

Datos abiertos sobre Covid-19 en las Comunidades Autónomas españolas: reutilización en estudios de epidemiología espacial

Open data on Covid-19 in the Spanish autonomous communities: reutilization in spatial epidemiology studies

José-Antonio Salvador-Oliván; Severino Escolano-Utrilla

Note: This article can be read in its English original version on:
<https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/87030>

Cómo citar este artículo.

Este artículo es una traducción. Por favor cite el original inglés:

Salvador-Oliván, José-Antonio; Escolano-Utrilla, Severino (2022). "Open data on Covid-19 in the Spanish autonomous communities: reutilization in spatial epidemiology studies". *Profesional de la información*, v. 31, n. 4, e310410.

<https://doi.org/10.3145/epi.2022.jul.10>

Artículo recibido el 23-06-2022
Aceptación definitiva: 08-07-2022



José-Antonio Salvador-Oliván ✉
<https://orcid.org/0000-0001-8568-3098>

Universidad de Zaragoza
Facultad de Medicina
Departamento de Ciencias de la
Documentación e Historia de la Ciencia
Pedro Cerbuna, 12. 50009 Zaragoza, España
jaso@unizar.es



Severino Escolano-Utrilla
<https://orcid.org/0000-0002-3489-0692>

Universidad de Zaragoza
Facultad de Filosofía y Letras
Departamento de Geografía y Ordenación
del Territorio
San Juan Bosco, 7. 50009 Zaragoza, España
severino@unizar.es

Resumen

La pandemia por Covid-19 ha puesto de manifiesto la necesidad de los gobiernos y de las administraciones sanitarias de todos los niveles de disponer de un registro de datos abiertos que facilite la toma de decisiones en la planificación y gestión de recursos sanitarios y proporcione información a la ciudadanía sobre la evolución de la epidemia. El concepto de "datos abiertos" comprende la posibilidad de reutilización por terceros. Por otro lado, el espacio y el tiempo son dimensiones básicas para estructurar e interpretar los datos de las variables referidas propiamente al estado de salud de las personas. De aquí que el objetivo principal de este estudio es evaluar si los ficheros de datos de las Comunidades Autónomas sobre Covid-19 son reutilizables para analizar la evolución de la enfermedad en unidades de análisis espaciales y temporales básicas a nivel autonómico y de nación. Para ello, de los portales de las Comunidades Autónomas se seleccionaron los ficheros de datos abiertos que contenían el número de casos diagnosticados de Covid-19 distribuidos en unidades espaciales básicas sanitarias o administrativas y en unidades temporales. Se evaluó principalmente la presencia de variables relacionadas con la infección, demográficas y temporales, así como el formato de descarga y metadatos. Se analizó también si la estructura de los ficheros era homogénea y adecuada para aplicar técnicas de análisis espacial. Los resultados revelan falta de normalización en la recogida de datos tanto en unidades espaciales como temporales y ausencia o ambigüedad en el significado de las variables motivada por la falta de metadatos. También se ha constatado una estructura no adecuada en ficheros de siete Comunidades Autónomas, lo que requeriría un proceso posterior de los datos para poder reutilizarlos y aplicar técnicas de análisis y modelado espacial, tanto a la hora de llevar a cabo un análisis global como al comparar patrones de evolución entre distintas regiones.

Palabras clave

Datos abiertos; Covid-19; Pandemias; Coronavirus; Reutilización de datos abiertos; Información de salud; Portales de datos de salud; Comunidades autónomas; España; Epidemiología espacial; Geolocalización.

Abstract

The Covid-19 pandemic has highlighted the need for governments and health administrations at all levels to have an open data registry that facilitates decision-making in the planning and management of health resources and provides information to citizens on the evolution of the epidemic. The concept of “open data” includes the possibility of reutilization by third parties. Space and time are basic dimensions used to structure and interpret the data of the variables that refer to the health status of the people themselves. Hence, the main objective of this study is to evaluate whether the autonomous communities’ data files regarding Covid-19 are reusable to analyze the evolution of the disease in basic spatial and temporal analysis units at the regional and national levels. To this end, open data files containing the number of diagnosed cases of Covid-19 distributed in basic health or administrative spatial units and temporal units were selected from the portals of the Spanish autonomous communities. The presence of infection-related, demographic, and temporal variables, as well as the download format and metadata, were mainly evaluated. Whether the structure of the files was homogeneous and adequate for the application of spatial analysis techniques was also analyzed. The results reveal a lack of standardization in the collection of data in both spatial and temporal units and an absence of, or ambiguity in, the meaning of the variables owing to a lack of metadata. An inadequate structure was also found in the files of seven autonomous communities, which would require subsequent processing of the data to enable their reuse and the application of analysis and spatial modeling techniques, both when carrying out global analyses and when comparing patterns of evolution between different regions.

Keywords

Open data; Covid-19; Pandemics; Coronavirus; Reutilization of open data; Health information; Health data portals; Autonomous communities; Spain; Spatial epidemiology; Geolocation.

Financiación

Este trabajo ha sido financiado por el Proyecto I+D+i concedido por el *Ministerio de Ciencia e Innovación*, titulado “La propagación espaciotemporal de la Covid-19 (Sars-CoV-2) en España y su relación multiescala con los patrones de movilidad cotidiana y vulnerabilidad sociodemográfica” (Ref. PID2020-115904RB-I00).

Conflicto de interés

Los autores declaran que no hay conflicto de interés.

1. Introducción

El término “datos abiertos” apareció por primera vez en 1995 en un documento de una agencia científica estadounidense que trataba de la divulgación de datos geofísicos y ambientales, y en el que los autores promovían un intercambio completo y abierto de información científica entre diferentes países. Años más tarde, en 2007, un grupo de activistas del software abierto estableció el concepto de datos abiertos, definiendo los datos públicos como un bien común basado en los principios de apertura, participación y colaboración (**Chignard**, 2013).

El *Open Government Working Group* recomendó ocho principios relacionados con el acceso y uso de datos públicos. Estos principios establecen que los datos públicos son abiertos cuando se reúnen las siguientes propiedades: 1) completos; 2) primarios; 3) oportunos; 4) accesibles; 5) procesables por máquina; 6) no discriminatorios; 7) no propietarios; y 8) sin licencia (**Dawes**, 2010).

Es importante distinguir disponibilidad de datos de usabilidad de datos, aunque el objetivo de una política de datos abiertos debería ser conseguir las dos propiedades. Ya que la mayoría de las políticas de datos abiertos vienen impuestas por organismos legislativos sin guías específicas sobre cómo deberían publicarse los datos de manera amigable para su uso público, los gestores frecuentemente se centran más en la disponibilidad de datos que en su usabilidad, lo que crea una brecha entre calidad de datos y calidad del sistema que incapacita a los usuarios utilizar todo el potencial de los datos (**Park; Gil-García**, 2022).

Atendiendo al concepto de datos abiertos, un portal de datos abiertos debería disponer de datos que pudieran ser reutilizados por profesionales, otros usuarios de datos y público en general. Entre las características mínimas que deben reunir los portales para que sus datos puedan ser utilizados por profesionales se encuentran: disponer de un mecanismo de actualización, utilizar un sistema de gestión de datos y disponibilidad de API, algo de lo que carecían más de la mitad de los portales en España en 2021 (**Abella; Ortiz-de-Urbina-Criado; De-Pablos-Heredero**, 2022).

Para evaluar el grado de reutilización de datos abiertos se han creado diferentes modelos. *Meloda 5* es una métrica basada en diversas dimensiones referidas a licencia legal, acceso a los datos, estándares técnicos, normalización, contenido relacionado con la geolocalización, frecuencia de actualización, difusión y reputación (**Abella; Ortiz-de-Urbina-Criado; De-Pablos-Heredero**, 2019). Otro modelo más sencillo fue desarrollado por Tim Berners-Lee. En él se mide la calidad de los datos mediante la asignación de cinco estrellas según su disponibilidad (niveles 1, 2 y 3) y capacidad de reutilización (niveles 4 y 5) (**Berners-Lee**, 2009). Otros requisitos que deben reunir los datos para su reutilización son: integridad (completos), accesibilidad y visibilidad, facilidad de uso y de comprensión, actualización, validez y utilidad, calidad y granularidad (datos primarios) (**Lourenço**, 2015).

