

DISPONIBILIDAD DE DATOS ABIERTOS RELACIONADOS CON LA COVID-19 EN LOS PORTALES WEB DE LAS COMUNIDADES AUTÓNOMAS Y DEL MINISTERIO DE SANIDAD

Almudena Martín Fernández (1), Mercedes Alfaro Latorre (1), Belén Crespo Sánchez-Eznarriaga (1) y Luz Fidalgo García (1)

(1) Subdirección General de Información Sanitaria. Ministerio de Sanidad. Madrid. España.

Las autoras declaran que no existe ningún conflicto de intereses.

RESUMEN

Fundamentos: Disponer de determinados datos abiertos supone que estos datos sean accesibles y estén disponibles para todo el mundo, sin restricciones legales ni técnicas. El sector público produce una gran variedad de información (social, económica, sanitaria, estadística) que es atractiva, completa y fiable al proceder directamente de fuentes oficiales. El objetivo de este estudio fue mostrar los datos abiertos relacionados con la COVID-19 publicados por las comunidades autónomas (en adelante CC. AA.) y el Ministerio de Sanidad, y comparar estos datos.

Métodos: Se accedió a cada uno de los portales abiertos y se realizó una observación de los datos publicados, analizando desde cuándo se publican y su periodicidad en la actualización. Se utilizó la herramienta de Tim Berners-Lee para la clasificación del grado de reutilización de los datos. En cuanto al tratamiento estadístico, se realizó un análisis de frecuencia, para analizar qué datos se dan en un mayor número de veces.

Resultados: 15 comunidades autónomas (CC. AA.) publicaron datos abiertos relacionados con la COVID-19 en diferentes formatos, así como el Ministerio de Sanidad. Los contenidos que se mostraron en los portales de datos abiertos variaron de una comunidad autónoma a otra, pero todas las CC.AA ofrecieron datos confirmados por PCR. Lo que variaba era la fecha de inicio de recogida de datos y la actualización de los mismos.

Conclusiones: Existe gran heterogeneidad entre las CC. AA., en cuanto a la publicación de la información de casos de COVID-19, en relación a indicadores, fechas, etc. Se pone de manifiesto una gran existencia de datos y en formatos reutilizables, pero hay que ir a muchas fuentes para obtener información sobre la situación epidemiológica a nivel nacional.

Palabras clave: COVID-19, Datos abiertos, Publicación.

ABSTRACT

Availability of open data related to COVID-19 in the web portals of the Spanish regions and the Ministry of Health

Background: Having certain open data means that this data is accessible and available to everyone, without legal or technical restrictions. The public sector produces a wide variety of information (social, economic, health, statistical) that is attractive, complete and reliable, coming directly from official sources. The aim of this study was to show the open data related to COVID-19 published by the regions of Spain and the Ministry of Health, and compare these data.

Methods: Each of the open portals was accessed and an observation of the published data was made, analyzing since when they were published and their periodicity in the update. The Tim Berners-Lee tool was used to classify the degree of reuse of the data. Regarding the statistical treatment, a frequency analysis was carried out, to analyze which data are given in a greater number of times.

Results: 15 regions publish open data related to COVID-19, as well as the Ministry of Health. The contents shown in the open data portals varied from one autonomous community to another, but the regions provided data confirmed by PCR. The data collection start date and the updating of the data varied.

Conclusions: There is an enormous diversity among de regions in terms of the publication of COVID data, in relation to indicators, dates. It shows a great existence of data and in reusable format, but you have to go to many sources to obtain information of the epidemiological situation at a national level.

Key words: COVID-19, Open data, publication.

INTRODUCCIÓN

El brote de neumonía que tuvo lugar en Wuhan, China, se produjo en diciembre de 2019, y nos ha familiarizado a todos con el nuevo coronavirus de 2019 (2019-nCoV). Este nuevo coronavirus, denominado 2019-nCoV, pertenece a la subfamilia *Orthocoronavirinae*, distinta de la del virus causante del síndrome respiratorio de Oriente Medio-coronavirus y de la del coronavirus del síndrome respiratorio agudo severo (SARS-CoV)⁽¹⁾. La Organización Mundial de la Salud (OMS) alertó la emergencia pública el 30 de enero de 2020 y el 11 de marzo la declaró pandemia. El acrónimo COVID-19 (*CO*Rona *VI*rus *D*isease) se acuñó el 11 de febrero⁽²⁾ para desvincularlo de cualquier origen geográfico, animal u otros modos de coronavirus.

La COVID-19, enfermedad causada por el virus SARS-CoV-2, ha cambiado sustancialmente la forma en que la sociedad interactúa con los datos de salud. Se ha convertido en un foco enorme y dominante para el interés del público y de los medios de comunicación y ahora guía una gran parte del esfuerzo de investigación. Los medios de comunicación están dominados por actualizaciones diarias sobre nuevos datos, visualizaciones y discusiones. Al mismo tiempo, muchos investigadores han cambiado o reorientado su trabajo hacia el estudio de la pandemia en curso. A medida que ha aumentado el interés y la preocupación, también ha aumentado nuestra necesidad de información para aprender más e informarnos de cara a la toma de decisiones. Al mismo tiempo, la producción y el uso de datos abiertos y prácticas más amplias de ciencia abierta se ha acelerado a un ritmo increíble. La pandemia actual de COVID-19 plantea preguntas importantes sobre la apertura, el intercambio y el uso de datos, y destaca los desafíos asociados con el uso de datos. Si algo ha puesto de manifiesto la pandemia generada por la COVID-19 es que los esfuerzos colectivos son fundamentales

para hacer frente a una amenaza global de esta magnitud.

Desde la notificación del primer caso de COVID-19 se ha publicado una enorme cantidad de estudios que han intentado aclarar las principales incógnitas acerca de la sintomatología, pruebas de detección, medidas de prevención o tratamiento de esta nueva enfermedad. La pandemia por COVID-19 ha dado lugar a la mayor explosión de publicaciones científicas conocida, donde cada 15 días se duplican las referencias. Una de las respuestas colectivas por parte de las editoriales ante esta avalancha, ha sido la creación de centros de recursos que unifican en una única web y en acceso abierto todo aquello que se va publicando sobre la COVID-19. Muchas de estas investigaciones han estado caracterizadas por la escasa validez externa y la falta de consistencia. La inmediatez por conocer los interrogantes del virus ha generado alguna retracción de artículos científicos, así como cambios diarios sobre la patogenia, la prevención y el tratamiento de la COVID-19. Cabría preguntarse si ante tanta inmediatez es posible garantizar la calidad de las investigaciones publicadas⁽³⁾.

En los últimos diez años, la tecnología de datos ha progresado aún más, lo que permite que los datos desempeñen un papel indispensable en la lucha contra la COVID-19 en la actualidad. El *big data*, el aprendizaje automático y otras tecnologías han dado forma a nuestro mundo médico y han aumentado las oportunidades para combatir y prevenir con éxito los virus, al apoyar a los humanos en la primera línea para determinar la mejor preparación y respuesta⁽⁴⁾.

La rápida propagación, la alta mortalidad, el rango de gravedad y otros factores desconocidos han generado una gran incertidumbre. Esto significa que existe una necesidad urgente de comprender las características del virus y

desarrollar estrategias para reducir su impacto. Para tomar las mejores decisiones en esta situación en desarrollo, necesitamos información.

Se están realizando esfuerzos dispares en todo el mundo, necesarios para llenar el vacío dejado por la falta de una infraestructura establecida para informes en tiempo real y de datos para poder compartir en formato abierto en el curso crisis de salud⁽⁵⁾. Para facilitar el trabajo con estos datos es importante que exista un sistema de informes estandarizado y armonizado para poder recopilar, visualizar y compartir datos.

