

Ritmo de crecimiento diario de la producción científica sobre Covid-19. Análisis en bases de datos y repositorios en acceso abierto

Daily growth rate of scientific production on Covid-19. Analysis in databases and open access repositories

Daniel Torres-Salinas

Cómo citar este artículo:

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

Artículo recibido el 12-04-2020
Aceptación definitiva: 14-04-2020



Daniel Torres-Salinas

<https://orcid.org/0000-0001-8790-3314>

Universidad de Granada
Departamento de Información y Comunicación,
Medialab UGR, Unidad de Excelencia Iberlab, y
Ec3metrics spin off.
Campus Cartuja, s/n. 18071 Granada, España
torressalinas@go.ugr.es

Resumen

La comunidad científica se encuentra ante uno de sus mayores retos para resolver un problema sanitario de alcance global como es la pandemia del Covid-19. Esta situación ha generado un volumen de publicaciones sin precedentes, pero ¿cuál es el volumen, en términos de publicaciones, de la investigación sobre Covid-19? Conseguir una visión global sobre el crecimiento diario de la producción científica sobre Covid-19 en diferentes fuentes de información (*Dimensions*, *Web of Science Core Collection*, *Scopus*, *PubMed* y ocho repositorios) es el objetivo general de este trabajo de investigación. En relación a los resultados obtenidos, *Dimensions* indexa un total de 9.435 publicaciones (69% con revisión por pares y 2.677 preprints) muy por encima de *Scopus* (1.568) y *WoS* (718). Nos encontramos ante un fenómeno clásico bibliométrico de crecimiento exponencial ($R^2 = 0,92$). El ritmo de crecimiento a nivel global es de 500 publicaciones diarias en los últimos días y la producción de duplica cada 15 días. En el caso de *PubMed* el crecimiento semanal está en torno a las 1.000 publicaciones, tanto en *Dimensions* como en *LitCovid*. De los ocho repositorios analizados se sitúan a la cabeza *PubMed Central*, *medRxiv* y *SSRN*, a pesar de la enorme contribución de éstos las revistas siguen siendo el núcleo de la comunicación científica. Finalmente se ha establecido que tres de cada cuatro publicaciones sobre el Covid-19 están disponibles en acceso abierto. Estas cifras exigen una respuesta de los profesionales de la información ante una explosión de información sin precedentes que nos sitúa en el centro de la pandemia informativa.

Palabras clave

Comunicación científica; Análisis bibliométrico; Publicaciones científicas; Crecimiento exponencial; Acceso abierto; Bases de datos científicas; Repositorios; Covid-19; 2019-nCoV; SARS-CoV-2; Coronavirus; Información de salud; Pandemias.

Abstract

The scientific community is facing one of its greatest challenges in solving a global health problem: the Covid-19 pandemic. This situation has generated an unprecedented volume of publications. What is the volume of scientific publications on Covid-19? The general objective of this research work is to obtain a global vision of the daily growth of scientific production on Covid-19 in various databases (*Dimensions*, *Web of Science Core Collection*, *Scopus*, *PubMed*, and eight repositories). In relation to the results obtained, *Dimensions* indexes a total of 9,435 publications (69% with peer review and 2,677 preprints) well above *Scopus* (1,568) and *WoS* (718). This is a classic bibliometric phenomenon of exponential growth ($R^2 = 0.92$). The daily global growth rate is 500 publications and the production doubles every 15 days. In the case

of *PubMed* the weekly growth is around 1,000 publications. Of the eight repositories studied, *PubMed Central*, *medRxiv* and *SSRN* are the leaders, but despite their contribution, journals continue to be the core of scientific communication. Finally, it has been established that three of every four publications on the Covid-19 are available in open access. The information explosion which places us, information professionals, at the centre of the information pandemic demands a serious and coordinated response.

Keywords

Scholarly communication; Scientific communication; Bibliometric analysis; Scientific output; Exponential growth; Open access; Scientific databases; Repositories; Covid-19; 2019-nCoV; SARS-CoV-2; Coronavirus; Health information; Pandemics.

