los datos de las personas a las que se les ha practicado PCR y han dado positivo.
- Catalunya ofrece los datos totales diarios, pero no los acumulados. En las defunciones no se indica el municipio y tampoco se indica si son sólo los hospitalarios o si tienen también en cuenta los fallecidos en residencias, en casa...

“ Existe fragilidad de la confianza en los datos en los momentos de crisis ”

- En Galicia no queda claro si el número que aparece es de activos o acumulados. Las cifras de ingresos hospitalarios, fallecimientos, altas, UCIs y casos domiciliarios son globales para la Comunidad Autónoma y no se desglosan ni por provincias ni por distrito sanitario. Hay discrepancias entre lo publicado por el *Ministerio* y la publicado por la *Xunta*.

- En Euskadi tampoco figuran los datos de fallecidos por provincia. Existen también discrepancias entre lo que algunos gobiernos autonómicos publican en *Twitter* y lo que figura en sus informes diarios.

« Estamos asistiendo a una evolución desde la *datafilia* detectada en un principio a una *datafobia* en la que se descalifica de forma absoluta la validez de los datos »

Todos estos inconvenientes, además de la falta de información acerca del sistema de recogida en las distintas comunidades autónomas, hace difícil sistematizar una plataforma de visualización de datos de la evolución de la Covid-19 a nivel nacional, así como hacer previsiones de la evolución.

3. El portal covidDATA-19

Es uno de los portales sobre Covid-19 para el área española construido con datos abiertos desde el 27 de febrero. Es una iniciativa del grupo *MADφ-UPV* y apoyada por la *Cátedra de Transparencia y Gestión de Datos* que desde el 19 de marzo de 2020 ofrece una representación visual de datos fiables con los que ensayar algoritmos de predicción de la pandemia (*Anastassopoulou et al., 2020; Calabuig et al., 2020*). La primera versión se difundió el 3 de abril y la segunda versión el 14. <https://indicame.upv.es/coviddata-19>

La fuente de datos utilizada para el portal ha sido *Datadista*, medio de comunicación dedicado a la investigación periodística.

<https://datadista.com>

Con respecto a la Covid-19 *Datadista* ha obtenido información de las fuentes de datos que reseñan, normalizándola y actualizándola diariamente según el calendario y ritmo de publicación del *Ministerio de Sanidad*, reseñando además sus notas aclaratorias. Son accesibles desde el repositorio *GitHub*:

<https://github.com/datadista/datasets/tree/master/COVID%2019>.

Están siendo utilizados por 39 proyectos, ya que obtener los datos tras la normalización que realiza *Datadista* asegura mayor éxito en la representación de la información.

Para la construcción del portal se ha utilizado el programa estadístico *R* (*R: The R Project for Statistical Computing, 1993*). Su librería *Shiny* ha permitido elaborar la plataforma de visualización interactiva a través de la librería *argonDash*, una adaptación a *R* de las plantillas de *Creative Tim*.

<https://www.creative-tim.com>

Para las visualizaciones interactivas se han utilizado dos librerías:

- *Ramcharts*, basada en el paquete *amcharts.js* escrito en *javascript*;
- *Plotly*.



Figura 5. Página de inicio de covidDATA-19 <https://indicame.upv.es/coviddata-19>

Para el mapa interactivo se ha utilizado la librería *Leaflet* y para las tablas interactivas la librería *Reactable*, basada a su vez en una librería escrita en *javascript* llamada *React Table*.

Los cálculos de los datos diarios y de su variación porcentual se realiza a partir de los datos acumulados de *Datadista*. La página principal presenta el resumen de datos (a modo de cuadro de mando) y la posibilidad de segmentar la información por comunidades autónomas. Permite también la selección del tipo de incidencia hospitalaria: datos de fallecimientos, altas e ingresos en UCI (en su primera versión se incluían también los datos de hospitalizados). En la columna central se pueden ver los datos numéricos: el último dato y la evolución.