Los científicos tienen la necesidad de disponer de datos en formatos con los que puedan trabajar. En este punto se muestra como imprescindible la utilización de datos abiertos. El concepto de datos abiertos (*open data*, en inglés) es una filosofía y práctica que persigue que determinados tipos de datos estén disponibles de forma libre para todo el mundo, sin restricciones de derechos de autor, de patentes o de otros mecanismos de control. El objetivo de las iniciativas de datos abiertos es publicar datos gubernamentales de la forma más interoperable y rica, para que ciudadanos, científicos y otras instituciones puedan construir aplicaciones interesantes y realizar análisis profundos con los datos que se publican.

Ha habido una cantidad impresionante de iniciativas inmediatas en respuesta a la COVID-19. Por ejemplo, se han creado repositorios de datos de acceso abierto relacionados con COVID-19⁽⁶⁾; se han desarrollado herramientas de visualización de datos en tiempo real proporcionadas por varios actores (por ejemplo, Universidad John Hopkins, 2020; Roser, Ritchie y Ortiz-Ospina, 2020; OMS, 2020).

Las tecnologías de la información posibilitan el procesamiento de los datos que se muestran en los documentos públicos haciendo que éstos sean más accesibles y permitan la creación

de productos y servicios. Al mismo tiempo, la apertura de los datos públicos es una exigencia de transparencia y presupuesto previo para el efectivo ejercicio de los derechos de participación política en un sistema democrático.

Estos valores encuentran su sitio en el Modelo de Gobierno Abierto que se extiende en nuestro país y en los países de nuestro entorno, y que persigue mejorar el actual modelo de democracia representativa promoviendo que los ciudadanos sean los protagonistas de la definición de las políticas públicas.

Existe una normativa y procedimiento de reutilización, entendiéndose por reutilización, según el artículo 3 de la *Ley 37/2007*⁽⁷⁾, como el uso de documentos que obran en poder de las Administraciones y organismos del sector público, por personas físicas o jurídicas, con fines comerciales o no comerciales, siempre que dicho uso no constituya una actividad administrativa pública.

En junio de 2019, se adoptó la Directiva de la Unión Europea sobre datos abiertos y la reutilización de la información del sector público, también conocida como la *Directiva de datos abiertos*⁽⁸⁾. Con esta directiva, Europa tiene como objetivo alentar a los Estados miembros a facilitar la reutilización de los datos del sector público, incluidos los datos de investigación financiados con fondos públicos, con una restricción legal, técnica y financiera mínima o nula. Además, la Directiva ha iniciado el proceso mediante el cual se pondrán a disposición conjuntos de datos de alto valor para su reutilización. Los Estados miembros tendrán que transponer la Directiva a la legislación nacional para el verano de 2021. Es interesante imaginar cuántos más datos sobre la COVID-19 habrían estado disponibles hoy si la Directiva ya se hubiera transpuesto e implementado en los Estados miembros. De cara a entender algunos conceptos, mencionar que esta directiva define “*formato legible por máquina*”

al formato de archivo estructurado que permita a las aplicaciones informáticas identificar, reconocer y extraer con facilidad datos específicos, incluidas las declaraciones fácticas y su estructura interna.

La información que las diferentes comunidades autónomas están aportando al Ministerio de Sanidad no está siendo del todo homogénea, debido por ejemplo a que los casos confirmados, así como los fallecimientos a causa de la COVID-19 sólo sean aquellos a los que se ha realizado PCR que ha resultado positiva, lo que va a provocar diferencias en función del número de test que se realicen en cada autonomía, también se puede producir una diferencia importante en la información relativa a los casos que han precisado de hospitalización y/o ingreso en UCI. También puede darse el caso que haya regiones que informen del número total de casos que desde el inicio de la epidemia han requerido de ingreso hospitalario y/o UCI, mientras que otras lo que informen sea la prevalencia, personas que están ingresadas a día del informe⁽⁹⁾.

Es necesario hacer este trabajo debido a que la pandemia de la COVID-19 ha llegado en un momento en que la mayoría de las Administraciones, ya exponían sus datos en portales de datos abiertos, y ha puesto de relieve la importancia de disponer de datos sólidos y sistemas de información de salud que proporcionen datos oportunos, fiables y procesables^(10,11).

El presente trabajo tuvo como objetivo principal estudiar los datos publicados, en el marco de la pandemia de coronavirus, en los distintos portales de datos abiertos a nivel autonómico y nacional.

Con esto se pretendió conocer qué datos se publican en cada portal, el período de publicación, la actualización de los datos y los formatos en los que están disponibles.

El propósito de analizar estos datos fue comparar el estado actual de la publicación de datos abiertos relacionados con el coronavirus, y valorar si estos datos son útiles en relación a su disponibilidad y al formato en el que se publican.

MATERIALES Y MÉTODOS

Las unidades de análisis que se utilizaron fueron el Ministerio de Sanidad y las CC. AA. En la obtención de resultados se realizó un análisis de frecuencia, para analizar qué datos se dan en un mayor número de veces.

En relación a la metodología, se localizó y accedió a cada uno de los portales autonómicos de datos abiertos, exceptuando los de las Ciudades Autónomas de Ceuta y Melilla, englobadas en el Instituto Nacional de Gestión Sanitaria (en adelante INGESA), Extremadura y Galicia que no disponían en ese momento de un portal de datos abiertos, y se analizaron los indicadores que se publicaron en estos portales. A nivel nacional, se accedió al portal oficial de Datos abiertos del Gobierno de España (datos.gob.es). El acceso a estos recursos se realizó en el período comprendido entre julio de 2020 y enero de 2021. En el **anexo I** se muestran las *urls* de acceso a los portales consultados.

Para el análisis de los datos se tuvieron en cuenta la *Orden SND/234/2020, de 15 de marzo*⁽¹²⁾, sobre adopción de disposiciones y medidas de contención y remisión de información al Ministerio de Sanidad ante la situación de crisis sanitaria ocasionada por la COVID-19, publicada en marzo de 2020 por el Ministerio de Sanidad.

En el anexo de esta orden se especifican, entre otras informaciones, los datos que las CC. AA., debían remitir al ministerio en cuanto a la situación epidemiológica y a la situación de la capacidad asistencial.

En concreto, se especificó que la información de la situación epidemiológica se remitiera de forma agregada por comunidad autónoma y ciudades autónomas, con los datos que se muestran a continuación:

- Nº de casos confirmados.
- Nº de casos hospitalizados.
- Nº de casos ingresados en camas de críticos (UCI, REA, etc.).
- Nº de casos dados de alta.
- Nº de casos dados de alta por defunción.
- Nº casos en las últimas 24 horas.
- Nº de pruebas diagnósticas (PCR) realizadas.

Se realizó una selección de los datos más comunes publicados en los portales de las CC. AA., y así realizar un análisis de los mismos. Estos datos fueron:

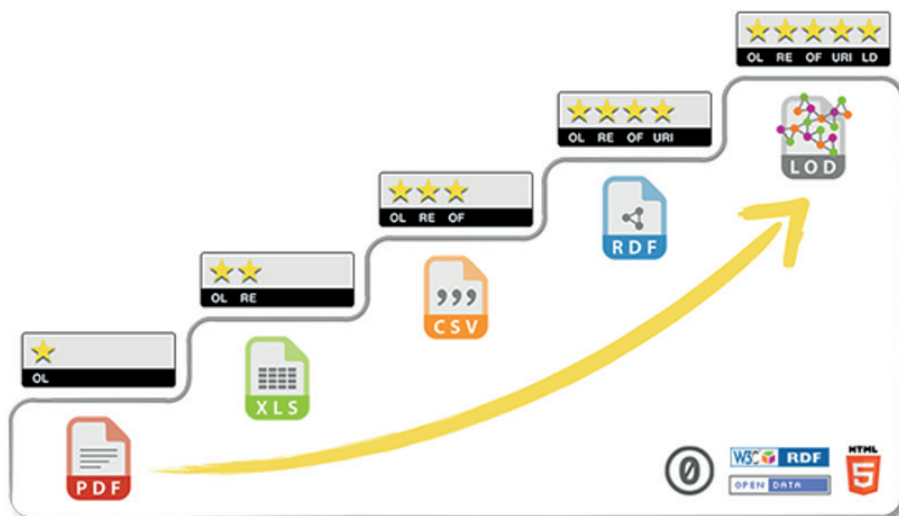
- Fecha de inicio de datos.
- Datos acumulados por provincias.
- Casos confirmados por PCR.
- Datos sobre profesionales de riesgo.
- Datos sobre residencias.
- Datos sobre colegios.
- Casos hospitalizados.
- Casos en UCI.
- Altas.
- Fallecidos.