1. Introducción

La comunidad científica se encuentra ante uno de sus mayores retos para resolver un problema sanitario de alcance global como es la pandemia del Covid-19. La situación en la que nos encontramos inmersos requiere un esfuerzo científico colectivo que se refleja diariamente en la publicación de cientos de documentos y recursos de todo tipo (artículos, preprints, guías clínicas, protocolos, etc.) que involucran a todas las áreas del conocimiento. Probablemente estamos asistiendo a la mayor concentración de recursos científicos para la resolución de un problema concreto superando con creces otros precedentes como pudieran ser el proyecto *Manhattan* o la misión *Apolo*. En este contexto la eficacia del sistema de comunicación y publicación científica y sus elementos (revistas, revisores, bases de datos, repositorios) están siendo puestos a prueba ante una enorme cantidad de conocimiento generado en un breve lapso de tiempo, que no tiene precedentes (Kupferschmidt, 2020).

El primer problema inmediato al que ha debido enfrentarse el universo de la publicación es la avalancha de artículos y preprints y la necesidad que éstos sean accesibles. Una de las respuestas colectivas por parte de las editoriales 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 Covid-19. Así, las multinacionales de la edición de como *Elsevier*, *Springer* o *Emerald* han adoptado esta política. Las grandes revistas científicas, especialmente las de biomedicina, también están haciendo frente a este escenario compartiendo todas sus publicaciones: *JAMA*, *BMJ*, *Science*, *Oxford*, *Cambridge* o *New England*. Asimismo, los repositorios, por su rapidez y eficacia en transmitir la información científica, están en el centro de las miradas y la mayor parte facilita en sus páginas de entrada consultas rápidas a sus preprints, como ocurre en *arXiv*, y otros, como *Zenodo*, han creado comunidades que recopilan los trabajos más relevantes. Ante esta avalancha surge una pregunta básica: ¿cuál es el volumen, en términos de publicaciones, de la investigación sobre el Covid-19?

El sistema de comunicación y publicación científica está siendo puesto a prueba ante la enorme cantidad de conocimiento generado en tan breve lapso de tiempo

Debido a la multidisciplinariedad del frente de investigación del Covid-19 y la multitud de fuentes de datos, una cuantificación urgente y una caracterización global, aunque sea de carácter descriptivo, ayudaría a los profesionales de la información a comprender y visualizar un fenómeno informativo al que tenemos que enfrentarnos en los próximos meses. Si bien es cierto que ya se están realizando los primeros análisis bibliométricos, éstos están centrados en analizar las fuentes tradicionales como *Web of Science*, *Scopus* o *PubMed* y en describir a los productores de las publicaciones (por ejemplo, Chahrour, 2020; Hossain, 2020; Alba, 2020). Sin embargo estos trabajos obvian la importancia de otras fuentes como son los repositorios en sus cuantificaciones y ofrecen una visión fragmentada del fenómeno. Conseguir una visión global sobre el crecimiento diario de la producción científica sobre el Covid-19 en diferentes fuentes es el objetivo general de este trabajo. Más específicamente se han establecido dos objetivos:

- Cuantificar de forma global cuál es el volumen de la producción científica sobre Covid-19 en varias bases de datos multidisciplinares como *Web of Science*, *Scopus*, y *Dimensions*, y especializadas como *Medline*, y determinar su crecimiento diario.
- Cuantificar del mismo modo el número de preprints diarios que se publican en los repositorios y describir el fenómeno del acceso abierto

Los resultados que se alcancen en este trabajo no sólo tienen un valor descriptivo de un fenómeno informativo singular, sino que permitirá a los profesionales de la información tener un mapa objetivo y global de aquellas de las fuentes de información y bases de datos más útiles para enfrentarse al Covid-19 y asesorar a los investigadores.