La expansión y efectos de la pandemia por Covid-19, y el desconocimiento inicial sobre muchos aspectos de la infección, puso de manifiesto el valor de disponer de datos adecuados y disponibles públicamente a una escala espacial y temporal adecuadas para poder aprovecharlos con fines de planificación y modelado (**Gardner et al.**, 2021).

La *Organización Mundial de la Salud (OMS)*, el *European Centre for Disease Prevention and Control (ECDC)* y la *Johns Hopkins University*, desarrollaron rápidamente conjuntos de datos para analizar las tendencias de la pandemia y evaluar los riesgos de propagación y contagio en varios países y regiones de todo el mundo (**Ashofteh; Bravo**, 2020). A nivel internacional se ha proporcionado una gran cantidad de datos en formato abierto para ser utilizados o reutilizados por investigadores con la finalidad de predecir la propagación de la enfermedad e identificar poblaciones o áreas potencialmente vulnerables (**Ball**, 2020). Son muchos los recursos y repositorios de datos abiertos relacionados con la Covid-19 creados por gobiernos e instituciones (**Hu et al.**, 2020; **Pecoraro; Luzzi**, 2021). **Alamo et al.** (2020) presentan los principales conjuntos de datos disponibles en internet relacionados con la Covid-19 y enumeran las instituciones más relevantes que presentan información actualizada acompañadas de una descripción de sus características.

En España son las Comunidades Autónomas (en adelante CCAA) las que gestionan la información sanitaria en sus territorios y tienen la capacidad y responsabilidad de publicar datos relativos a la pandemia en sus portales de datos abiertos específicos sobre coronavirus (**Díez-Garrido; Melero-Lázaro**, 2022). En *covidDATA-19* se recogen datos de casos diarios confirmados, fallecidos, casos UCI y altas por CCAA a partir de la fuente de datos *Datadista* (**Ferrer-Sapena et al.**, 2020). También el *Gobierno de España*, a través del *Instituto de Salud Carlos III*, publica datos abiertos sobre el número de casos por técnica diagnóstica por provincia de residencia, así como el número de hospitalizaciones, ingresos en UCI y defunciones por sexo, edad y provincia (*ISCIII*, 2022).

A pesar del enorme esfuerzo que supone la creación de estos portales, los datos se utilizan para describir la incidencia en formato de informes, visualizaciones y descargas solo por provincias. Sin embargo, son insuficientes para aplicar técnicas con las que construir modelos espaciales y temporales que permitan pronosticar la evolución de la pandemia y la eficacia de medidas de contención en unidades básicas más precisas, homogéneas y comparables, como municipios o Zonas Básicas de Salud (ZBS). Por ello, es necesario acudir a los portales de datos abiertos de las CCAA para averiguar si es posible reutilizar datos de Covid-19 a los que se puedan aplicar técnicas de análisis espacial que permitan identificar patrones espaciotemporales para modelar o pronosticar la propagación e incidencia en unidades espaciales con resolución de municipios o ZBS (**Escolano-Utrilla; Salvador-Oliván**, 2022).

Algunos estudios de los portales de datos abiertos sobre coronavirus en las CCAA han evaluado qué datos contienen y el formato de descarga (**Díez-Garrido; Melero-Lázaro**, 2022; **Martín-Fernández et al.**, 2021), pero ninguno de ellos ha evaluado si es posible su reutilización en investigaciones de disciplinas especializadas y, más concretamente, en estudios que busquen aplicar técnicas de análisis espacial.

El desarrollo de modelos espaciotemporales sobre la incidencia, riesgo y propagación requiere la existencia de datos que cumplan con determinadas características en sus contenidos y forma de expresión. Por ello, el objetivo de este estudio es averiguar si los portales de las comunidades autónomas contienen ficheros de datos sobre Covid-19 reutilizables para analizar la evolución de la pandemia en unidades de análisis espaciales y temporales básicas, tanto a escala de comunidad autónoma como nacional.

2. Epidemiología espacial y geografía de la salud

Los estudios epidemiológicos tienen como objetivo explicar la distribución de la enfermedad en una población mediante la identificación de las personas en situación de riesgo, el momento de aparición de la enfermedad y los lugares donde se encuentran (**Rothman; Greenland; Lash**, 2011). El lugar geográfico se ha convertido en un elemento esencial para comprender la dinámica y evolución de una epidemia y, como resultado, surgió la epidemiología espacial como un enfoque novedoso para la comprensión y control de las epidemias actuales.

Por otra parte, la geografía de la salud aborda también el estudio de la distribución espacial y la incidencia de las enfermedades pero con un enfoque territorial; es decir, el territorio como espacio físico y como construcción se sitúa en el centro de la investigación. De aquí que los dominios temáticos de la geografía de la salud sean más amplios que los de la epidemiología espacial, pues incluyen también contenidos de los sistemas de salud (distribución de equipamientos, recursos y servicios sanitarios, etc), demográficos, ambientales y de planificación (**Souris**, 2019).

Los estudios de epidemiología espacial y de geografía de la salud han contribuido notablemente al conocimiento de los procesos de propagación de la pandemia por Covid-19 y de los factores que influyen en su desigual incidencia y riesgo social y territorial. Entre otros aspectos, la investigación desarrollada en este campo ha mejorado el conocimiento de los patrones espaciotemporales de propagación geográfica y de las rutas de transmisión de la pandemia (**Franch-Pardo et al.**, 2021), ha permitido identificar puntos críticos locales (**Hu et al.**, 2020), calcular el número de nuevos contagios por caso (número R) así como valorar la incidencia de múltiples factores socioeconómicos, demográficos, ambientales, y de riesgo de infección (**Elliott; Wartenberg**, 2004; **Kang et al.**, 2020; **Xu; Kraemer**, 2020; **Kobayashi et al.**, 2021; **Paez**, 2021).

Aunque la epidemiología espacial y la geografía de la salud tienen diferentes ámbitos y objetivos de indagación, comparan numerosos métodos, técnicas, datos y medios de análisis espacial y espaciotemporal, por cuanto el espacio geográ-

fico se sitúa en el centro de ambas disciplinas (Souris, 2019). Para que los datos sobre eventos de salud –cambios en el estado fisiológico o sanitario de los individuos de una población– puedan ser utilizados en estas disciplinas, además de la información temática deben incluir también la localización.

La mayor parte de los estudios de epidemiología espacial y geografía de la salud aplican procedimientos de análisis espacial implementados en programas de sistemas de información geográfica (SIG) o similares capaces de analizar información geográfica digital (Kirby; Delmelle; Elbert, 2017). El modelo de datos más extendido en los SIG utiliza puntos, líneas, polígonos, píxeles y otros objetos digitales para representar la localización y forma de los fenómenos y objetos reales (personas, territorios, variables ambientales, etc.) que estudian. La descripción de la localización suele seguir algún estándar de datos y metadatos, y se conecta con la información temática, generalmente contenida en tablas relacionales, mediante un geocódigo. La información espacial es considerada como una infraestructura pública, mantenida por instituciones dedicadas específicamente a la creación y difusión de información geográfica (Ministerio de Transportes, Movilidad y Agenda Urbana, s/f).

Los datos sobre eventos de salud se toman en individuos, de los que se puede registrar su localización. Los datos individuales permiten reconstruir trayectorias y comportamientos con gran detalle, en contextos locales, pero solo están disponibles de forma excepcional. La forma habitual de publicar los datos individuales sobre eventos de salud consiste en agregar la información en unidades espaciales, con la mayor resolución posible. A su vez, estas unidades pueden fusionarse en otras mayores –lo que produce pérdida de resolución espacial– en función de los objetivos de cada estudio o del ámbito espacial de la planificación sanitaria. Desde el punto de vista práctico es conveniente que los datos estén referidos a unidades espaciales específicas, como Zonas Básicas de Salud (ZBS), Áreas de Salud u otras, o administrativas (secciones censales, distritos, municipios) representadas por su geocódigo y cuya información espacial esté actualizada y accesible. Esta opción facilita, además, el uso de otra información (demográfica, socioeconómica, ambiental) referida a estas mismas unidades espaciales.