Asimismo, se analizaron los formatos en los que las CC. AA., publicaron sus datos en sus portales de datos abiertos, siguiendo el esquema de las cinco estrellas de Tim Berners-Lee⁽¹³⁾ (figura 1).

Tim Berners-Lee, inventor de la Web, propuso en 2010 un sistema para medir la calidad de los datos abiertos según el nivel de reutilización que permiten. Este sistema, conocido como el esquema de 5 estrellas de datos abiertos, establece una serie de niveles, que van de menor a mayor facilidad de reutilización de los datos según la siguiente escala:

- Una estrella: este nivel se alcanza con la mera publicación de los datos, independientemente de su formato. Aquí nos encontraríamos con datos abiertos no estructurados (documentos en formato PDF, imágenes, vídeos, etc.), difíciles de manipular y de reutilizar.
- Dos estrellas: en este nivel los datos cuentan con un formato estructurado, para lo cual se ha utilizado un *software* propietario. Sería el caso de las hojas de cálculo de un conocido proveedor de *software*.
- Tres estrellas: en este caso los datos abiertos estructurados no utilizan un *software* propietario, sino un *software* libre, como *OpenOffice* o CSV.
- Cuatro estrellas: este nivel supone la existencia de datos estructurados adecuadamente con URI que identifican los recursos. Para ello utilizan el estándar RDF, recomendado por *World Wide Web Consortium* (W3C), que facilita la interoperabilidad.
- Cinco estrellas: es el nivel de mayor madurez, donde los datos, además de cumplir todos los requisitos anteriores, están enlazados con otros datos similares publicados por otras organizaciones. Esto dota de contexto y optimiza las búsquedas, facilitando la reutilización.

Figura 1
Esquema de las cinco estrellas de Berners-Lee.



<https://5stardata.info/en/>

RESULTADOS

El Ministerio de Sanidad publicó datos en el portal oficial de Datos abiertos del Gobierno de España (<https://datos.gob.es>), bajo el título *Evolución de enfermedad por el coronavirus (COVID-19)*, que incluía lo siguiente:

- Actualización diaria por el Centro de Coordinación de Alertas y Emergencias Sanitarias (formato PDF).
- Estudio Nacional de sero-epidemiología de la infección por SARS-CoV-2 en España ENE-COVID-19 (formato HTML).
- Informes COVID-19 del Centro Nacional de Epidemiología (formato PDF).
- Serie histórica de casos por Comunidad Autónoma (formato CSV).

- Serie histórica de casos por Provincia (formato CSV).
- Situación y evolución de la pandemia de COVID-19 en España (formato HTML).

En este portal sólo está actualmente disponible la última actualización de la situación del coronavirus en formato PDF, donde sí que aparecen muchos datos, pero en formato no reutilizable.

En los ficheros CSV *Serie histórica de casos por comunidad autónoma*, y *Serie histórica de casos por Provincia*, existen datos desde el 31 de enero de 2020 a día de hoy, pero sólo del número de casos, desglosado en función de la prueba diagnóstica realizada. Estas series históricas se recogen del SERLAB (Sistema español de resultados de laboratorio para pruebas diagnósticas de SARS-CoV-2), según la resolución de 19 de junio de 2020, por la que se establece

el Sistema de Información a través del cual se deberá remitir la información sobre pruebas diagnósticas de laboratorio necesaria para el seguimiento de la pandemia ocasionada por el COVID-19 al que se refiere el artículo 25 del *Real Decreto-Ley 21/2020, de 9 de junio*.

El panel que a día de hoy se muestra en el Ministerio de Sanidad se obtiene a partir de la declaración de los casos COVID-19 a la Red Nacional de Vigilancia Epidemiológica (RENAVE), a través de la plataforma informática vía Web SiViES (Sistema para la Vigilancia en España) que gestiona el Centro Nacional de Epidemiología (CNE). Esta información procede de la encuesta epidemiológica de caso que cada comunidad autónoma cumplimenta ante la identificación de un caso de COVID-19.

En el panel COVID-19 se presenta actualmente información geográfica e indicadores de evolución de la pandemia. Hasta el 10 de mayo, se incluyeron casos diagnosticados por una prueba diagnóstica positiva de infección activa, así como todos aquellos casos hospitalizados, ingresados en UCI y defunciones; a partir del 11 de mayo se incluyeron casos confirmados por PCR, o por pruebas de antígeno. La población utilizada para el cálculo de las tasas de incidencia procede de las cifras oficiales de población disponibles en el padrón municipal a 1 de enero del Instituto Nacional de Estadística de 2020.

Los CSVs sí son un formato reutilizable, por lo que interesaba saber el contenido de las columnas:

- Ccaa_iso o provincia_iso: el código ISO de la CA o la provincia de residencia.
- Fecha: desde el inicio de la pandemia hasta el 10 de mayo, la fecha de inicio de síntomas o, en su defecto, la fecha de diagnóstico menos 6 días. A partir del 11 de mayo, la fecha de inicio de síntomas, o en su defecto, la fecha de diagnóstico

menos 3 días, o la fecha de diagnóstico para los casos asintomáticos.

- Num_casos: el número de casos totales, confirmados o probables.
- Num_casos_prueba_pcr: el número de casos con prueba de laboratorio PCR o técnicas moleculares.
- Num_casos_prueba_test_ac: el número de casos con prueba de laboratorio de test rápido de anticuerpos.
- Num_casos_prueba_prueba_ag: Número de casos con prueba de laboratorio de test de detección de antígeno.
- Num_casos_prueba_elisa: Número de casos con prueba de laboratorio de serología de alta resolución (ELISA/ECLIA/CLIA).
- Num_casos_prueba_desconocida: el número de casos sin información sobre la prueba de laboratorio.

El Ministerio de Sanidad también publicó datos abiertos de capacidad asistencial (<https://www.mscbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/capacidadAsistencial.htm>). El fichero histórico de capacidad asistencial y actividad COVID de hospitales contiene los datos que se muestran a continuación y la serie comprende desde el 19 de agosto de 2020 hasta el día de actualización y se actualizaba diariamente a partir de las 15:30, excepto sábados, domingos y festivos.

- Para cada día (la fecha se refiere a la del envío de datos en que se informa sobre la situación del día previo), por provincia y tipo de unidad (hospitalización convencional, unidades de críticos con respirador y unidades de críticos sin respirador) se totalizaban los datos relativos a número de camas (número de camas

ocupadas por casos COVID-19 y número de camas ocupadas por casos no COVID-19), número de ingresos (COVID-19) y número de altas (COVID-19).

– Para cada fecha se totalizaban los datos enviados hasta las 15 horas por los hospitales; en el caso de que un hospital no hubiese actualizado sus datos a dicha hora se contabilizan los últimos datos disponibles de los tres días anteriores.

Cada comunidad autónoma, que dispuso de un portal de datos abiertos, publicó la información que consideró y en distintos formatos, como se puede ver en la **tabla 1**, donde se pudo observar si la CC. AA. publicó datos abiertos y en qué formatos estaban disponibles. Las CC. AA., de Extremadura y Galicia disponían de un portal de datos abiertos, pero en estos portales no se publicaban datos epidemiológicos en relación a la COVID-19.

Tabla 1
CC. AA. que disponen de datos abiertos, en qué formatos y su clasificación en función del esquema de estrellas de Berners-Lee.