Los repositorios, por su rapidez y eficacia en transmitir la información científica, están en el centro de las miradas

2. Material y métodos

Para la realización de este trabajo se han utilizado varios recursos. El primero es *Dimensions*, una base de datos multidisciplinar que integra publicaciones, datos, ensayos clínicos, etc. Indexa contenido de revistas científicas, bases de datos como *PubMed* y preprints procedentes de ocho repositorios en acceso abierto. No se ha consultado *Dimensions* directamente, sino que se ha hecho del dataset público con las referencias a todas las publicaciones sobre Covid-19:

<https://tinyurl.com/qqzncmv>

Este fichero fue sometido a un proceso de normalización en diversos campos.

También se han consultado las bases de datos *Web of Science Core Collection* completa y *Scopus*. La recuperación de información en *Dimensions*, *WoS* y *Scopus* se ha realizado mediante la ecuación de búsqueda ofrecida por *Dimensions* en su dataset:

PUBLICATION YEAR: 2020; FULL DATA SEARCH: "2019-nCoV" OR "Covid-19" OR "SARS-CoV-2" OR ("coronavirus" OR "corona virus") AND (Wuhan OR China)

Tal y como puede observarse, las publicaciones se refieren exclusivamente a Covid-19 y no se incluyen otros términos tangenciales. Se ha hecho uso de la misma ecuación en las tres bases de datos. Las consultas se realizaron durante los días 11 y 12 de abril de 2020.

Una cuarta fuente que se ha empleado es *LitCovid*, una web centralizada de publicaciones sobre Covid-19, curada y mantenida por la *National Library of Medicine / National Institutes of Health (NIH)* (Chen, Allot & Lu, 2020). Contiene publicaciones verificadas sobre Covid-19 sin los errores que pueda provocar una consulta mal definida. Esta fuente ayuda a cuantificar de forma realista las publicaciones indexadas en *PubMed* y a verificar las publicaciones de *PubMed* indexadas en *Dimensions*. Los resultados que se presentan son indicadores básicos de producción, considerando todos los tipos documentales.

3. Resultados

3.1. Bases de datos científicas

A la hora de resolver un problema bibliográfico el primer lugar donde acudir son las grandes bases de datos científicas *Web of Science* y *Scopus*. En la tabla 1 se presentan los resultados de la búsqueda. *Web of Science* tiene actualmente indexadas 764 publicaciones y *Scopus* 1.568, una cantidad no demasiado grande que se reduce especialmente si consideramos solamente los tipos citables. Sin embargo, no es posible con estas fuentes trazar el crecimiento diario ya que no ofrecen el día exacto de publicación o al menos de inclusión en la base de datos. Sin embargo si se puede limitar la búsqueda a los registros ingresados durante las últimas semanas. Si utilizamos el filtro semanal los dos productos nos ofrecen resultados coincidentes mostrando que el 50% de las publicaciones se han incorporado en las dos últimas semanas, es decir los días comprendidos entre el 30 de marzo y el 12 de abril. En el caso de *WoS* el 32% son de la última semana (6-12 de abril), una cifra que se reduce al 20% en *Scopus*. En cualquier caso, si bien estas cifras son groseras y limitadas por la fecha de adición de cada base de datos, ya nos apuntan a un crecimiento importante y muy concentrado en los últimos días en relación al tema del Covid-19. Para aproximarnos con mayor concreción a dicho crecimiento hemos utilizado también *Dimensions*, una base de datos mixta con fuentes de información de distinta naturaleza.

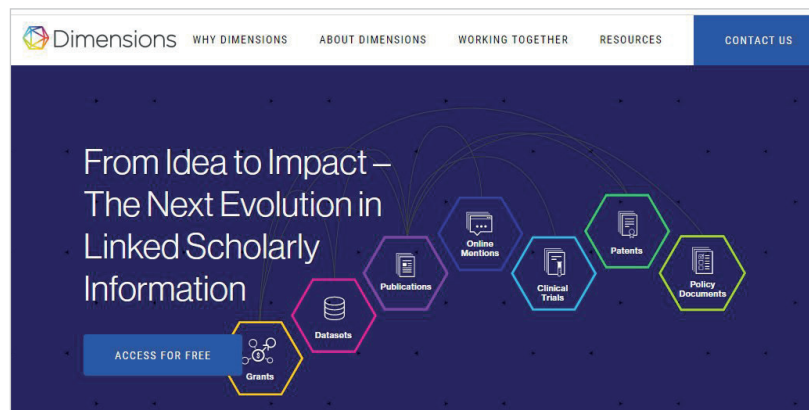


Figura 1. *Dimensions*, parte de la empresa *Digital Science*, integra información científica de diversos tipos y fuentes.