3. Método

Fuentes de datos

Se emplearon como unidades de análisis los portales de datos abiertos sobre coronavirus de las Comunidades Autónomas españolas. Se han seleccionado estos portales porque su objetivo es divulgar conjuntos de datos en formatos reutilizables para que otros ciudadanos e investigadores los puedan emplear (García-García; Curto-Rodríguez, 2018).

Se localizaron mediante una búsqueda en Google con los términos “datos abiertos” coronavirus y el nombre de la comunidad autónoma, y se accedió a la web del portal de datos abiertos de cada comunidad autónoma durante el mes de mayo de 2022.

Debido a que estos portales contienen múltiples ficheros de datos sobre la Covid-19, se seleccionaron aquellos que cumplieran con los siguientes criterios de inclusión, adecuados para alcanzar el objetivo de este estudio:

- Ficheros con datos sobre casos diagnosticados de Covid-19 en unidades espaciales básicas sanitarias (ZBS, distritos o centros sanitarios) o administrativas (municipios, comarcas, provincias).
- Ficheros con datos sobre casos diagnosticados de Covid-19 en unidades temporales (datos recogidos en días, semanas o en los distintos períodos u olas de la pandemia).

La razón para disponer de información con elevada resolución espacial y temporal se debe a su valor estratégico para la gestión de los recursos sanitarios y para la adopción de medidas de control de la pandemia a escala local, que es a la que operan los sistemas sanitarios.

Cuando alguna CCAA no disponía de ficheros que satisficieran los dos criterios de inclusión, se seleccionaron aquellos que cumplieran con alguno de ellos.

Se aplicaron los siguientes criterios de exclusión:

- Ficheros con datos globales para toda la CCAA, no desagregados en unidades espaciales.
- Ficheros exclusivamente con datos acumulados finales, o solo con cifras de los últimos 7 o 14 días.

VARIABLES Y EXTRACCIÓN DE DATOS

Basándonos en la métrica *Meloda 5* para la reutilización de conjuntos de datos abiertos (Abella; Ortiz-de-Urbina-Criado; De-Pablos-Heredero, 2019) y en variables empleadas para obtener modelos epidemiológicos (Alamo et al., 2020), se seleccionaron las variables necesarias para analizar y representar espacialmente la evolución espaciotemporal y situación de la pandemia en unidades espaciales básicas.

Las variables se clasificaron en seis categorías:

- (1) datos de identificación de los portales;
- (2) variables geográficas, de capital importancia ya que el lugar donde vive un individuo es considerado como un potencial determinante de morbilidad y mortalidad (Diez-Roux, 2001);
- (3) datos sobre la infección por Covid-19;
- (4) variables temporales;

- (5) variables demográficas, importantes también para conocer la densidad y estructuras demográficas y socioeconómicas de la población ya que se han reconocido como un factor de riesgo en la transmisión de la enfermedad y en la tasa de incidencia (Priyadarsini; Suresh, 2020; Sy; White; Nichols, 2021); y
- (6) otras variables como el formato de descarga o la existencia de metadatos que faciliten la interpretación de los datos que hay en los portales.

De los ficheros descargados de las CCAA que cumplían los criterios de inclusión, se evaluó la presencia o no de las variables y se recogieron los datos en un fichero *Excel*. Se evaluó también la estructura de los ficheros de datos abiertos para averiguar si eran o no apropiados para aplicar procedimientos de estadística espacial. Prácticamente todos los sistemas de información geográfica utilizan para el análisis de datos espaciotemporales una estructura de datos en “lista” que tiene al menos tres columnas:

- en una, se registran los códigos de las unidades espaciales (que se repiten por cada medida de tiempo);
- en otra, se guardan las medidas de la variable tiempo (expresadas en unidades y formato adecuados);
- otra, contiene los valores de la variable de interés (fenómeno estudiado).

Los datos en forma de tabla, con las unidades espaciales en las filas y las unidades temporales en las columnas se pueden utilizar para análisis sincrónicos y también se pueden transformar en formato “lista” mediante una operación de trasposición. Por esta razón, se clasificaron los ficheros seleccionados en tres tipos:

- estructura adecuada si cada fila contenía una unidad territorial identificable (municipio, distrito sanitario, ZBS) y su número de casos según una distribución temporal (días o semanas); (
- estructura inadecuada, pero se puede reutilizar con tratamiento de datos;
- estructura inadecuada por falta de datos esenciales: no es posible reutilizarlos.

El listado completo de variables seleccionadas y el fichero *Excel* están disponibles en *Open Science Farmework (OSF)*: <https://osf.io/r6byj>

4. Resultados

4.1. Características generales

Todas las CCAA, excepto Extremadura y las Ciudades Autónomas de Ceuta y Melilla, tienen ficheros de datos abiertos relativos a la pandemia del coronavirus. De las 16, en solo 9 consta que el tipo de licencia para el uso de los ficheros de datos es *Creative Commons* siempre que se reconozca la autoría (*Attribution* o *By*), y solo una, Castilla la Mancha, especifica que debe compartirse igual que la licencia original.

La mitad de las CCAA con datos abiertos presentan metadatos que describen el significado y contenido de las variables de los ficheros. Son Andalucía, Baleares, Canarias, Castilla y León, Cataluña, Comunidad Valenciana, Navarra y País Vasco.

Todas las CCAA publican datos en formato legible por máquina y no propietario, por lo que no existen barreras para reutilizar los datos disponibles. El formato csv está presente en todas, y tan solo Baleares y Cataluña ofrecen sus datos también en formato rdf.

Tabla 1. Características generales de los ficheros de datos abiertos de las CCAA

CCAA	Licencia legal	Metadatos	Formato de datos								
			rdf	csv	json	PC-axis	ods	xml	rss	tsv	Excel
Andalucía		✓		✓	✓	✓	✓				✓
Aragón				✓							✓
Asturias	CC BY			✓							
Baleares		✓	✓	✓	✓			✓	✓	✓	
Canarias	CC BY	✓		✓							
Cantabria	CC BY			✓							✓
Castilla-La Mancha				✓							
Castilla y León	CC BY-SA	✓		✓	✓			✓			✓
Cataluña	CC BY	✓	✓	✓				✓	✓	✓	
C. Valenciana		✓		✓	✓			✓		✓	
Galicia	CC BY			✓							
La Rioja				✓	✓			✓			✓
Madrid	CC BY			✓	✓						
Murcia	CC BY			✓							
Navarra		✓		✓	✓			✓		✓	✓
País Vasco		✓		✓	✓		*	✓			✓

4.2. Datos geográficos y temporales de los ficheros

En la tabla 2 se resumen las principales características del fichero de cada CCAA que presenta mayor nivel de granularidad espacial y temporal, declaración de los casos y completitud de los datos (período de tiempo cubierto).

Tabla 2. Características del fichero de datos sobre Covid-19 con mayor granularidad y completitud en cada Comunidad Autónoma

CCAA	Unidades espaciales	Casos diagnosticados	Resolución temporal	Período
Andalucía	Distritos sanitarios	PDIA	Semanas	Completo
Aragón	ZBS	Totales	Días	Completo
Asturias	Áreas sanitarias	Totales	Semanas	Hasta 12/2021
Baleares	ZBS	Totales	Días	Completo
Canarias	Municipios	Totales	Días	01/2021-3/2022
Cantabria	Municipios	Totales	No	Fase aguda
	No	Totales	Días	Fase aguda
Castilla-La Mancha	Provincias	Totales	No	Fase aguda
	No	PDIA	Días	Fase aguda
	Municipios	PDIA (> 60)	Dos semanas	3/2022-
Castilla y León	ZBS	Totales, PDIA y PCR	Días	Completo
Cataluña	Área Básica de Salud	Totales, PDIA	Días	Completo
C. Valenciana	Deptos de Salud	PDIA	Días	Completo
Galicia	Áreas sanitarias	Totales	Días	Completo
La Rioja	ZBS	Totales	No	Completo
Madrid	ZBS	Totales	Días/semanas	Completo
Murcia	ZBS	Totales	Días	Fase aguda
Navarra	ZBS	Totales	Días	Fase aguda
País Vasco	ZBS	Totales	Días	13/05/2020-actualidad

Nota: PDIA = Pruebas diagnósticas de infección activa

Andalucía proporciona el número de casos diagnosticados por PDIA (pruebas diagnósticas de infección activa - PCR y tests de antígenos) por distritos sanitarios, con sus nombres y códigos INE y por semanas desde marzo de 2020 hasta la marzo de 2022. A partir de marzo de 2022, están recogidos los datos de mayores de 60 años en otro fichero (que incluye desde el inicio de la pandemia). Por lo tanto, los dos ficheros cubren todo el período. Los ficheros contienen la descripción de las variables incluidas.