Comunidad autónoma	Datos abiertos en CA	Formato disponible	Esquema de las cinco estrellas de Berners-Lee			
			★	★★	★★★	★★★★
Andalucía	Sí	PDF, XLS, PC-AXIS, CSV, OpenOffice	HTML	XLS		
Aragón	Sí	HTML, CSV y XLS	HTML	XML	CSV	
Asturias	Sí	Datos a CSV y XLS. Gráficas a PNG, CSV y XLS	PNG	XLS	CSV	
Islas Baleares	Sí	PDF. Mapa con la situación	PDF			
Canarias	Sí	Mapa en arc.gis. No descargable	-			
Cantabria	Sí	CSV y PNG	-		CSV	
Castilla-La Mancha	Sí	CSV	-		CSV	
Castilla y León	Sí	CSV, JSON, XLS. GeoJSON, Shapefile, KML	-	XLS	CSV, JSON, KML	
Cataluña	Sí	CSV, RDF, JSON, RSS, TSV y XML	-		CSV, JSON, KML	RDF
C. Valenciana	Sí	CSV, XML, JSON	-	XML	CSV, JSON	
Extremadura	No	-	-			
Galicia	No	-	-			
La Rioja	Sí	XLS, XML, CSV, JSON. Dispone de API	-	XLS	XML, CSV, JSON	
Comunidad de Madrid	Sí	CSV, JSON. Dispone de API	-			
Región de Murcia	Sí	PDF, CSV	PDF		CSV	
Navarra	Sí	CSV, ODS, XLS	-	XLS		
País Vasco	Sí	XML, XLS, CSV, JSON. Dispone de punto SPARQL.	-	XLS	CSV, JSON, KML	SPARQL
Ceuta y Melilla	No	-	-			
Ministerio de Sanidad	Sí	HTML, PDF, CSV	PDF		CSV	

Fuente: elaboración propia.

Atendiendo al tipo de formato utilizado en la publicación de los datos, pudimos observar que de las 15 CC. AA. que publicaron datos abiertos, 13 de ellas publicaron datos en formato CSV, lo que supuso un el 86% de las CC. AA. publicaron sus datos en CSV, el 33% de las CC. AA. publicaron los datos en XLS, 33% de las CC. AA. publicaron en JSON/GeoJSON, el 20% de las CC. AA. publicaron en PDF, el 13% publicaron en PNG, el 13% dispuso de una API, el 6% dispuso de un punto SPARQL, el 6% de las CC. AA. (que supuso sólo una CC. AA. en cada formato), publicaron en los formatos en PC, AXIS, HTML, KML, SHAPEFILE, RDF, TSV, y ODS (tabla 2). En la tabla 2 se muestran, asimismo, las frecuencias absolutas y relativas para cada formato. Castilla y León, Cataluña, La Rioja y País Vasco fueron las CC. AA., que más formatos aportaron.

En sus portales las CC. AA., publicaron información sobre casos confirmados de COVID-19, en muchos casos desagregada por provincias, distritos y municipios o zonas básicas de salud, además de otras variables.

Las variables más comunes que se publicaron en los portales de las CC. AA., se recogen en la tabla 3. Algunas CC. AA., publicaron datos por zonas básicas de salud, datos sobre todas las pruebas realizadas (PCR, anticuerpos, antígenos), y esta información se pudo mostrar de forma acumulada o diaria.

En relación a la fecha de actualización, en algunos portales se indicaba que la actualización era diaria, y luego al entrar en la información disponible se comprobó que no era así (Aragón). De las 15 CC. AA., que contaban con

Tabla 2
Formatos utilizados por las CC. AA.

Comunidad autónoma	Formatos																TOTAL
	CSV	XLS	JSON, GEOJSON	XML	PDF	PNG	API	HTML	PC-AXIS	ARC GIS	KML	SHAPEFILE	RDF	TSV	ODS	SPARQL	
Castilla y León	X	X	X								X	X					5
Cataluña	X		X	X									X	X			5
La Rioja	X	X	X	X			X										5
País Vasco	X	X	X	X												X	5
Andalucía	X	X					X	X									4
Aragón	X	X					X										3
Asturias	X	X				X											3
C. Madrid	X		X				X										3
Navarra	X	X												X			3
C. Valenciana	X		X	X													3
Cantabria	X					X											2
Region de Murcia	X				X												2
Islas Baleares					X												1
Canarias									X								1
Castilla-La Mancha	X																1
Frecuencia absoluta	13	7	7	4	3	2	2	3	1	1	1	1	1	1	1	1	
Frecuencia relativa	0,87	0,47	0,47	0,27	0,20	0,13	0,13	0,20	0,07	0,07	0,07	0,07	0,07	0,07	0,07	0,07	

Tabla 3
Contenidos que ofrecen las CC. AA. en sus portales de datos abiertos.

Comunidad autónoma	Actualización diaria	Datos acumulados por provincias	Casos confirmados por PCR	Datos sobre profesionales de riesgo	Datos sobre residencias	Datos sobre colegios	Casos hospitalizados	Casos en UCI	Altas	Fallecidos	Total datos publicados
Castilla y León	NO (de lunes a viernes)	SÍ	SÍ	SÍ	SÍ	SÍ	SÍ	SÍ	SÍ	SÍ	9
Andalucía	NO	SÍ	SÍ	SÍ	SÍ	NO	SÍ	SÍ	SÍ	SÍ	8
Aragón	SÍ	SÍ	SÍ	SÍ	NO	SÍ	SÍ	SÍ	SÍ	SÍ	8
Cantabria	SÍ	NA	SÍ	SÍ	SÍ	NO	SÍ	SÍ	SÍ	SÍ	7
Islas Baleares	SÍ	NA	SÍ	SÍ	SÍ	NO	SÍ	SÍ	SÍ	SÍ	7
Canarias	SÍ	SÍ	SÍ	NO	NO	NO	SÍ	SÍ	SÍ	SÍ	6
La Rioja	SÍ	NA	SÍ	NO	SÍ	NO	SÍ	SÍ	SÍ	SÍ	6
País Vasco	SÍ	SÍ	SÍ	NO	NO	NO	SÍ	SÍ	SÍ	SÍ	6
Navarra	SÍ (última actualización 29/05/2020)	NA	SÍ	SÍ	NO	NO	SÍ	SÍ	SÍ	SÍ	6
C. Valenciana	NO (de lunes a viernes)	NO	SÍ (acumulados)	NO	NO	NO	SÍ	SÍ	SÍ (acumulados)	SÍ (acumulados)	5
Castilla-La Mancha	SÍ (última actualización 5/05/2021)	Sólo n ^o casos por semanas	SÍ	NO	NO	NO	SÍ	SÍ	NO	SÍ	5
Cataluña	SÍ	NO	NO	NO	NO	SÍ	SÍ	SÍ	SÍ	SÍ	5
Asturias	SÍ	NA	SÍ	NO	NO	NO	SÍ	SÍ	NO	NO	3
Región de Murcia	SÍ	NA	SÍ	NO	NO	NO	SÍ (sólo gráfica)	SÍ (sólo gráfica)	NO	NO	3
C. Madrid	NO (semanal)	NA	SÍ (sin especificar técnica)	NO	NO	NO	NO	NO	NO	NO	1

un portal de datos abiertos, 11 CC. AA., tenían una actualización diaria de la información, 3 CC. AA., los actualizaban de lunes a viernes, y 1 CC. AA., los actualizaba semanalmente.

La única variable que publicaron todas las CC. AA., fue el número de casos hospitalizados. 5 CC. AA., publicaron datos acumulados por provincias, lo que supuso el 62,5%. 14 CC. AA., el 93,3% publicaron los casos confirmados por PCR. 6 CC. AA., el 40%, publicaron información relativa a los casos en profesionales de riesgo; 5 CC. AA., 33,3%, publicaron información relativa a los casos en residencias; 3 CC. AA., el 20%, publicaron información relativa a los casos en colegios; 14 CC. AA., el 93,3%,

publicaron los casos en UCI; 11 CC. AA., el 73,3%, publicaron el número de altas, y 12 CC. AA., el 80%, publicaron el número de fallecidos.

La fecha de inicio de la recogida de datos se muestra en la **figura 2**, aunque esta fecha varía dependiendo de la variable. Así las variables básicas, como número de casos, número de pruebas PCR, etcétera, tienen una fecha de inicio distinta de por ejemplo el número de altas o las hospitalizaciones. Aragón y Cataluña comenzaron publicando sólo el número de casos.