El menú se complementa con tres opciones más:

- en la primera se presenta un mapa nacional coroplético en el que se pueden ver los últimos datos y la evolución de los últimos días;
- en la segunda se presentan varias comparativas: evolución de los datos agrupados por incidencia, distribución por edad y sexo, el ranking general por CCAA, y otros datos como son la distribución de mascarillas en España. Todos estos datos pueden filtrarse y descargarse en formato gráfico (png, jpg, svg y pdf) o en formato de tabla (csv, json, xls) para su reutilización;
- en la última opción del menú se presentan los datos en forma de tabla. Actualmente se presenta la tabla de acumulados con posibilidad de ver la evolución diaria, así como la tabla de variación porcentual diaria (figura 6).

Creemos que tiene dos puntos fuertes con respecto a otros portales:

- posibilidad de descargar sólo una parte del gráfico (en los cuatro formatos mencionados anteriormente) así como de escribir anotaciones sobre éste;
- descarga de los datos en formato *json*.

En el momento de escribir este artículo mostraba algunas series interrumpidas, consecuencia de las incidencias en los datos proporcionados por el *Ministerio* y descritas en la sección anterior. Esto, lejos de ser una debilidad, pensamos que es una fortaleza porque una de las utilidades que tiene el portal es la detección de fallos o inconsistencias en los datos (véase por ejemplo la figura 3).

Por último estamos estudiando enriquecer las visualizaciones con una distribución por provincias (*Providencialdata19, 2020*) o ampliar su cobertura con datos de otros países, así como proporcionar otro tipo de comparativas como, por ejemplo, los casos con la población.

Si bien el objetivo de este trabajo no es valorar la situación actual de la puesta a disposición de los datos sobre Covid-19 por parte de las CCAA a los ciudadanos, presentamos una tabla final que describe la situación de los datos abiertos para cada CCAA. Esta tabla consta de una plantilla con los portales web de las CCAA con su correspondiente enlace de acceso, en el que se ha analizado si cada uno contiene o no datos recogidos sobre la Covid-19. Además, se realiza un comenta-



Figura 6. Panel con las otras opciones: fallecidos, distribución de mascarillas y mapas.

rio sobre la información recogida, de dónde proceden los datos (sanidad, prensa, otros), si hay actualizaciones diarias o no, qué tipo de datos se ofrecen (acumulados, casos activos, fallecidos...), si presentan visualización de datos y gráficas, si los datos están accesibles y se pueden descargar. Otras informaciones relevantes relacionadas son: porcentaje de tests realizados, situación de los enfermos en residencias de ancianos, ERTE, etc.

Hay escasez de fuentes oficiales fácilmente reutilizables

4. Lecciones aprendidas

Utilizar una fuente de datos integral, actualizada y fiable como *Datadista* marca una diferencia significativa en una situación de urgencia, ya que ahorra efectuar la limpieza de datos. En el momento de escribir este artículo (abril de 2020), no existe otra fuente oficial tan amigable para reutilizar, si bien desde el *Instituto de Salud Carlos III (ISCIII)* pueden descargarse los acumulados por comunidad autónoma.

Este proyecto ha puesto de manifiesto una vez más la dificultad de trabajar con datos abiertos directamente desde fuentes oficiales. Algunos datos no están disponibles, otros no cuentan con formatos fácilmente reutilizables, y existen múltiples formas de recoger los datos, obligando a procesarlos, lo que en momentos de urgencia produce retrasos.

Pero quizá el problema más relevante es la no determinación desde el inicio de la recogida de información: qué es lo que abarca cada uno de los indicadores de la muestra, y cómo y cuándo debe ser recogido.