Para Aragón se puede descargar un fichero de datos con el número total de casos distribuidos diariamente y por Zonas Básicas de Salud desde el principio de la pandemia hasta la actualidad. A partir de mediados de mayo de 2020 los casos confirmados se diagnostican a partir de PCR y tests de antígenos, con lo que se produce un aumento con respecto a días anteriores que solo se diagnosticaban con PCR. El inconveniente es que hay que descargarse un fichero por cada día hasta marzo de 2022, y a partir de esta fecha, los ficheros se descargan por semanas, recogiendo los datos diarios de esa semana. No existen metadatos.

Asturias dispone de un fichero con el número total de casos (no se indica cómo se diagnostican) distribuidos semanalmente por áreas sanitarias desde el principio de la pandemia hasta diciembre de 2021.

Baleares tiene un fichero de datos con el número total de casos diagnosticados por PCR y tests de antígenos distribuidos por ZBS (con sus geocódigos) y días, según grupos de edad y sexo, desde el principio de la pandemia hasta la actualidad. Contiene metadatos.

Canarias tiene un fichero con el número de casos y fallecidos, según grupos de edad y sexo, distribuidos por municipios (tanto de residencia como de asignación) y días, desde enero de 2021 hasta marzo de 2022. Por lo tanto, no cubre todo el período de la fase aguda de pandemia. Cada registro corresponde a un caso, del que se indica el sexo y grupo de edad.

Cantabria no dispone de un fichero único con datos desagregados en unidades espaciales y temporales. Mantiene dos ficheros; en uno de ellos está recogido el número total de casos acumulado final por municipios; en otro, está el número de casos totales por días pero para toda la comunidad. En ambos casos, el período de tiempo abarca hasta marzo de 2022.

Para Castilla-La Mancha tampoco hay un fichero único desagregado en unidades espaciales y temporales. Existen dos ficheros; en uno, aparece el número total de casos acumulado final y fallecidos por provincias; en otro, se incluye el número de casos confirmados por PDIA distribuidos diariamente para toda la comunidad autónoma. Ambos cubren la fase aguda de pandemia. Conviene señalar que, desde marzo de 2022, han creado otro archivo de datos con el número de casos en las dos últimas semanas en personas mayores de 60 años distribuidos por municipios.

Castilla y León proporciona un fichero con el número total de casos y los diagnosticados por PDIA y PCR, distribuidos por días y centros de salud (con sus códigos y coordenadas GPS), así como el número total de pruebas realizadas, cubriendo desde el inicio de la pandemia hasta la actualidad.

Para Cataluña hay cuatro ficheros con datos desagregados en unidades espaciales y temporales. El de mayor utilidad es el que contiene el número total de casos y los diagnosticados por PDIA (PCR y test de antígenos) distribuidos por días en Áreas Básicas de Salud (ABS) durante todo el período de pandemia hasta la actualidad. Contiene metadatos.

La Comunidad Valenciana dispone de un fichero con casos confirmados por PDIA, según grupos de edad y sexo, distribuidos diariamente por Departamentos de Salud durante toda la pandemia hasta la actualidad.

En Galicia el fichero contiene todos los casos diagnosticados por PDIA y autotests distribuidos por Áreas Sanitarias y días durante todo el período de pandemia hasta la actualidad.

La Rioja dispone de varios ficheros de datos que no se descargan, otros llevan a una página de error 404 o contienen datos imprecisos (por ejemplo, en número de casos aparece <10 en lugar del valor exacto). El fichero de datos más útil es el que recoge el número de casos positivos acumulados por ZBS desglosados por PCR, antígenos y CLIA (Test de ELISA/anticuerpos), con el número total de pruebas realizadas. No se especifica el inicio del período, que llega hasta la actualidad.

Para Madrid, hay dos ficheros. Uno de ellos contiene todos los casos diagnosticados distribuidos por ZBS y días hasta el 1 de julio de 2020; a partir del 2 de julio, los datos se distribuyen por semanas hasta marzo de 2022. A partir de marzo, solo se registran los casos en personas mayores de 60 años distribuidos por ZBS y días hasta la actualidad.

En Murcia existen dos ficheros (uno hasta el 30 de noviembre de 2020, y otro desde el 1 de diciembre de 2020 hasta marzo de 2022) con el número total de casos distribuidos por ZBS y días.

Navarra dispone de un fichero con el número total de casos acumulados distribuidos por ZBS y días desde el principio de la pandemia hasta marzo de 2022. Sería necesario procesar los datos para averiguar el número de casos nuevos por día y ZBS.

En el País Vasco hay un fichero que recoge el número total de casos distribuidos por ZBS con sus códigos y por días, desde el 15 de mayo de 2020 hasta la actualidad. En este fichero, las columnas son los días, a diferencia del resto de las CCAA, que ocupan las filas. Contiene el número de casos diagnosticados por PCR, tests de antígenos y tests de anticuerpos, así como el número de pruebas realizadas, número de fallecidos, hospitalizados y casos en la UCI, por grupos de edad y sexo.

El nivel más alto de desagregación territorial se ha observado en 9 CCAA que proporcionan datos a escala de ZBS, incluida Cataluña, aunque su denominación sea diferente (Área Básica de Salud). En Andalucía, Asturias, Comunidad Valenciana y Galicia los datos están agregados en unidades de menor resolución, con denominaciones propias de cada comunidad. Tres CCAA (Canarias, Cantabria y Castilla-La Mancha) no proporcionan datos por áreas de gestión de servicios sanitarios sino en unidades administrativas como municipios o provincias (figura 1).