6 CC. AA., tenían estratificado el número de casos según tramo de edad y sexo. 1 CC. AA., sólo lo mostró según tramo de edad (**tabla 4**).

Figura 2**Línea temporal de publicación de datos abiertos en las CC. AA. para nº de casos y nº de PCR.****Tabla 4****Estratificación según tramo de edad y sexo del indicador número de casos.**

Comunidad autónoma	Estratificación de nº de casos	
	Por sexo	Por tramos de edad
Andalucía	SÍ	SÍ
Aragón	SÍ	SÍ
Asturias	NO	NO
Islas Baleares	SÍ	SÍ
Canarias	NO	NO
Cantabria	NO	NO
Castilla-La Mancha	NO	NO
Castilla y León	SÍ	SÍ
Cataluña	NO	NO
Comunidad Valenciana	SÍ	SÍ
La Rioja	NO	SÍ
Comunidad de Madrid	NO	NO
Región de Murcia	NO	NO
Navarra	NO	NO
País Vasco	SÍ	SÍ

En el portal datos.gob.es, las CC. AA., de Islas Baleares, Castilla y León, Castilla-La Mancha, Comunidad de Madrid, La Rioja, País Vasco y Comunidad Valenciana, publican también datos abiertos en varios formatos reutilizables (anexo II).

DISCUSIÓN

Existe una enorme diversidad entre las CC. AA., en cuanto a la publicación de la información de casos de COVID-19. Por ejemplo, en algunas CC. AA. la información se ofrece a través de notas de prensa en PDF con datos agregados por provincia, en otros sólo a nivel de CC. AA., en otras a través de servicios de *OpenData* o servicios web, en otras simplemente no se puede consultar. Además, en todos los casos el modelo de información no es el mismo en lo que se refiere a terminología, como casos confirmados, afectados, recuperados, fallecidos, etc. Es un verdadero esfuerzo recopilar todos los datos en un formato útil para trabajar (como CSV o JSON), con el añadido de que en cada CC. AA. los datos que se muestran son distintos, y la fecha desde la que se muestran los datos también varía, lo que pone de manifiesto una falta de estandarización. En esta situación, adquiere gran importancia la existencia de una coordinación entre las CC. AA., y el Gobierno de España, que permita el acceso a todos estos datos desde un único lugar y de forma homogénea, independientemente que cada CC. AA. lo publique en sus webs respectivas⁽¹⁴⁾.

A nivel nacional, se pone de manifiesto una gran existencia de datos y en formatos reutilizables, pero hay que ir a muchas fuentes para obtener información sobre la situación epidemiológica a nivel nacional. Un aspecto a tener en cuenta, es la forma en la que se contabilizan las distintas variables. En muchos informes, se muestran leyendas informando de este aspecto, por ejemplo, en cuanto a los fallecidos, en algunos casos se contabilizan los fallecidos con

sospecha de COVID-19, o con confirmación de COVID-19, según certificado de defunción. Lo mismo sucede con la fecha utilizada para la presentación de los casos, que puede ser la fecha de inicio de síntomas, o en su defecto, la fecha de diagnóstico menos tres días (mediana entre la fecha de inicio de síntomas y la fecha de diagnóstico).

Otro aspecto importante a tener cuenta es la consolidación de los datos. Por ejemplo, en los datos publicados por la Red Nacional de Vigilancia Epidemiológica, se resalta que *“todos los resultados son provisionales y deben interpretarse con precaución, porque se ofrece información de cada variable con la información disponible para cada una de ellas en el momento de la extracción de datos”*. Estos resultados pueden variar con cada informe por la dinámica propia de declaración de datos por las CC. AA. a la Red Nacional de Vigilancia Epidemiológica (RENAVE) a través de SiViES.

A nivel europeo, desde el Centro europeo para la prevención y el control de enfermedades (ECDC), a partir del 14 de diciembre de 2020 publican informes semanales para la situación de COVID-19 en todo el mundo y en la UE/EEE y el Reino Unido sobre el número de casos y muertes reportados en todo el mundo y agregados por semana todos los jueves. Los datos semanales estarán disponibles como archivos descargables en los siguientes formatos: XLSX, CSV, JSON y XML.

A la vista de la cantidad de datos que se publican a nivel mundial, cabe preguntarse: ¿Cómo podemos avanzar contra la pandemia? ¿Estamos progresando en la lucha contra la pandemia? Es cierto que para responder a estas preguntas necesitamos datos, pero incluso los mejores datos no son suficientes, y están lejos de ser perfectos. El objetivo de disponer de unos buenos datos es doble:

- Proporcionar datos confiables, globales y abiertos sobre cómo se está propagando la pandemia de COVID-19, qué impacto tiene la pandemia, y si las medidas que están tomando los países tienen efecto o no.
- Constituir una infraestructura que permita a los investigadores, y a todas las personas interesadas, navegar y comprender estos datos.

El intercambio temprano y abierto de datos, como publicaciones, conjuntos de datos, software, código y otro material científico, juega un papel esencial para permitir una respuesta rápida a epidemias o pandemias⁽⁴⁾. En sentido se ha visto un aumento de los datos abiertos, como se constata en el crecimiento de la publicación de acceso abierto, la disponibilidad del genoma de COVID-19, y la cantidad de iniciativas de datos abiertos para combatir COVID-19 de manera colaborativa.

Algunos países de la UE, han desarrollado distintas alternativas de datos abiertos, bien creando un repositorio en *GitHub*⁽¹⁵⁾, como en el caso de Italia, creando un *COVIDHub*⁽¹⁶⁾, como en Alemania. La Unión Europea también dispone de un portal de datos abiertos, en los que se publican datos a nivel europeo e internacional en los formatos CSV, XLSX, XLM, JSON y HTML. Más específicamente de ámbito europeo el Centro Europeo para la Prevención y Control de Enfermedades (ECDC), tiene una actualización diaria, y ofrece diversos mapas, tablas y cuadros de mando (más conocidos como *dashboards*). Destacable, el gráfico de evolución de casos para nada menos que 30 países, con su color correspondiente para cada uno en función de los casos notificados a 14 días por 100.000 habitantes.

La relación completa de países y la Unión Europea, y sus URL se encuentran en el [anexo III](#).

También existen diversos recursos a nivel internacional, y de iniciativas colaborativas ([anexo IV](#)). Una de las fuentes seguramente más conocidas y de las primeras en aparecer es el *mapamundi* de datos de John Hopkins University^(17,18), que ofrece datos actualizados varias veces al día sobre número de casos totales, personas fallecidas y personas recuperadas, a nivel mundial, por países y por regiones dentro de países.

La Universidad de Oxford dispone de un rastreador de respuestas gubernamentales (*Coronavirus government response tracker*)⁽¹⁹⁾, en el que puede consultar un repositorio en *GitHub*. Este rastreador recopila sistemáticamente información sobre varias respuestas políticas diferentes que los gobiernos han adoptado para responder a la pandemia en 18 indicadores, como el cierre de escuelas y las restricciones de viaje. Ahora tiene datos de más de 180 países.

Worldometer⁽²⁰⁾ es un proveedor de estadísticas globales de COVID-19 para muchas personas interesadas en todo el mundo.

La Organización Mundial de la Salud (OMS) también muestra datos, si bien desde el 31 de diciembre de 2019 hasta el 21 de marzo de 2020, la OMS recopiló el número de casos confirmados de COVID-19 y muertes a través de comunicaciones oficiales bajo el *Reglamento Sanitario Internacional*⁽²¹⁾, complementado con el seguimiento de los sitios web y cuentas de redes sociales de los ministerios de salud oficiales, desde el 22 de marzo de 2020, los datos mundiales se compilan a través de paneles de control específicos de la región de la OMS y/o los datos de recuento agregados se informan a la sede de la OMS a diario.