En conclusión, estamos asistiendo a una evolución desde la *datafilia* detectada en un principio a una *datafobia* en la que se descalifica de forma absoluta la validez de los datos. Obviamente el método de contabilización del número de contagiados y fallecidos hace cambiar los resultados (Llaneras, 2020). Hasta ahora se registraban como casos Covid-19 quienes habían sido diagnosticados con un test. Dado que no se efectúa al 100% de la población, mientras no contemos con datos sobre el número de pruebas, la dimensión de la pandemia y su letalidad es desconocida.

5. Referencias

Anastassopoulou, Cleo; Russo, Lucia; Tsakris, Athanasios; Siettos, Constantinos (2020). "Data-based analysis, modelling and forecasting of the Covid-19 outbreak". *PLoS one*, v. 15, n. 3.

<https://doi.org/10.1371/journal.pone.0230405>

Bai, Yan; Yao, Lingsheng; Wei, Tao; Tian, Fei; Jin, Dong-Yan; Chen, Lijuan; Wang, Meiyun (2020). "Presumed asymptomatic carrier transmission of Covid-19". *Jama*, v. 323, n. 14, pp. 1406-1407.

<https://doi.org/10.1001/jama.2020.2565>

Calabuig, José-Manuel; García-Raffi, Luis-Miguel; García-Valiente, Albert; Sánchez-Pérez, Enrique-Alfonso (2020). "Kaplan-Meier type survival curves for Covid-19: a health data based decision-making tool". Preprint.

<https://arxiv.org/abs/2005.06032>

Cobarsí-Morales, Josep (2020). "Covid-19: fuentes de información cuantitativa". *Anuario ThinkEPI*, v. 14, e14d02.

<https://doi.org/10.3145/thinkepi.2020.e14d02>

Dong, Ensheng; Du, Hongru; Gardner, Lauren (2020). "An interactive web-based dashboard to track Covid-19 in real time". *The lancet infectious diseases*, v. 20.

[https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1)

España (2020). "Real decreto 463/2020, de 14 de marzo, por el que se declara el estado de alarma para la gestión de la situación de crisis sanitaria ocasionada por el Covid-19". *BOE*, n. 67, 14 marzo.

<https://www.boe.es/buscar/act.php?id=BOE-A-2020-3692>

Huang, Chaolin; Wang, Yeming; Li, Xingwang; Ren, Lili; Zhao, Jianping; Hu, Yi; Zhang, Li; Fan, Guohui; Zu, Jiuyang; Gu, Ziaoying; Cheng, Zhenshun; Yu, Ting; Xia, Jiaan; Wei, Yuan; Wu, Wenjuan; Xie, Xuelei; Yin, Wen; Li, Hui; Liu, Min; Xiao, Yan; Gao, Hong; Guo, Li; Xie, Jungang; Wang, Guangfa; Jiang, Rongmeng; Gao, Zhancheng; Jin, Qi; Wang, Jianwei; Cao, Bin (2020). "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China". *The lancet*, v. 395, n. 10223, pp. 497-506.

[https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)

Instituto de Salud Carlos III (2020). Covid-19.

<https://cnecovid.isciii.es/covid19/#documentaci%C3%B3n-y-datos>

Llaneras, Kiko (2020). *Los datos del #coronavirus: la serie se ha roto: sabemos los muertos de hoy pero no los de anteayer.*

<https://www.youtube.com/watch?v=6lqrKH1So0I&feature=youtu.be>

Madjid, Mohammad; Safavi-Naeini, Payam; Solomon, Scott D.; Vardeny, Orly (2020). "Potential effects of coronaviruses on the cardiovascular system: A review". *Jama cardiology*, 27 March.