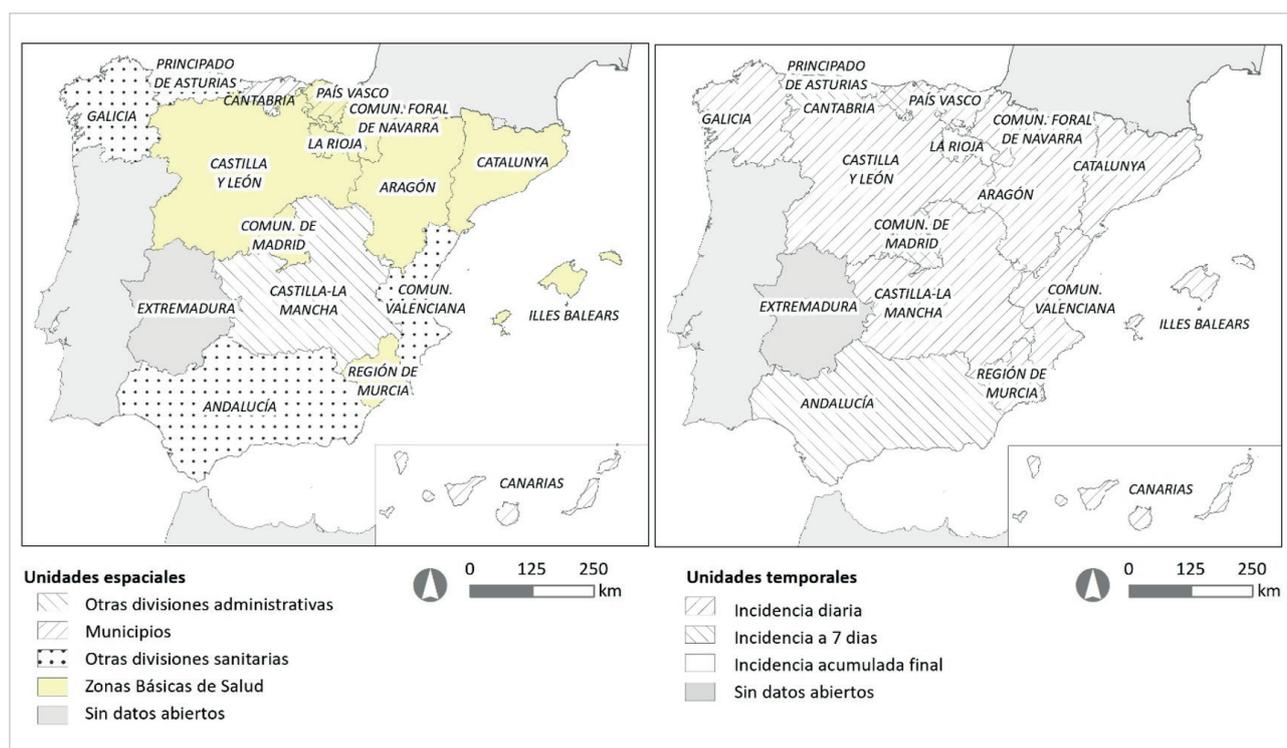


Figura 1. Representación de unidades espaciales (izquierda) y temporales (derecha) de los registros de datos abiertos de la Covid-19 por comunidades autónomas

Respecto a la distribución temporal, el número de casos aparece registrado por días en todas las CCAA salvo en Andalucía y Asturias, que se consignan por semanas. Es preciso mencionar que en Castilla la Mancha los datos diarios no se presentan por unidades espaciales y, cuando aparecen por semanas, los datos son los casos totales

“ Datos abiertos son aquellos que no solo se comparten sino que se pueden reutilizar por terceras partes para llevar a cabo estudios en sus disciplinas científicas ”

acumulados en las dos últimas semanas solo en mayores de 60 años. Igualmente ocurre en Cantabria, donde los datos diarios se refieren a toda la CCAA, y cuando los datos están por municipios, son datos acumulados finales. En La Rioja, los datos tampoco están distribuidos en unidades temporales y registran los casos acumulados finales por ZBS (figura 1).

4.3. Estructura de los ficheros

En la tabla 3 se muestra la estructura de los ficheros de las CCAA, así como su valoración y requisitos de tratamiento de los datos respecto a su uso en aplicaciones de técnicas de análisis espacial y si es necesario un proceso posterior de los datos.

Tabla 3. Posibilidad de reutilización de los ficheros de datos de las CCAA para aplicar técnicas de análisis espacial

Comunidad autónoma	Estructura	Requiere proceso de los datos
Andalucía	×	Información redundante
Aragón	×	Hay que descargar un fichero por día o semana. Hay que incluir la variable DIA en el fichero global
Asturias	×	Las Áreas Básicas de Salud hay que pasarlas a filas
Baleares	✓	
Canarias	×	Cada fila corresponde a un solo caso. Habría que sumar el número de casos en un día
Cantabria	××	Los datos no están registrados en unidades espaciales
Castilla-La Mancha	××	Faltan datos
Castilla y León	✓	
Cataluña	✓	
C. Valenciana	×	Los Departamentos de Salud hay que pasarlos a filas (transposición de los datos)
Galicia	✓	
La Rioja	××	Faltan datos
Madrid	✓	
Murcia	✓	
Navarra	×	Presenta el número de casos acumulados para cada día y ZBS. Habría que restar para averiguar número de casos diarios y sumarlos para agregarlos en semanas u otros períodos
País Vasco	×	Pasar a filas los días (están en columnas)

(✓ Estructura adecuada; × Requieren proceso de los datos; ×× No se pueden reutilizar)

Solo 6 CCAA disponen de un fichero con la estructura adecuada, donde cada fila corresponde al número de casos por ZBS y día u otro período. En 5 CCAA, sería necesario procesar los datos para adaptar su estructura. De ellas, el archivo correspondiente a Andalucía es el que menos dificultades presenta, ya que el proceso consistiría simplemente en eliminar filas con información redundante (aparecen los distritos sanitarios con el número de casos y también las provincias, que contienen el número total de casos de los distritos que pertenecen a cada provincia). Para Aragón, hay que descargar tantos ficheros como días, y a partir de marzo de 2022, tantos como semanas; en cada fichero, hay una hoja con varias tablas, por lo que hay que copiar y pegar en un fichero único los datos del número de casos y ZBS, y habría que incluir como variable el día al que pertenece el fichero.

En el fichero de Canarias, cada fila corresponde a un único caso por municipio con la fecha de diagnóstico, por lo que habría que sumar el número de casos en el mismo día y en el mismo municipio. Sin embargo, en Navarra se presenta el número de casos acumulados para cada día y ZBS, con lo que habría que obtener el número de casos diarios en cada ZBS por día restando, al día actual, el valor de casos acumulado del día anterior.

En el fichero de Asturias, hay 8 columnas que corresponden al número de casos en cada una de las 8 Áreas de Salud; sería necesario convertir las 8 columnas a una sola con los nombres de todas las Áreas de Salud, repitiéndose para cada día. En la Comunidad Valenciana ocurre algo parecido, ya que hay 24 columnas, cada una de ellas recoge el número de casos por Departamento de Salud, con lo que habría que convertir todas las columnas a una sola que recogiera los nombres de los Departamentos. Y, por último, en el País Vasco ocurre la situación contraria, ya que cada fila es una ZBS y las columnas son los días, por lo que en este caso, para la representación y análisis temporal con sistemas de información geográfica, también hay que pasar todos los días a una única columna con múltiples filas, tantas como días.

En los ficheros de Cantabria, Castilla la Mancha y la Rioja no se podrían utilizar los datos ya que faltan variables, bien las relacionadas con el territorio o con la distribución temporal de los casos.

5. Discusión

Los resultados de nuestro estudio han revelado la falta de estándares y heterogeneidad de los modelos de datos compartidos entre las CCAA. Principalmente afectan a:

- procedimiento de diagnóstico de los casos (PCR, tests de antígenos, tests de anticuerpos);
- registro temporal del número de casos (diarios, semanales, por olas, o acumulados);
- unidades espaciales (ZBS, Distritos Sanitarios, Áreas de Salud, municipios, comarcas, provincias);
- período de tiempo cubierto (desde el inicio de la pandemia hasta la actualidad, fase aguda de pandemia, desde marzo de 2022 con la nueva estrategia de vigilancia y control);

además de otros datos que complementan el análisis epidemiológico de la pandemia, como número de fallecidos, ingresos en UCI, número de pruebas realizadas, población total o por grupos de edad y sexo, etc.

En general, los problemas detectados en nuestro estudio en los portales de datos abiertos son consistentes con los obtenidos por **Gardner et al.** (2021) en la recopilación de datos de Covid-19:

- ambigüedad de las definiciones de los parámetros;
- diferente frecuencia de actualización de variables entre los portales;
- deficiencias en la arquitectura del sistema, que debe ser de naturaleza jerárquica, obteniendo los datos a nivel local (ZBS o municipios) y agregándolos hasta Áreas de Salud o provincias.

Además, se han identificado otros inconvenientes que dificultan la reutilización de los datos y que son coherentes con los señalados por **Alamo et al.** (2020):

- falta de estandarización en la recopilación de datos,
- datos incompletos: al incluir solo los casos confirmados por una prueba de laboratorio, y que atañe también al número de fallecidos;
- falta de precisión en la medida de variables clave: por ejemplo, no se recogen los casos asintomáticos, que desempeñan un papel importante en la transmisión del virus.

Aunque hay una variedad de formatos para descargar los datos, en todas las CCAA se pueden descargar en un formato adecuado para procesarlos posteriormente en una base de datos o en aplicaciones de hojas de cálculo electrónicas como *Excel* y similares (xls, xlsx, csv).

Se discuten a continuación los problemas que plantea la falta de normalización de datos entre CCAA y su deficiente registro en cada una de ellas, centrándonos en las variables nucleares en epidemiología espacial: casos con infección por Covid, localización geográfica y distribución temporal.