De cara a futuras pandemias, es imperativo que se establezca un sistema de notificación

estandarizado de informes para recopilar, visualizar y compartir sistemáticamente datos de alta calidad sobre enfermedades infecciosas emergentes y de notificación obligatoria en tiempo real⁽⁵⁾.

Los datos deben estar disponibles en un espacio y escala temporal que sea suficiente para que sea para la planificación y con fines de modelado. Un componente crítico de la propuesta es la democratización de los datos; toda la información recopilada (observando normas de privacidad necesarias) debe ser puesta a disposición del público de inmediato en el momento del lanzamiento, debe ser legible en varios formatos y basados en normas de datos abiertos.

La disponibilidad generalizada de datos en la actualidad durante la COVID-19 hace que sea difícil imaginar que se vuelva a vivir una epidemia o una pandemia sin la oportunidad de acceder a datos en tiempo real y a los conocimientos de investigación más recientes. Esta situación ha propiciado la puesta en marcha de nuevas estrategias de vigilancia motivadas por la COVID-19. Por nombrar algunos ejemplos, la Junta de Andalucía promulgó el Acuerdo de 18 de mayo de 2021, del Consejo de Gobierno, por el que se aprueba la formulación de la *Estrategia de Vigilancia y Respuesta en Salud Pública de Andalucía*⁽²²⁾, en el que se destaca el establecimiento de mecanismos de coordinación, comunicación y difusión accesibles de la información, e impulsar una infraestructura digital para el desarrollo de la salud pública, acorde con los avances tecnológicos, que permita proporcionar información accesible a todos aquellos que la necesitan, cuando la necesiten y de forma que les permita actuar en consecuencia. O la *Estrategia 2021-2027* del Centro Europeo para la Prevención y Control de Enfermedades (ECDC)⁽²³⁾, en la que se hace

mención a la salud digital y se plantea como estrategia ayudar con el análisis de datos avanzados, con la estandarización en la recopilación de datos y la introducción de métodos de agregación para facilitar las comparaciones y el análisis entre los países. A nivel nacional tenemos la *Estrategia de Vigilancia en Salud Pública* del Sistema Nacional de Salud.

Como limitaciones a esta investigación, destacar que puede darse el caso que alguna información que se muestra se haya modificado, o que en algunos portales de datos abiertos, ahora se muestren más indicadores de los que se han recogido en este artículo. E incluso que algunas de las plataformas que se indican, hayan quedado obsoletas.

Para concluir, reiterar la idea de que la pandemia de la COVID-19 tiene un impacto importante en todo el mundo, y en respuesta, muchos gobiernos e instituciones han estado creando y reuniendo una gran cantidad de datos para monitorear la propagación de la enfermedad y su impacto en la salud. Dado que estos datos están principalmente en manos de organismos públicos, a menudo se ponen a disposición del público como datos abiertos.

Sin embargo, encontrar datos no siempre está exento de desafíos, y al utilizar estos recursos se puede experimentar dificultades para navegar a través de tal abundancia de datos y encontrar comodidad con respecto a su confiabilidad. Crear, mantener y publicar datos requiere esfuerzo y tiempo. Cuando se dispone de las habilidades, herramientas y canales para publicar y descubrir datos, los datos se pueden recopilar y reutilizar más rápido. Además, la difusión de datos a través de diferentes fuentes con sus propias definiciones y estándares hace que sea difícil combinarlos.

BIBLIOGRAFÍA

1. Zhu N, Zhang D, Wang W *et al.* A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med.* 2020;382:727-733. Disponible en: <https://www.nejm.org/doi/full/10.1056/nejmoa2001017>
2. WHO (2020). Rolling updates on coronavirus disease (COVID-19). Author. Acceso: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen>
3. Torres-Salinas D. (2020). Ritmo de crecimiento diario de la producción científica sobre Covid-19. Análisis en bases de datos y repositorios en acceso abierto. *Profesional De La Información*, 29(2). <https://doi.org/10.3145/epi.2020.mar.15>
4. European Data Portal. 2020. The Power of Data in the Everlasting Battle: Humanity vs. Virus (1/2). 06/05/2020. Disponible en: <https://www.europendataportal.eu/en/impact-studies/covid-19/power-data-everlasting-battle-humanity-vs-virus-12>
5. Gardner L *et al.* A need for open public data standards and sharing in light of COVID-19. *The Lancet* (Agosto 2020). Disponible en <https://www.thelancet.com/action/showPdf?pii=S1473-3099%2820%2930635-6>
6. Xu B, Kraemer MUG; Open COVID-19 Data Curation Group. Open access epidemiological data from the COVID-19 outbreak. *Lancet Infect Dis.* 2020 May;20(5):534. doi: 10.1016/S1473-3099(20)30119-5. Epub 2020 Feb 19. PMID: 32087115; PMCID: PMC7158984. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7158984/>
7. Boletín Oficial del Estado. Real Decreto 1495/2011, de 24 de octubre, por el que se desarrolla la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público, para el ámbito del sector público estatal. BOE núm. 269, de 08/11/2011. Disponible en: <https://www.boe.es/buscar/act.php?id=BOE-A-2011-17560>
8. Directiva (UE) 2019/1024 del Parlamento Europeo y del Consejo de 20 de junio de 2019 relativa a los datos abiertos y la reutilización de la información del sector público (versión refundida). Disponible en: <https://www.boe.es/doue/2019/172/L00056-00079.pdf>
9. Arganda C. “Covid-19: los datos que aportan las CCAA, ni homogéneos ni comparables”. *DiarioFarma* (Abril 2020). Disponible en: <https://www.diariofarma.com/2020/04/05/covid-19-los-datos-que-aportan-las-ccaa-ni-homogeneos-ni-comparables>
10. World Health Organization (2020). Coronavirus press conference 11 February. World Health Organization. Disponible en: https://www.who.int/docs/default-source/coronaviruse/transcripts/who-audio-emergencies-coronavirus-full-press-conference-11feb2020-final.pdf?sfvrsn=e2019136_2
11. World Health Organization. 2020. SCORE for health data technical package: global report on health data systems and capacity, 2020. Disponible en: https://cdn.who.int/media/docs/default-source/world-health-data-platform/score/who_2021-01-31_global-report-score_tb_v2.pdf?sfvrsn=cf86a4fb_3&download=true
12. Boletín Oficial del Estado. Orden SND/234/2020, de 15 de marzo, sobre adopción de disposiciones y medidas de contención y remisión de información al Ministerio de Sanidad ante la situación de crisis sanitaria ocasionada por el COVID-19. BOE núm. 68, de 15 de marzo de 2020. Disponible en: https://www.boe.es/diario_boe/txt.php?id=BOE-A-2020-3702
13. Berners-Lee, Tim (2009). Linked Data. Disponible en: <http://www.w3.org/DesignIssues/LinkedData.html>
14. Martín Fernández A. Papel de los sistemas de información y la salud electrónica en la pandemia de COVID-19. Una llamada a la acción. *Rev Esp Salud Pública.* 2021; 95: 21 de enero e202101006. Disponible en: https://www.mscbs.gob.es/biblioPublic/publicaciones/recursos_propios/resp/revista_cdrom/VOL95/C_ESPECIALES/RS95C_202101006.pdf
15. GitHub. Disponible en : <https://github.com/>
16. COVID-19-Hospitalisierungen in Deutschland Disponible en: https://github.com/robert-koch-institut/COVID-19-Hospitalisierungen_in_Deutschland