<https://doi.org/10.1001/jamacardio.2020.1286>

Providencialdata19 (2020). Covid19.
<https://code.montera34.com:4443/numeroteca/covid19>

R: The R Project for Statistical Computing (s.f.).
<https://www.r-project.org>

Rothe, Camila; Schunk, Mirjam; Sothmann, Peter; Bretzel, Gisela; Froeschl, Guenter; Wallrauch, Claudia; Zimmer, Thorbjörn; Thiel, Verena; Janke, Christian; Guggemos, Wolfgang; Seilmaier, Michael; Drosten, Christian; Vollmar, Patrick; Zwirgmaier, Katrin; Zange, Sabine; Wölfel, Roman; Hoelscher, Michael (2020). "Transmission of 2019-NCOV infection from an asymptomatic contact in Germany". *New England journal of medicine*, v. 382, n. 10, pp. 970-971.
<https://doi.org/10.1056/NEJMc2001468>

Salas-Rueda, Ricardo-Adán; Salas-Rueda, Rodrigo-David (2019). "Análisis sobre el uso de la red social Facebook en el proceso de enseñanza-aprendizaje por medio de la ciencia de datos". *Revista de comunicación de la SEECI*, n. 50, pp. 1-26.
<http://doi.org/10.15198/seeci.2019.50.1-26>

Shiny (s.f.).
<https://shiny.rstudio.com>

The New York Times (2020). "Coronavirus map: Tracking the global outbreak". *The New York Times*, 30 June.
<https://www.nytimes.com/interactive/2020/world/coronavirus-maps.html>

Torres-Salinas, Daniel (2020). "Ritmo de crecimiento diario de la producción científica sobre Covid-19. Análisis en bases de datos y repositorios en acceso abierto". *El profesional de la información*, v. 29, n. 2, e290215.
<https://doi.org/10.3145/epi.2020.mar.15>

Torres-Salinas, Daniel; Robinson, Nicolás; Castillo-Valdivieso, Pedro A. (2020 en evaluación). *Open access and altmetrics in the pandemic age: Forecast analysis on COVID-19 related literature*.
<https://doi.org/10.5281/zenodo.3763140>

WHO (2020). *Coronavirus press conference 11 February*. World Health Organization.
https://www.who.int/docs/default-source/coronaviruse/transcripts/who-audio-emergencias-coronavirus-full-press-conference-11feb2020-final.pdf?sfvrsn=e2019136_2

Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, I-Jsbrand-Jan; Appleton, Gabrielle; Axton, Myles; Baak, Arie; Blomberg, Niklas; Boiten, Jan-Willem; Bonino-da-Silva-Santos, Luiz; Bourne, Philip T.; Bouwman, Jildau; Brookes, Anthony J.; Clark, Tim; Crosas, Mercè; Dillo, Ingrid; Dumon, Olivier; Edmunds, Scott; Evelo, Chris T.; Finkers, Richard; González-Beltrán, Alejandra; Gray, Alasdair J. G.; Groth, Paul; Goble, Carole; Grethe, Jeffrey S.; Heringa, Jaap; Hoen, Peter A. C.; Hooft, Rob; Kuhn, Tobias; Kok, Ruben; Kok, Joost; Lusher, Scott J.; Martone, Maryann E.; Mons, Albert; Packer, Abel L.; Persson, Bengt; Rocca-Serra, Philippe; Roos, Marco; Van-Schaik, Rene; Sansone, Susanna-Assunta; Schultes, Erik; Sengstag, Thierry; Slater, Ted; Strawn, George; Swertz, Morris A.; Thompson, Mark; Van-der-Lei, Johan; Van-Mulligen, Erik; Velterop, Jan; Waagmeester, Andra; Wittenburg, Peter; Wolstencroft, Katherine; Zhao, Jun; Mons, Barend (2016). "The FAIR guiding principles for scientific data management and stewardship". *Scientific data*, v. 3, n. 1, 160018.
<https://10.1038/sdata.2016.18>

Worldometer (2020). *About Worldometer Covid-19 data*.
<https://www.worldometers.info/coronavirus/about/#sources>