<https://www.dimensions.ai>

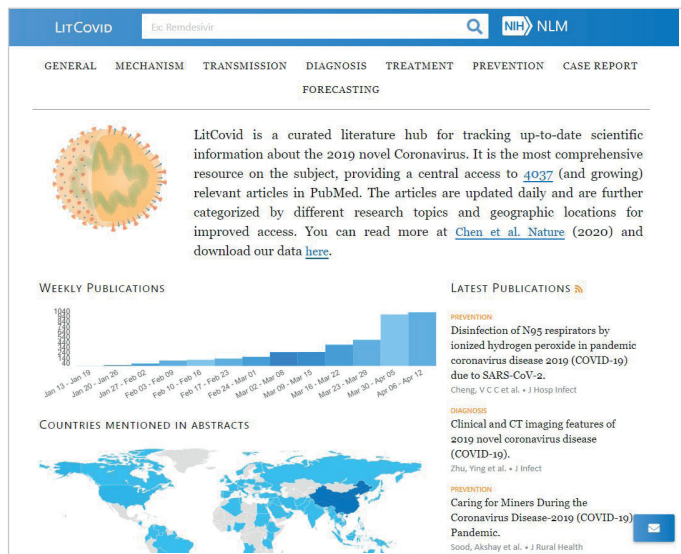


Figura 2. *LitCovid* selecciona información científica actualizada sobre el nuevo coronavirus de 2019. Producido por la *National Library of Medicine* de los Estados Unidos.

<https://www.ncbi.nlm.nih.gov/research/coronavirus>

Empleando el dataset de *Dimensions* el panorama cambia sustancialmente (gráfico 1). Existe un total de 9.435 publicaciones científicas incluyéndose tanto publicaciones en revistas como en repositorios (por ejemplo: *PMC*, *arXiv*...) y mezclando, por tanto, contribuciones que han pasado el proceso de revisión con otras que no lo han hecho como el caso de los preprints. Esta base de datos hace las funciones más de agregador de fuentes que de base de datos tradicional. La fusión de múltiples fuentes y la inclusión del campo de la fecha de indexación en *Dimensions* permiten trazar un retrato más preciso del crecimiento diario de la producción científica sobre Covid-19. En el momento en que se declaró la emergencia sanitaria internacional por la *World Health Organization*, el 30 de enero, *Dimensions* ya contaba con 340 publicaciones indexadas. En un mes esa cifra se disparó y alcanzó un corpus significativo de 1.466 publicaciones (24 de febrero). Desde ese momento el número de publicaciones se duplica con un promedio de 15 días, como ocurrió entre el 14 de marzo y el 30 de marzo cuando se pasaron de las 3.041 publicaciones a las 6.022. Nos encontramos con un crecimiento cuyo mejor ajuste es el modelo de crecimiento exponencial con un valor de R^2 de 0,92. En cuanto a la distribución del total de publicaciones hay que señalar que 6.549 corresponden a publicaciones en revistas científicas estando 5.715 de las mismas indexadas en los productos de los *NIH PubMed* y *PubMed Central*. Del conjunto, 2.677 (31%) corresponden a repositorios en acceso abierto de preprints.