17. John Hopkins University. (2020). Coronavirus resource center. Acceso: <https://coronavirus.jhu.edu/>
18. Coronavirus Covid-19 global cases by Center for System Science and Engineering. Center for Systems Science and Engineering at Johns Hopkins University. Disponible en: <https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6>
19. Coronavirus government response tracker. University of Oxford. Disponible en: <https://www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-tracker>
20. Worldometers. Disponible en: <https://www.worldometers.info/coronavirus/>
21. Organización Mundial de la Salud. 2005. Reglamento Sanitario Internacional Disponible en: <https://www.who.int/ihr/publications/9789241580496/es/>
22. Boletín Oficial de la Junta de Andalucía. Acuerdo de 18 de mayo de 2021, del Consejo de Gobierno, por el que se aprueba la formulación de la Estrategia de Vigilancia y Respuesta en Salud Pública de Andalucía. BOJA nº 96 de 21/05/2021. Disponible en: https://www.juntadeandalucia.es/boja/2021/96/BOJA21-096-00005-8631-01_00192511.pdf
23. European Centre for Disease Prevention and Control (ECDC). ECDC Strategy 2021–2027 To improve lives in Europe and globally apply scientific excellence, thus empowering Member States, the European Commission and other partners to drive public health policy and practice. Disponible en: <https://www.ecdc.europa.eu/sites/default/files/documents/ECDC-Strategy-2021-2027.pdf>
24. Leiva-Cepas F, Romero-Rodríguez E, Barroso Sevillano M. (2020). Las revistas científicas ante la pandemia por COVID-19 [Scientific journals in the face of the COVID-19 pandemic]. *Semergen*, 46 Suppl 1, 1–2. <https://doi.org/10.1016/j.semerg.2020.06.012>. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7328551/>
25. Roser M, Ritchie H, Ortiz-Ospina E. (2020). Coronavirus disease (COVID-19)—Statistics and research. Acceso: <https://ourworldindata.org/coronavirus>
26. Orden SND/234/2020, de 15 de marzo, sobre adopción de disposiciones y medidas de contención y remisión de información al Ministerio de Sanidad ante la situación de crisis sanitaria ocasionada por el COVID-19. Disponible en: https://www.boe.es/diario_boe/txt.php?id=BOE-A-2020-3702
27. Directiva 2003/98/CE del Parlamento Europeo y del Consejo, de 17 de noviembre de 2003, relativa a la reutilización de la información del sector público. Disponible en: <https://www.boe.es/doue/2003/345/L00090-00096.pdf>
28. Directiva 2013/37/UE del Parlamento Europeo y del Consejo, de 26 de junio de 2013, por la que se modifica la Directiva 2003/98/CE relativa a la reutilización de la información del sector público. Disponible en: <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:0001:0008:ES:PDF>
29. Dong E, Du H, Gardner L. (2020). An interactive web bases dashboard to track COVID-19 in real time. *The Lancet* Vol 20 May 2020. Disponible en: <https://www.thelancet.com/action/showPdf?pii=S1473-3099%2820%2930120-1>
30. European Centre for Disease Prevention and control. Disponible en: <https://www.ecdc.europa.eu/en/covid-19/data-collection>
31. European Data Portal. 2020. The Power of Data in Everlasting Battle: Humanity vs. Virus (2/2). 11/05/2020. Disponible en: <https://www.europeandataportal.eu/en/impact-studies/covid-19/power-data-everlasting-battle-humanity-vs-virus-22>
32. Open Data Watch. 2020. Data in the time of COVID-19. Open data, data use, and COVID-19. Disponible en: <https://opendatawatch.com/what-is-being-said/data-in-the-time-of-covid-19/>
33. Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público. Disponible en: <https://boe.es/buscar/act.php?id=BOE-A-2007-19814>
34. Jamali D, Barkemeyer R, Leigh J, Samara G. Open Access, Open Science, and Coronavirus: Mega trends with historical proportions. *Bus Ethics*. 2020 May 19. PMID: 32836923. Disponible en: <https://onlinelibrary.wiley.com/doi/10.1111/beer.12289>

Anexo I	
Fuentes de datos abiertos por comunidades autónomas.	
Comunidad autónoma	Fuente de datos abiertos
Andalucía	https://www.juntadeandalucia.es/institutodeestadisticaycartografia/badea/informe/anal?CodOper=b3_2314&idNode=42348
Aragón	https://opendata.aragon.es/datos/catalogo/dataset/publicaciones-y-anuncios-relacionados-con-el-coronavirus-en-aragon
Asturias	https://app.transparenciaendatos.es/v/#!/5eb4344e16b9fc465933d217
Islas Baleares	https://www.caib.es/sites/transparencia/es/informacia_diaria_del_servei_daepidemiologia_sobre_la_situacia_a_les_illes_balears_de_la_covid-19_agosttipoaifa/
Canarias	https://grafcan1.maps.arcgis.com/apps/opsdashboard/index.html#/156eddd4d6fa4ff1987468d1fd70efb6 . https://www3.gobiernodecanarias.org/sanidad/scs/content/dcb400c5-6504-11ea-9a8e-719d4b52bf6c/InformeCasosCOVID-19.pdf
Cantabria	https://www.icane.es/covid19/dashboard/home/home
Castilla-La Mancha	https://datosabiertos.castillalamancha.es/dataset/evolución-de-coronavirus-covid-19-en-castilla-la-mancha
Castilla y León	https:// analisis.datosabiertos.jcyl.es/explore/?sort=modified&refine.keyword=COVID-19
Cataluña	https://analisi.transparenciacatalunya.cat/es/Salut/Incidencia-de-la-COVID-19-a-Catalunya/623z-r97q/data
C. Valenciana	https://dadesobertes.gva.es/va/dataset/covid-19-total-acumulat-casos-confirmats-PCR-altres-persones-mortes-i-hospitalitzades-i-en-uci-2020
La Rioja	https://web.larioja.org/dato-abierto/datoabierto?filtros=%7B%22tema%22:%22Salud%22%7D#listado
C. Madrid	https://datos.comunidad.madrid/catalogo/dataset/covid19_tia_zonas_basicas_salud
Región de Murcia	https://www.murciasalud.es/pagina.php?id=458869
Navarra	https://gobiernoabierto.navarra.es/es/open-data/datos/datos-evolutivos-covid-19
País Vasco	https://opendata.euskadi.eus/catalogo/-/evolucion-del-coronavirus-covid-19-en-euskadi/
Ministerio de Sanidad	https://datos.gob.es

Anexo II		
Datos publicados en datos.gob.es por comunidades autónomas.		
Comunidad autónoma	Datos publicados en datos.gob.es	Formato
C. Valenciana	Datos de casos y personas fallecidas, por grupo de edad y sexo, acumulados desde el 31-01-2020 notificados en 2020 y 2021.	CSV
	Casos confirmados por PCR+ en los últimos 14 días y personas fallecidas por departamento de salud de la Comunitat Valenciana.	CSV
	Serie de casos con Pruebas Diagnósticas de Infección Activa (PDIA), positivas en la Comunitat Valenciana, según fecha de diagnóstico de laboratorio. Las PDIA incluyen dos pruebas de detección de infección activa, la PCR (siglas en inglés de <i>polymerase chain reaction</i> , reacción en cadena de la polimerasa) y el test de antígenos.	CSV
	Total acumulado de casos confirmados, PCR+, altas epidemiológicas, personas muertas y datos del día de personas hospitalizadas y en UCI.	CSV
	Serie de personas fallecidas en la Comunitat Valenciana según fecha de defunción.	CSV
Castilla y León	Personas vacunadas en residencia de mayores, tanto personal como residentes.	XLS, CSV
	Vacunas recibidas por las delegaciones territoriales de salud.	XLS, CSV
	Número reproductivo básico medio de COVID-19 por provincias.	XLS, CSV
	Información semanal sobre el número de existencias y el consumo del material sanitario. Información suministrada desde el 25 de septiembre de 2020.	XLS, CSV
	Información diaria y acumulada de PCR positivas en alumnos y profesorado de centros no universitarios, así como las aulas y centros en cuarentena. Desde el 9 de septiembre de 2020.	XLS, CSV
	Mortalidad por COVID-19 por tramos de edad y sexo en cada una de las provincias de Castilla y León. Datos absolutos y tasas por 1.000 tarjetas sanitarias. Actualización diaria.	XLS, CSV
	Porcentaje de personas fallecidas total (no solo por COVID19) por número de tarjetas sanitarias en cada zona básica de salud. Datos acumulados desde el 1 de marzo de 2020. Actualización diaria.	XLS, CSV
	Número de test rápidos y PCR realizadas en las Comunidad de Castilla y León. Actualización diaria.	XLS, CSV
	Porcentajes de test realizados y los resultados de los mismos en los profesionales sanitarios organizados por categorías profesionales.	XLS, CSV
	Datos de número de pacientes por provincia, sexo y tramo de edad. Actualización de lunes a viernes.	XLS, CSV
	Situación de los enfermos por coronavirus detectados en atención primaria en Castilla y León. Actualización de lunes a viernes. Evolución provincial del número de pacientes total y tasa por cada 100.000 personas.	XLS, CSV
	Situación enfermos en residencias por coronavirus. Datos correspondientes a 1.214 centros, incluye residencias, centros de personas discapacitadas y viviendas tuteladas.	XLS, CSV
Incidencia diaria de pacientes enfermos y de porcentaje de personas enfermas por número de tarjetas sanitarias en cada zona básica de salud. También indican los enfermos y los porcentajes para los últimos 7 y 14 días.	XLS, CSV	