6. Anexo

Portal web	URL	Contiene datos	No contiene datos	Comentarios
Datos abiertos Junta de Andalucía	https://www.juntadeandalucia.es/datosabiertos/portal/catalogo.html		X	Al buscar Covid-19 se muestran 28 <i>datasets</i> , ninguno es de sanidad, son de contratos menores, educación y otros. Sólo 3 corresponden a Salud y no tienen relación con la Covid-19, son sobre donaciones de sangre en Córdoba
Portal Covid de Andalucía	https://portalandaluciacovid19.es		X	Sólo dispone de documentos técnicos y comunicados oficiales en html en https://www.juntadeandalucia.es/organismos/saludyfamilias/areas/salud-vida/paginas/coronavirus-comunicados-anteriores.html
Datos abiertos Gobierno de Aragón	https://opendata.aragon.es	X		A la pregunta Covid-19 en el buscador salen 0 resultados. Pero sí aparecen por coronavirus. Todos los <i>datasets</i> son de actualización diaria
Gobierno de Aragón. Departamento de Sanidad. Dir. Gral. Salud Pública	https://www.aragon.es/documentos/20127/38742837/casos_coronavirus_aragon.xlsx	X		Contiene un xls con todos los datos acumulados de actualización diaria sobre coronavirus

Datos abiertos del Principado de Asturias	https://cutt.ly/BsO6W0c		X	No aparece información al buscar Covid-19 y coronavirus. Únicamente información sobre las medidas económicas y laborales
Portal Coronavirus Principado de Asturias	https://coronavirus.asturias.es/inicio	X		Se pueden descargar los datos del portal. Contiene información acumulada de total de casos, activos, curaciones, fallecimientos, hospitalizados y en UCI
Portal de Datos abiertos del Gobierno Balear	http://www.caib.es/sites/opendata-caib/ca/inici_home/?campa=yes		x	La búsqueda en el catálogo de datos tanto para Covid-19 como para coronavirus sale 0
Portal Covid-19 Gobierno Balear	http://www.caib.es/sites/covid-19/es/covid-19		X	Sólo aparecen datos en algunos de los comunicados de prensa
Datos abiertos Gobierno de Canarias	https://opendata.gobiernodecanarias.org/opendata			Al preguntar en el catálogo por Covid-19 salen como resultado 51 datasets pero ninguno tiene relación con la Covid-19. Por coronavirus no sale ningún resultado
Portal Covid-19 Gobierno de Canarias	https://www3.gobiernodecanarias.org/sanidad/scs/tematica.jsp?idCarpeta=e01092c2-7d66-11ea-871d-cb574c2473a4	X		Disponen de un mapa que visualiza los casos y la evolución con gráficas, pero los datos no son descargables. La web lee la visualización de datos. Lo hace el Instituto Cartográfico de Canarias. https://grafcan1.maps.arcgis.com/apps/opsdashboard/index.html#/156eddd4d6fa4ff1987468d1fd70efb6
Gobierno de Cantabria	https://www.cantabria.es/sanidad			Dispone de los datos en*: https://www.scsalud.es/coronavirus https://experience.arcgis.com/experience/a2c5986c293d4b9eaa35a7f573506973 Dispone del histórico en*: https://www.scsalud.es/documents/2162705/9255280/2020_covid19_historico.xls/68dc2943-bb0a-80c5-a4ee-34522f6d1330 y de municipios en*: https://www.scsalud.es/documents/2162705/9255280/2020_covid19_historico.xls/68dc2943-bb0a-80c5-a4ee-34522f6d1330
Junta de Castilla-León	https://www.jcyl.es	X		Dispone información sobre la Covid-19 y visualización. Los datos se encuentran en el portal de datos abiertos
Junta de Castilla La Mancha	https://www.castillalamancha.es		X	La información se suministra a partir de notas de prensa en html* https://sanidad.castillalamancha.es/ciudadanos/enfermedades-infecciosas/coronavirus/actualizacion-de-casos-en-castilla-la-mancha
Portal de datos abiertos de Castilla La Mancha	https://datosabiertos.castillalamancha.es/search/type/dataset		X	
Generalitat de Catalunya	http://aquas.gencat.cat/ca/actualitat/ultimes-dades-coronavirus	X		Se encuentran las visualizaciones de datos de: 1. Mapa interactivo de casos positivos por municipio. http://aquas.gencat.cat/ca/actualitat/ultimes-dades-coronavirus/mapa-per-municipis 2. Mapa interactivo de casos positivos por ABS http://aquas.gencat.cat/.content/IntegradorServeis/mapa_covid/atlas.html 3. Análisis de los últimos datos* https://app.powerbi.com/view?r=eyJrIjoizTKyNTcwNjgtNTQ4Yi00ZTg0LTk1OTctNzY3ZGEzNWE4OTIwIiwidCI6IjNiOTQyNzRjLWQzMGUtNDNiYy04YzA2LWZmNzI1MzY3NmZlYyYlMmMiOj9 4. Mapa de casos Stop Covid-19 https://aquas-gencat.carto.com/u/aquas-gencat-admin/builder/9396196e-da82-4850-a85c-7f24b3e82388/embed
Comunidad Valenciana	http://coronavirus.san.gva.es/es/inicio	X		Disponen de un portal del Coronavirus con datos y su visualización. El problema es que únicamente aparecen los del día de la visualización y las gráficas de evolución
Junta de Extremadura	http://www.juntaex.es/con05	X		Aparecen datos diarios sobre la Covid-19 en noticias
Portal de datos abiertos de la Junta de Extremadura	http://gobiernoabierto.juntaex.es/datos		X	No aparecen datasets al preguntar en el catálogo de datos por Covid-19 o coronavirus. Tampoco en los datasets de la Consejería de salud