5.1. Variables Covid

Uno de los problemas constatados es la ausencia y/o ambigüedad en la definición de estas variables, motivado por la falta de metadatos o descripción insuficiente de su significado. La variable temática clave para el análisis de cualquier epidemia es el número de casos infectados. La *Organización Mundial de la Salud* define caso confirmado a toda aquella persona en la que se ha confirmado la infección mediante prueba de laboratorio (*WHO*, 2022). Este hecho condiciona la estimación real de casos infectados debido a que no todos infectados sintomáticos podían acceder a pruebas sobre todo durante los primeros meses de pandemia y a que los asintomáticos o con síntomas leves no se hacían la prueba.

En la mayoría de los ficheros de las CCAA aparece el número total de casos, sin indicar cómo se han diagnosticado y si se han confirmado a través de todas pruebas diagnósticas (PCR, test de antígenos, test serológicos), alguna de ellas o varias, con lo que el análisis de la evolución de la pandemia a lo largo del tiempo (por aplicar una nueva metodología en la contabilidad de casos) y la comparación entre CCAA hace que sea más compleja.

A partir del 28 marzo de 2022, tras la fase aguda de la pandemia, entró en vigor la nueva *Estrategia de Vigilancia y Control frente a Covid-19* que supone un cambio importante en el método de registro ya que solo se incluyen los casos confirmados por PDIA en población de 60 años o más, personas inmunodeprimidas y mujeres embarazadas (*Ministerio de Sanidad*, 2022). Sin embargo, hay CCAA en las que se siguen registrando a partir de esta fecha todos los casos y en otras solo los mayores de 60 años.

Existe también falta de uniformidad con el número de pruebas de laboratorio, pudiendo referirse al número total de pruebas realizadas o al número de personas examinadas, sin que se describa con precisión el significado de los recuentos (**Alamo et al.**, 2020).

Para llevar a cabo cualquier estudio es necesario saber qué mide una variable y cómo se ha medido. Los ficheros de datos deben proporcionar metadatos de calidad bien estructurados, organizados y detallados que garanticen la interpretación correcta y apropiada de los datos

“ La aplicación de técnicas de análisis espaciotemporal a datos abiertos sobre Covid-19 es de gran utilidad para obtener patrones y modelos geográficos que expliquen la dinámica, transmisión e incidencia de la enfermedad y para identificar potenciales factores de riesgo ”

(**Barcellos; Bernardini; Viterbo**, 2022), de manera que puedan ser entendidos y utilizados por otros investigadores (**Kubler et al.**, 2018; **Wu; Ma; Zhang**, 2021).

En este sentido, los principios de datos FAIR (localizables, accesibles, interoperables y reutilizables) definen las características que los datos, herramientas, vocabularios e infraestructuras deben presentar para ayudar a la reutilización por parte de terceros, garantizando un acceso rápido y apropiado a datos completos y organizados (**Wilkinson et al.**, 2016). También cabe destacar la existencia de la herramienta *Metadata Quality Assessment (MQA)*, desarrollada por el consorcio de *data.europa.eu* para evaluar la calidad de los metadatos de los datos abiertos públicos y cómo mejorarlos (*Publications Office of the European Union*, 2020).

5.2. Distribución temporal

Existe falta de homogeneidad en el período de tiempo registrado en las series de cada CCAA. Aunque en la mayoría los registros se extienden desde el inicio de la pandemia hasta la actualidad, en algunas cubren únicamente la fase aguda (desde febrero-marzo de 2020 hasta marzo de 2022); o solo hasta diciembre de 2021; o desde enero de 2021 hasta marzo de 2022; o desde mayo de 2020. Esta disparidad hace imposible que se pueda llevar a cabo un análisis global y comparar, desde el punto de vista sincrónico, el comportamiento de la pandemia entre CCAA.

La resolución temporal de los casos también es variable entre las CCAA; en algunas la disponibilidad de datos es diaria, en otras semanal o incluso quincenal. Este problema se ha observado principalmente cuando se obtienen datos de múltiples fuentes (**Alamo et al.**, 2020).

Es importante que las fechas de notificación y registro de los casos sean un fiel reflejo de la realidad. El retraso en la comunicación de los casos y la falta de actualización regular de los datos, tardando varios días o un fin de semana, puede afectar a la exactitud y fiabilidad de los datos en las fechas registradas y a cambios en los patrones o tendencias no atribuibles al fenómeno observado. También falta información sobre si la fecha empleada para el registro de los casos en el fichero tiene el mismo significado en todas las CCAA o es diferente, pudiendo hacer referencia a la fecha de inicio de los síntomas, fecha de realización de la prueba diagnóstica o a la fecha de diagnóstico menos tres días (**Martín-Fernández et al.**, 2021).

5.3. Resolución espacial

Otro rasgo de los datos abiertos sobre la Covid-19 es la falta de homogeneidad de las unidades espaciales entre las diferentes CCAA. Uno de los factores identificados que han limitado el uso de técnicas espaciales para detectar patrones de propagación de la pandemia ha sido la falta de disponibilidad de datos expresados en unidades espaciales de elevada resolución y el sesgo de datos de Covid-19, junto con la escasez de datos demográficos a escala grande o con elevada resolución espacial (**Fatima et al.**, 2021). Los resultados de nuestro estudio revelan una falta de normalización en la disponibilidad de los datos en unidades espaciales en diferentes escalas.

El artículo 62 de la Ley 14/1986, de 25 de abril, General de Sanidad (*España*, 1986), señala que

“para conseguir la máxima operatividad y eficacia en el funcionamiento de los servicios a nivel primario, las Áreas de Salud se dividirán en Zonas Básicas de Salud”.

Para llevar a cabo estudios espaciales y temporales sobre la propagación, incidencia, riesgo y otros aspectos de interés sobre la pandemia, que sean comparables entre CCAA y que abarquen todo el territorio nacional, es imprescindible disponer de información expresada en unidades espaciales equivalentes, como las ZBS u otras de tipo administrativo de mayor resolución, que sean lo más estables posible. Y también es una condición necesaria para la gestión eficiente de los recursos sanitarios y para la adopción de medidas de control de la pandemia.

El registro en unidades espaciales de mayor resolución, como las secciones censales o los municipios e, incluso las ZBS, posibilita su agregación, manual o automática, en otras unidades de menor resolución y mayor tamaño (sectores sanitarios u otras) para realizar análisis a otra escala. En cambio, operar en sentido inverso (por ejemplo, de provincias a sectores sanitarios o de ZBS a municipios) no es directo y requiere adoptar determinadas suposiciones sobre la distribución del fenómeno estudiado y aplicar técnicas analíticas complejas.

Para la mayoría de las CCAA se dispone de datos completos sobre el número de casos desglosados por ZBS y días. Solo en Castilla y León se anotan las ZBS con coordenadas GPS, habiéndose descrito como la comunidad con mejores resultados en cuanto a transparencia e información sobre la situación epidemiológica de la Covid-19 y la peor, Castilla-La Mancha (**Díez-Garrido; Melero-Lázaro**, 2022). Esto concuerda con nuestro análisis de resultados ya que esta última contiene el número de casos acumulado final por provincias o el número total de casos por días para toda la comunidad, con lo que no resulta de utilidad para una posterior reutilización de los datos. Además, contiene errores en la fecha, ya que las 559 filas corresponden a todos los meses del año 2022, y la descarga del fichero se realizó en mayo.

“ En los portales de datos abiertos sobre Covid-19 de las Comunidades Autónomas españolas existe una falta de normalización en los datos recogidos y carecen de una estructura homogénea que dificulta su reutilización ”

Resulta difícil trabajar con datos abiertos directamente de fuentes oficiales de las CCAA, ya que algunas no publican datos, se han utilizado múltiples formas de recoger los datos, y no están estructurados de manera uniforme, lo que obliga a procesarlos de manera adecuada (Ferrer-Sapena *et al.*, 2020). Los ficheros de datos abiertos disponibles proceden de diversas fuentes, contienen diferentes variables (nivel de agregación y significado) organizadas y estructuradas de manera distinta (en filas o columnas); estas deficiencias dificultan su uso a nivel global y para análisis sincrónicos y diacrónicos en todo el territorio nacional. Además, la existencia de múltiples ficheros para cada CCAA con datos repetidos en algunos de ellos requiere más tiempo y recursos (de personas y económicos) para crearlos y mantenerlos.