Anexo II (continuación) Datos publicados en datos.gov.es por comunidades autónomas.		
Comunidad autónoma	Datos publicados en datos.gov.es	Formato
Castilla y León	Prevalencia coronavirus en cada zona básica de salud.	XLS, CSV
	Número de rastreadores COVID.	XLS, CSV
	Personas por lugar de fallecimiento. Datos correspondientes a 1.214 centros, incluye centros de personas discapacitadas y viviendas tuteladas.	XLS, CSV
	Datos diarios de confirmados, número de casos, altas y fallecimientos en hospitales debido al coronavirus (COVID-19) por provincias.	XLS, CSV
	Porcentaje de personas fallecidas COVID-19 (confirmados y compatibles con la enfermedad) por número de tarjetas sanitarias en cada zona básica de salud. Datos acumulados desde el 1 de marzo de 2020. Actualización diaria.	XLS, CSV
	Datos de ocupación hospitalaria tanto en planta como en UCI en todos los hospitales de Castilla y León. Actualización diaria.	XLS, CSV
	Porcentaje de test realizados y los resultados de los mismos en los profesionales sanitarios organizados por provincias.	XLS, CSV
	Relación de test PDIA realizadas a profesionales sanitarios por tramos de edad y sexo.	XLS, CSV
	Evolución del número de hospitalizados en planta, hospitalizados en unidades de críticos, altas y fallecimientos por hospital debidos al coronavirus (COVID-19) en Castilla y León. Actualización diaria.	XLS, CSV
	Situación profesionales sanitarios afectados por coronavirus. Desde 6 de abril de 2020.	XLS, CSV
La Rioja	Pruebas PCR de COVID-19 realizadas y resultados por localidades en las últimas 24 horas y acumuladas desde el inicio de la pandemia. Casos diagnosticados en los últimos 7 y 14 días por localidades. Incluye ubicación geográfica, número de positivos, negativos, totales y acumulados.	CSV, JSON, XLS, XML
	Horas transcurridas desde que se realizan las PCR de COVID-19 hasta que el laboratorio notifica el resultado.	CSV, JSON, XLS, XML
	Pruebas de COVID-19 realizadas en La Rioja agrupadas por grupo de edad, incluyendo datos acumulados a 14 días.	CSV, JSON, XLS, XML
	Incidencia de COVID-19 acumulada por Zona Básica de Salud, incluyendo los datos acumulados a los últimos 7 y 14 días.	CSV, JSON, XLS, XML
	Número de casos activos de COVID-19 en localidades de La Rioja de más de 100 hab.	CSV, JSON, XLS, XML
	Pruebas de COVID-19 realizadas por Zona Básica de Salud, incluyendo número de positivos, no positivos y datos acumulados.	CSV, JSON, XLS, XML
	Pruebas de COVID-19 realizadas por días, incluyendo número de positivos y no positivos.	CSV, JSON, XLS, XML
	Incidencia de COVID-19 acumulada por días, incluyendo los datos acumulados a los últimos 7 y 14 días.	CSV, JSON, XLS, XML
	Coronavirus: ritmo de reproducción (R0). Histórico hasta 17-08-2020.	CSV, JSON, XLS, XML
Coronavirus: confirmados PCR residencias de personas mayores de La Rioja. Histórico hasta 17-08-2020.	CSV, JSON, XLS, XML	
Islas Baleares	Mapa de casos activos confirmados por PCR.	HTML

Anexo II (continuación) Datos publicados en datos.gob.es por comunidades autónomas.		
Comunidad autónoma	Datos publicados en datos.gob.es	Formato
Cataluña	Datos semanales de COVID-19 por áreas de gestión asistenciales (AGA).	XML-APP, CSV, JSON, RDF-XML
	Datos semanales de COVID-19 por comarca.	XML-APP, CSV, JSON, RDF-XML
	Registro de defunciones por COVID-19 en Cataluña. Segregación por sexo y comarca.	XML-APP, CSV, JSON, RDF-XML
	Registro de defunciones por COVID-19 en Cataluña. Segregación por sexo y Área Básica de Salud.	XML-APP, CSV, JSON, RDF-XML
	Incidencia del COVID-19 en Cataluña. Del 24 de febrero al 22 de septiembre de 2020.	XML-APP, CSV, JSON, RDF-XML
	Registro de test de COVID-19 realizados en Cataluña. Segregación por sexo y edad.	XML-APP, CSV, JSON, RDF-XML
	Datos diarios de COVID-19 por comarca.	XML-APP, CSV, JSON, RDF-XML
País Vasco	Evolución del coronavirus (COVID-19) en Euskadi.	XML, JSON, XLSX, CSV
C. Madrid	COVID-19/TIA por Municipios y Distritos de Madrid. Casos confirmados, casos confirmados con infección activa y tasas de incidencia acumulada (TIA) por municipios y distritos de Madrid. Informe diario o semanal con datos de casos confirmados, casos confirmados con infección activa y tasas de incidencia acumulada de los últimos 14 días y desde el inicio de la epidemia (25 de febrero de 2020).	CSV, JSON, SHP
	COVID-19/TIA Zonas Básicas de Salud. Casos confirmados y tasa de incidencia (TIA) por Zonas Básicas de Salud. Desde 25 de febrero de 2020.	CSV, JSON, SHP
Castilla-La Mancha	Evolución de coronavirus COVID-19 en Castilla-La Mancha.	HTML

Anexo III Portales de datos abiertos en países de la Unión Europea.	
País	Fuente
Alemania	https://npgeo-corona-npgeo-de.hub.arcgis.com
Austria	https://info.gesundheitsministerium.at/opendata.html?lang=en
Eslovenia	https://covid-19.sledilnik.org/sl/data
Estonia	https://koroonaqaart.ee/en
Francia	https://www.data.gouv.fr/fr/reuses/tableau-de-bord-de-suiivi-de-lepidemie-de-coronavirus-en-france/
Holanda	https://data.rivm.nl/geonetwork/srv/dut/catalog.search#/metadata/2c4357c8-76e4-4662-9574-1deb8a73f724
Irlanda	https://data.gov.ie/dataset?q=covid&sort=score+desc%2C+metadata_created+desc
Italia	https://github.com/pcm-dpc/COVID-19
Lituania	https://open-data-ls-osp-sdg.hub.arcgis.com/datasets/45b76303953d40e2996a3da255bf8fe8_0?geometry=2.955%2C52.877%2C44.769%2C57.274
República Checa	https://onemocneni-aktualne.mzcr.cz/api/v2/covid-19
Unión Europea	https://data.europa.eu/euodp/en/data/dataset/covid-19-coronavirus-data-daily-up-to-14-december-2020
Centro europeo para el control de enfermedades	https://www.ecdc.europa.eu/en

Anexo IV	
Iniciativas colaborativas a nivel internacional.	
Institución	Fuente
John Hopkins University	https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6
COVID-19 Coronavirus Pandemic de Worldometer	https://www.worldometers.info/coronavirus/
Organización Mundial de la Salud	https://covid19.who.int
Proyecto COVID-19	https://covid19tracking.narrativa.com
Universidad de Oxford	https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker
Our world in data	https://ourworldindata.org/coronavirus;