Xunta de Galicia	https://coronavirus.sergas.gal	X	Ofrece visualizaciones de los datos diarios, pero no se pueden extraer los datos. 1. 17-04-2020: Taxa de letalidade https://saladecomunicacion.sergas.gal/Paginas/DetalleNova.aspx?idioma=es&idNova=10379 2. 17-04-2020: Galicia rexistra un total de 6.142 casos infectados por coronavirus https://saladecomunicacion.sergas.gal/Paginas/DetalleNova.aspx?idioma=es&idNova=10378	
Portal de datos abiertos de la Xunta de Galicia	https://abertos.xunta.gal/busca-de-datos		X	En la consulta a su catálogo de datos no figura información cuando se interroga por Covid-19 ni coronavirus
Región de Murcia	http://www.murciasalud.es/pagina.php?id=458440	X		Aparecen datos diarios publicados como imagen
Portal de datos abiertos de la Región de Murcia	https://transparencia.carm.es/portal-regional-de-datos-abiertos		X	Al preguntar por Covid-19 aparecen muchos <i>dataset</i> s pero ninguno relacionado con los datos de los afectados por el coronavirus. Aparecen menos resultados al buscar por coronavirus, pero igualmente los resultados no tienen información sobre los datos de los afectados
Comunidad Foral de Navarra	https://www.navarra.es/es/web/coronavirus/inicio	X		Aparecen los datos diarios como imagen. Son los únicos que tienen un documento explicativo de los datos que se publican: https://www.navarra.es/documents/3860554/3937821/Preguntas+y+respuestas+sobre+la+detección+de+COVID-19.pdf/7af10c20-6c29-33f2-647c-3e79c-5269f27?t=1585415567898
Portal de datos abiertos de Navarra	https://gobiernoabierto.navarra.es/es/coronavirus		X	Dispone de distintos <i>dataset</i> s: 1. Datos evolución Covid-19 https://gobiernoabierto.navarra.es/open-data/datos/datos-evolutivos-covid-19 2. Positivos Covid-19 por zonas básicas de salud https://gobiernoabierto.navarra.es/open-data/datos/datos-zonas-basicas-salud-covid-19 3. Evolución de la epidemia: Radiografía Covid-19 Navarra https://gobiernoabierto.navarra.es/sites/default/files/radiografia_covid_navarra.pdf Mapas: Número de casos Tasa por cada 1.000 habitantes https://tinyurl.com/t78ocex https://tinyurl.com/wquqgl7
Gobierno de Euskadi	https://www.euskadi.eus/nuevo-coronavirus-covid-19	X		Dispone de un mapa interactivo del coronavirus
Portal de datos abiertos de Euskadi	https://opendata.euskadi.eus/catalogo/-/evolucion-del-coronavirus-covid-19-en-euskadi	X		Tienen un <i>dataset</i> de actualización diaria creado el 21 de marzo. En la documentación asociada sí detallan algunas cuestiones de la metodología de recogida
Gobierno de La Rioja	https://web.larioja.org	X		Aparecen los datos del día y visualización de la evolución de los distintos datos. No figura la metodología
Portal de datos abiertos de La Rioja	https://www.larioja.org/tecnologia-comunica/es/proyectos-dgtic/open-data		X	No figuran datos cuando se pregunta por Covid-19 o coronavirus. El resultado que aparece no tiene relación con las cifras del virus