Tabla 1. Producción científica sobre Covid-19 en las bases de datos *Scopus* y *Web of Science Core Collection* a fecha de 12/04/2020

| | <i>Web of Science Core Collection</i> | | <i>Scopus</i> | |
|---|---------------------------------------|----------|---------------|----------|
| Total de publicaciones | 718 | | 1.568 | |
| Semanas | n | % | n | % |
| Añadidas en la última semana (6 abril-12 abril) | 231 | 32 | 310 | 20 |
| Añadidas en las últimas 2 semanas (30 de marzo-12 abril) | 384 | 53 | 757 | 48 |
| Añadidas en las últimas 4 semanas (16 de marzo-12 de abril) | 575 | 80 | 1.042 | 66 |

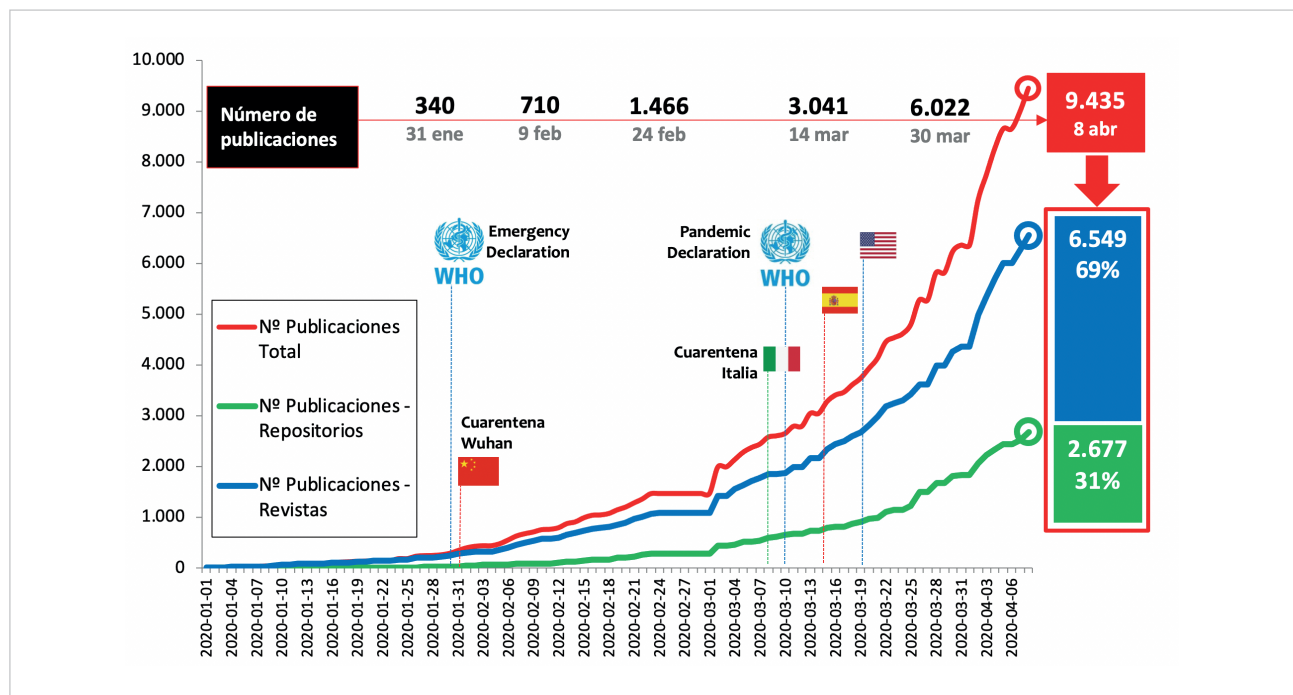


Gráfico 1. Evolución del número de publicaciones acumuladas sobre Covid-19 en *Dimensions* clasificadas según su fuente (repositorios y revistas)