“ No son necesarios tantos ficheros de datos sobre Covid-19 en cada Comunidad Autónoma; es suficiente con uno que registre los principales datos a mayor nivel de resolución temporal y espacial ”

5.4. Reutilización a nivel nacional y propuesta de estructura de fichero de datos

En la actualidad, para poder llevar a cabo estudios de incidencia y propagación de la Covid-19 en el conjunto del territorio nacional, se requiere agrupar los datos de las distintas CCAA en un único fichero. En este fichero no se incluirían las siguientes CCAA: Extremadura, Ceuta y Melilla (al no tener datos abiertos sobre Covid-19); Cantabria, Castilla-La Mancha y La Rioja, al faltar datos espaciales o temporales, y Canarias, ya que solo tiene datos por municipios. En el resto de las CCAA habría que someter los ficheros seleccionados en este estudio a un proceso de reestructuración y de cálculo de los datos ya descrito en el apartado de Resultados, y de normalización de las variables espaciales, que consistiría en agrupar las Áreas de Salud por Distritos Sanitarios; y temporales, agrupando los días en semanas, ya que es el nivel de desagregación que pueden compartir todos los ficheros.

Para poder llevar a cabo estudios de epidemiología espacial a nivel de país, todas las CCAA deberían disponer de datos abiertos en ficheros reutilizables con una estructura homogénea y presencia de metadatos que describieran con precisión el significado de las variables. Las variables esenciales del fichero deben estar relacionadas con:

- infección: número de casos diagnosticados distribuidos según distintas pruebas diagnósticas (PCR, tests de antígenos, tests de anticuerpos), así como el número de pruebas realizadas para confirmar el diagnóstico, además de otras variables como casos hospitalizados, en UCI, fallecimientos;
- localización espacial: unidades de mayor resolución como ZBS, con sus nombres y coordenadas o geocódigos;
- variables temporales: días (fecha completa);
- demográficas: grupos de edad y sexo.

De esta manera, cada fila del fichero correspondería al número de casos diagnosticados (distribuidos por prueba realizada y según grupo de edad y sexo) en una ZBS por día. El fichero cumpliría con características deseables relacionadas con:

- la integridad o completitud de los datos: tanto en lo que se refiere a las variables recogidas como al período de tiempo cubierto;
- exactitud: pruebas de diagnóstico empleadas para la declaración de casos confirmados;
- granularidad: datos proporcionados a una mayor escala de desagregación espacial, por ZBS, y temporal, por días;
- oportunidad: los datos se deberían actualizar diariamente, sin retrasos, para que no afecte de manera negativa al desarrollo de modelos predictivos.

6. Conclusiones

En este artículo se ha abordado uno de los principios fundamentales de los datos abiertos: las posibilidades de reutilización, por parte de terceros, para estudios de epidemiología espacial. Del análisis y evaluación de los datos abiertos relacionados con la Covid-19 disponibles y accesibles en los portales de las CCAA de España se han identificado varios problemas que imposibilitan su reutilización para el desarrollo de modelos espaciotemporales de la infección por coronavirus. De ellos, los más importantes son la inconsistencia de datos, desconocimiento y disparidad de criterios en el significado de algunas variables, unidades básicas de análisis no homogéneas, falta de actualización, series temporales diferentes, y cambio de criterio en el diagnóstico de casos confirmados. Estos problemas impiden comparar la evolución entre las comunidades autónomas.

Disponer de datos abiertos a través de portales es un activo importante, de gran utilidad para múltiples propósitos, pero deben reunir una serie de características que permitan un uso eficiente y, de esta manera, generar valor añadido en investigaciones y crear conocimiento útil en la toma de decisiones. No son necesarios tantos ficheros para cada CCAA; cuantos más haya, más tiempo y esfuerzo hay que dedicar para su actualización. Es suficiente con uno que tenga la misma estructura para todas las CCAA y que recoja todas las variables importantes, lo que facilitará que el mantenimiento sea más eficiente y redundará en una mayor confiabilidad de los datos. Las CCAA registran y publican los datos originales; si estos no son fiables y actuales, su uso por investigadores o como fuentes secundarias de información por otros organismos como el *Ministerio de Sanidad de España*, la *Unión Europea* o la *OMS* resultará completamente inútil.

7. Referencias

- Abella, Alberto; Ortiz-de-Urbina-Criado, Marta; De-Pablos-Heredero, Carmen** (2019). "Meloda 5: A metric to assess open data reusability". *Profesional de la información*, v. 28, n. 6, e280620.
<https://doi.org/10.3145/epi.2019.nov.20>
- Abella, Alberto; Ortiz-de-Urbina-Criado, Marta; De-Pablos-Heredero, Carmen** (2022). "Criteria for the identification of ineffective open data portals: pretender open data portals". *Profesional de la información*, v. 31, n. 1, e310111.
<https://doi.org/10.3145/epi.2022.ene.11>
- Alamo, Teodoro; Reina, Daniel G.; Mammarella, Martina; Abella, Alberto** (2020). "Covid-19: Open-data resources for monitoring, modeling, and forecasting the epidemic". *Electronics*, v. 9, n. 5, pp. 827.
<https://doi.org/10.3390/electronics9050827>
- Ashofteh, Afshin; Bravo, Jorge M.** (2020). "A study on the quality of novel coronavirus (Covid-19) official datasets". *Statistical journal of the IAOS*, v. 36, n. 2, pp. 291-301.
<https://doi.org/10.3233/SJI-200674>
- Ball, William P.** (2020). "Open data, open science and transparency in the time of Covid 19". *Radical statistics newsletter*, n. 127, pp. 4-9.
- Barcellos, Raissa; Bernardini, Flavia; Viterbo, José** (2022). "Towards defining data interpretability in open data portals: Challenges and research opportunities". *Information systems*, v. 106, 101961.
<https://doi.org/10.1016/j.is.2021.101961>
- Berners-Lee, Tim** (2009). *Linked data*.
<https://www.w3.org/DesignIssues/LinkedData.html>
- Chignard, Simon** (2013). "A brief history of open data". *Paris innovation review*.
<http://parisinnovationreview.com/articles-en/a-brief-history-of-open-data>
- Dawes, Sharon S.** (2010). "Stewardship and usefulness: Policy principles for information-based transparency". *Government information quarterly*, v. 27, n. 4, pp. 377-383.
<https://doi.org/10.1016/j.giq.2010.07.001>
- Díez-Garrido, María; Melero-Lázaro, Mónica** (2022). "La transparencia como medicina. La comunicación de la pandemia a través de los portales de datos abiertos de las autonomías". *Estudios sobre el mensaje periodístico*, v. 28, n. 1, pp. 65-80.
<https://doi.org/10.5209/esmp.77172>
- Diez-Roux, Ana V.** (2001). "Investigating neighborhood and area effects on health". *American journal of public health*, v. 91, n. 11, pp. 1783-1789.
<https://doi.org/10.2105/AJPH.91.11.1783>
- Elliott, Paul; Wartenberg, Daniel** (2004). "Spatial epidemiology: Current approaches and future challenges". *Environmental health perspectives*, v. 112, n. 9, pp. 998-1006.
<https://doi.org/10.1289/ehp.6735>
- Escolano-Utrilla, Severino; Salvador-Oliván, José-Antonio** (2022). "Global regularity and local variability of the space-temporal patterns of Covid 19 in Aragón (Spain)". *Boletín de la Asociación de Geógrafos Españoles*, n. 93.
<https://doi.org/10.21138/bage.3276>
- España* (1986). "Ley 4/1986, de 25 de abril, General de Sanidad". *BOE*, n. 102, 29 abril.
<https://www.boe.es/eli/es/l/1986/04/25/14/con>
- Fatima, Munazza; O'Keefe, Kara J.; Wei, Wenjia; Arshad, Sana; Gruebner, Oliver** (2021). "Geospatial analysis of Covid-19: A scoping review". *International journal of environmental research and public health*, v. 18, n. 5, 2336.
<https://doi.org/10.3390/ijerph18052336>
- Ferrer-Sapena, Antonia; Calabuig, José-Manuel; Peset, Fernanda; Sánchez-del-Toro, Isabel** (2020). "Trabajar con datos abiertos en tiempos de alarma: el caso de uso de covidDATA-19". *Profesional de la información*, v. 29, n. 4, e290421.
<https://doi.org/10.3145/epi.2020.jul.21>
- Franch-Pardo, Ivan; Desjardins, Michael R.; Barea-Navarro, Isabel; Cerdà, Artemi** (2021). "A review of GIS methodologies to analyze the dynamics of Covid-19 in the second half of 2020". *Transactions in GIS*, v. 25, n. 5, pp. 2191-2239.
<https://doi.org/10.1111/tgis.12792>
- García-García, Jesús; Curto-Rodríguez, Ricardo** (2018). "Divulgación de información pública de las comunidades autónomas españolas (2013-2017): portal de datos abiertos, portal de transparencia y web institucional". *Profesional de la información*, v. 27, n. 5, pp. 1051-1060.
<https://doi.org/10.3145/epi.2018.sep.09>