<p>Portal de datos abiertos de Catalunya</p>	<p>http://governobert.gencat.cat/ca/dades_obertes/dades-obertes-covid-19</p>	<p>X</p>	<p>Dispone de los datasets:</p> <ol style="list-style-type: none"> 1. Incidència de la Covid-19 a Catalunya https://analisi.transparenciacatalunya.cat/Salut/Incidencia-de-la-COVID-19-a-Catalunya/623z-r97q 2. Registre de test de Covid-19 per sexe i municipi https://analisi.transparenciacatalunya.cat/Salut/Registre-de-test-de-COVID-19-realitzats-a-Cataluny/jj6z-iyrp 3. Registre de test de Covid-19 per sexe i edat https://analisi.transparenciacatalunya.cat/ca/Salut/Registre-de-test-de-COVID-19-realitzats-a-Cataluny/qwj8-xpvk 4. Registre de test de COVID-19 per sexe i ABS 5. Evolució diària dels Expedients de Regulació Temporal d'Ocupació https://analisi.transparenciacatalunya.cat/Treball/Evolucio-di-ria-dels-Expedients-de-Regulaci-Tempora/atmi-6snp 6. Expedients de Regulació Temporal d'Ocupació (ERTOs) per comarca https://analisi.transparenciacatalunya.cat/Treball/Expedients-de-Regulaci%C3%B3-Temporal-d-Ocupaci%C3%B3-ERTOs-6fpi-ycak 7. Expedients de Regulació Temporal d'Ocupació (ERTOs) per divisió econòmica https://analisi.transparenciacatalunya.cat/ca/Treball/Expedients-de-Regulaci%C3%B3-Temporal-d-Ocupaci%C3%B3-ERTOs-hu7t-2x76
<p>Portal de datos abiertos de la Comunidad Valenciana</p>	<p>https://dadesobertes.gva.es/va/dataset/dades-covid-19-comunitat-valenciana-ultimes-dades*</p>	<p>X</p>	<p>Actualización diaria del dataset con los datos pero no figuran los datos de días anteriores. Los datasets que publican diariamente son:</p> <ol style="list-style-type: none"> 1. Porcentaje y número de casos por rango de edad y sexo 2020-04-16* https://dadesobertes.gva.es/es/dataset/dades-covid-19-comunitat-valenciana-ultimes-dades/resource/ed8fb727-b186-4ad8-bdcc-0df1fe515467 2. Número de casos y tasa por cada 100.000 habitantes por departamentos de salud 2020-04-16* https://dadesobertes.gva.es/es/dataset/dades-covid-19-comunitat-valenciana-ultimes-dades/resource/5173518e-3cfd-408e-9348-fc32a83b8dcf 3. Porcentaje y número de fallecidos por rango de edad y sexo 2020-04-16* https://dadesobertes.gva.es/es/dataset/dades-covid-19-comunitat-valenciana-ultimes-dades/resource/8ef3b49a-58bd-4a17-befe-ba7ecb09d808 4. Número de fallecidos y tasa por cada 100.000 habitantes por departamentos de salud 2020-04-16* https://dadesobertes.gva.es/es/dataset/dades-covid-19-comunitat-valenciana-ultimes-dades/resource/05cc07f1-ef0e-4c22-85e3-ef7facafa1e2