- Gardner, Lauren; Ratcliff, Jeremy; Dong, Ensheng; Katz, Aaron** (2021). "A need for open public data standards and sharing in light of Covid-19". *The lancet infectious diseases*, v. 21, n. 4, e80.
[https://doi.org/10.1016/S1473-3099\(20\)30635-6](https://doi.org/10.1016/S1473-3099(20)30635-6)
- Hu, Tao; Guan, Weihe-Wendy; Zhu, Xinyan; Shao, Yuanzheng; Liu, Lingbo; Du, Jing; Liu, Hongqiang; Zhou, Huan; Wang, Jialei; She, Bing; Zhang, Luyao; Li, Zhibin; Wang, Peixiao; Tang, Yicheng; Hou, Ruizhi; Li, Yun; Sha, Dexuan; Yang, Yifan; Lewis, Ben; Kakkar, Devika; Bao, Shuming** (2020). "Building an open resources repository for Covid-19 research". *Data and information management*, v. 4, n. 3, pp. 130-147.
<https://doi.org/10.2478/dim-2020-0012>
- ISCI (2022). *Covid-19 en España*. ISCI Centro Nacional de Epidemiología.
<https://cnecovid.isciii.es/>
- Kang, Dayun; Choi, Hyunho; Kim, Jong-Hun; Choi, Jungsoon** (2020). "Spatial epidemic dynamics of the Covid-19 outbreak in China". *International journal of infectious diseases*, v. 94, pp. 96-102.
<https://doi.org/10.1016/j.ijid.2020.03.076>
- Kirby, Russell S.; Delmelle, Eric; Erberth, Jan M.** (2017). "Advances in spatial epidemiology and geographic information systems". *Annals of epidemiology*, v. 27, n. 1.
<https://doi.org/10.1016/j.annepidem.2016.12.001>
- Kobayashi, Shinji; Falcón, Luis; Fraser, Hamish; Braa, Jorn; Amarakoon, Pamod; Marcelo, Alvin; Paton, Chris** (2021). "Using open source, open data, and civic technology to address the Covid-19 pandemic and infodemic". *Yearbook of medical informatics*, v. 30, n. 1, pp. 38-43.
<https://doi.org/10.1055/s-0041-1726488>
- Kubler, Sylvain; Robert, Jérémy; Neumaier, Sebastian; Umbrich, Jürgen; Le-Traon, Yves** (2018). "Comparison of meta-data quality in open data portals using the Analytic Hierarchy Process". *Government information quarterly*, v. 35, n. 1, pp. 13-29.
<https://doi.org/10.1016/j.giq.2017.11.003>
- Lourenço, Rui-Pedro** (2015). "An analysis of open government portals: A perspective of transparency for accountability". *Government information quarterly*, v. 32, n. 3, pp. 323-332.
<https://doi.org/10.1016/j.giq.2015.05.006>
- Martín-Fernández, Almudena; Alfaro-Latorre, Mercedes; Crespo-Sánchez-Eznarriaga, Belén; Fidalgo-García, Luz** (2021). "Disponibilidad de datos abiertos relacionados con la Covid-19 en los portales web de las comunidades autónomas y del Ministerio de Sanidad". *Revista española de salud pública*, v. 95, e202112190.
<https://dialnet.unirioja.es/servlet/articulo?codigo=8254679>
- Ministerio de Sanidad (2022). *Estrategia de vigilancia y control frente a Covid-19 tras la fase aguda de la pandemia*.
https://www.sanidad.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/documentos/Nueva_estrategia_vigilancia_y_control.pdf
- Ministerio de Transportes, Movilidad y Agenda Urbana. *Infraestructura de Datos Espaciales de España (IDEE)*.
<https://plataforma.idee.es/web/idee/inicio>
- Paez, Antonio** (2021). "Reproducibility of research during Covid-19: Examining the case of population density and the basic reproductive rate from the perspective of spatial analysis". *Geographical analysis*, first online.
<https://doi.org/10.1111/gean.12307>
- Park, Sora; Gil-García, J. Ramón** (2022). "Open data innovation: Visualizations and process redesign as a way to bridge the transparency-accountability gap". *Government information quarterly*, v. 39, n. 1, 101456.
<https://doi.org/10.1016/j.giq.2020.101456>
- Pecoraro, Fabrizio; Luzi, Daniela** (2021). "Open data resources on Covid-19 in six European countries: Issues and opportunities". *International journal of environmental research and public health*, v. 18, n. 19, 10496.
<https://doi.org/10.3390/ijerph181910496>
- Priyadarsini, S. Lakshmi; Suresh, M.** (2020). "Factors influencing the epidemiological characteristics of pandemic Covid-19: A TISM approach". *International journal of healthcare management*, v. 13, n. 2, pp. 89-98.
<https://doi.org/10.1080/20479700.2020.1755804>
- Publications Office of the European Union (2020). *Metadata quality assessment methodology. How data.europa.eu measures the quality of all stored metadata*.
<https://data.europa.eu/mqa/methodology?locale=en>
- Rothman, Kenneth J.; Greenland, Sander; Lash, Timothy L.** (2011). *Modern epidemiology. Third edition*. Philadelphia: Lippincott Williams & Wilkins. ISBN: 978 1 45115920 2

Souris, Marc (2019). *Epidémiologie et géographie: Principes, methods et outils de l'analyse spatiale*. London: ISTE. ISBN: 978 1 784065 73 7

Sy, Karla-Therese L.; White, Laura F.; Nichols, Brooke E. (2021). "Population density and basic reproductive number of Covid-19 across United States counties". *PLoS one*, v. 16, n. 4, pp. 1-11.
<https://doi.org/10.1371/journal.pone.0249271>

WHO (2022). *Public health surveillance for Covid-19: interim guidance*. World Health Organization.
<https://www.who.int/publications/i/item/who-2019-nCoV-surveillanceguidance-2022.1>

Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, Ijsbrand-Jan; Appleton, Gabrielle; Axton, Myles; Baak, Arie; Blomberg, Niklas; Boiten, Jan-Willem; Da-Silva-Santos, Luiz-Bonino; Bourne, Philip E.; Bouwman, Jildau; Brookes, Anthony; 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://doi.org/10.1038/sdata.2016.18>

Wu, Dan; Ma, Le; Zhang, Hui (2021). "Evaluation of open health data portals for Covid-19 from the perspective of the user experience". *The electronic library*, v. 39, n. 2, pp. 296-317.
<https://doi.org/10.1108/EL-01-2021-0011>

Xu, Bo; Kraemer, Moritz U. G. (2020). "Open access epidemiological data from the Covid-19 outbreak". *The lancet infectious diseases*, v. 20, n. 5, pp. 534.
[https://doi.org/10.1016/S1473-3099\(20\)30119-5](https://doi.org/10.1016/S1473-3099(20)30119-5)



Te esperamos en

www.sedic.es

c/Rodríguez San Pedro 2,
oficina 606. 28015 Madrid

Tfno: +34 915 934 059
secretaria@sedic.es

Sociedad Española de Documentación e Información Científica



 <https://twitter.com/SEDIC20>

 <https://www.facebook.com/AsociacionSEDIC>

 <https://www.linkedin.com/groups?home=&gid=5060038>