<p>Portal de datos abiertos de la Junta de Castilla-León</p>	<p>https:// analisis.datosabiertos.jcyl.es/explore/?sort=modified</p>	<p>X</p>	<p>Dispone información sobre:</p> <ol style="list-style-type: none"> 1. Situación de solicitudes de ertes coronavirus https:// analisis.datosabiertos.jcyl.es/explore/dataset/situacion-de-solicitudes-de-ertes-coronavirus/table/?disjunctive.provincia 2. Ocupación de camas en hospitales https:// analisis.datosabiertos.jcyl.es/explore/dataset/ocupacion-de-camas-en-hospitales 3. Tasa de enfermos acumulados por zonas básicas de salud https:// analisis.datosabiertos.jcyl.es/explore/dataset/tasa-enfermos-acumulados-por-areas-de-salud/table 4. Situación enfermos en residencias por coronavirus https:// analisis.datosabiertos.jcyl.es/explore/dataset/situacion-enfermos-en-residencias-por-coronavirus/table 5. Porcentaje de test de coronavirus realizados a profesionales por categorías https:// analisis.datosabiertos.jcyl.es/explore/dataset/porcentaje-de-test-de-coronavirus-realizados-a-profesionales-por-categorias/table 6. Porcentaje de test de coronavirus realizados a profesionales por provincias https:// analisis.datosabiertos.jcyl.es/explore/dataset/porcentaje-de-test-de-coronavirus-realizados-a-profesionales-por-provincias/table 7. Situación de hospitalizados por coronavirus en Castilla y León https:// analisis.datosabiertos.jcyl.es/explore/dataset/situacion-de-hospitalizados-por-coronavirus-en-castilla-y-leon/table 8. Prevalencia coronavirus https:// analisis.datosabiertos.jcyl.es/explore/dataset/prevalencia-coronavirus/table 9. Situación enfermos por coronavirus detectados en atención primaria por tramos de edad y sexo https:// analisis.datosabiertos.jcyl.es/explore/dataset/situacion-enfermos-por-coronavirus-detectados-en-atencion-primaria-por-tramos-d0/table 10. Situación enfermos por coronavirus detectados en atención primaria https:// analisis.datosabiertos.jcyl.es/explore/dataset/situacion-afectados-por-coronavirus-en-atencion-primaria/table/?disjunctive.provincia 11. Situación epidemiológica coronavirus (Covid-19) en Castilla y León https:// analisis.datosabiertos.jcyl.es/explore/dataset/situacion-epidemiologica-coronavirus-en-castilla-y-leon/table/?disjunctive.provincia 12. Situación profesionales sanitarios afectados por coronavirus https:// analisis.datosabiertos.jcyl.es/explore/dataset/situacion-profesionales-sanitarios-por-coronavirus/table/?disjunctive.provincia&disjunctive.categoria 13. Tasa mortalidad por centros de salud https:// analisis.datosabiertos.jcyl.es/explore/dataset/tasa-mortalidad-por-centros-de-salud/table
--	--	----------	--

* Estos urls han sido desactivados al finalizar el estado de alarma, pero se incluyen en esta tabla porque dan muestra del interés de las instituciones en publicar datos de la Covid-19.