Durante los cinco días, comprendidos entre el 2 y el 6 de abril, se ha añadido un promedio de 514 trabajos diarios a *Dimensions*. Para complementar estos datos una de las bases de datos fundamentales es *PubMed*. Para analizar el número de publicaciones en dicha base de datos hemos tomado como referencia todos aquellos registros del dataset de *Dimensions* que contaban con el identificador PubMedId. Una vez aislados dichos registros obtenemos un total de 4.291 y si los contabilizamos desde el 20 de enero se reducen a 3.844 (tabla 2). Los datos de *Dimensions* los hemos comparado con los ofrecidos oficialmente por *LitCovid*, una fuente de los *NIH* que extrae y selecciona los artículos más relevantes de *PubMed*. En *LitCovid* la cifra total de publicaciones es de 3.922, por tanto, ambas bases de datos recogen un número similar de publicaciones. Aunque en las primeras semanas hay diferencias sustanciales en el número de trabajos indexados a favor *Dimensions* éstas se reducen las últimas semanas, especialmente las comprendidas entre el 23 marzo-29 marzo y el 30 marzo-5 abril cuando la diferencia es del 7% y 4% respectivamente. La cobertura similar de dos fuentes diferentes nos permite afirmar que actualmente la producción científica sobre Covid-19 en *PubMed* se sitúa en torno a las 1.000 publicaciones semanales. El cre-

“ A principios de abril de 2020 la producción científica sobre Covid-19 en *PubMed* está en torno a las 1.000 publicaciones semanales ”

cimiento de la producción reproduce el esquema exponencial ya señalado con los datos de el gráfico 1, en ambos casos el gran salto se produce en la semana del 30 de marzo al 5 abril cuando *Dimensions* y *LitCovid* incrementan su número de registros en un 91% y un 97% respectivamente en relación a la semana anterior.

Tabla 2. Evolución de las publicaciones sobre Covid-19 en la base de datos *PubMed* a través de los registros recogidos en *Dimensions* y en *LitCovid*

| Semana | Número de publicaciones | | | Incrementos semanales | |
|-------------------------|---------------------------------|-----------------------------------|---|---------------------------------------|---|
| | Nº pubs. <i>PubMed LitCovid</i> | Nº pubs. <i>PubMed Dimensions</i> | % diferencia a favor de <i>Dimensions</i> | % crecimiento semanal <i>LitCovid</i> | % crecimiento semanal <i>Dimensions</i> |
| 20 enero - 26 enero | 13 | 68 | 81 | | |
| 27 enero - 02 febrero | 45 | 103 | 56 | 246 | 51 |
| 03 febrero - 09 febrero | 102 | 172 | 41 | 127 | 67 |
| 10 febrero - 16 febrero | 116 | 194 | 40 | 14 | 13 |
| 17 febrero - 23 febrero | 139 | 235 | 41 | 20 | 21 |
| 24 febrero - 01 marzo | 173 | 300 | 42 | 24 | 28 |
| 02 marzo - 08 marzo | 263 | 284 | 7 | 52 | -5 |
| 09 marzo - 15 marzo | 266 | 358 | 26 | 1 | 26 |
| 16 marzo - 22 marzo | 406 | 562 | 28 | 53 | 57 |
| 23 marzo - 29 marzo | 499 | 539 | 7 | 23 | -4 |
| 30 marzo - 05 abril | 985 | 1.029 | 4 | 97 | 91 |
| 06 abril - 12 abril | 915 | *317 | -- | -1 | 0 |
| Totales | 3.922 | 3.844 | -- | -- | -- |

*Datos hasta el 8 de abril de 2020

3.2. Datos de repositorios y acceso abierto

En el gráfico 2 se presentan los datos de las publicaciones indexadas en ocho repositorios. El total de publicaciones es de 4.075; de éstas, 1.398 (34%) corresponden a *PMC* que presenta una actividad significativa desde los primeros días de la pandemia. Aunque *PMC* no es un repositorio de preprints como el resto, sí ofrece todos los contenidos de sus revistas en acceso abierto. Al ser en su mayor parte con revisión por expertos se corresponden con la ruta dorada. Los siete repositorios restantes sólo indexan preprints y pertenecen, por tanto, a la ruta verde.

“ Tres de cada cuatro publicaciones que aparecen en revistas sobre Covid-19 están en acceso abierto, a las que habría que sumar los preprints de los repositorios ”

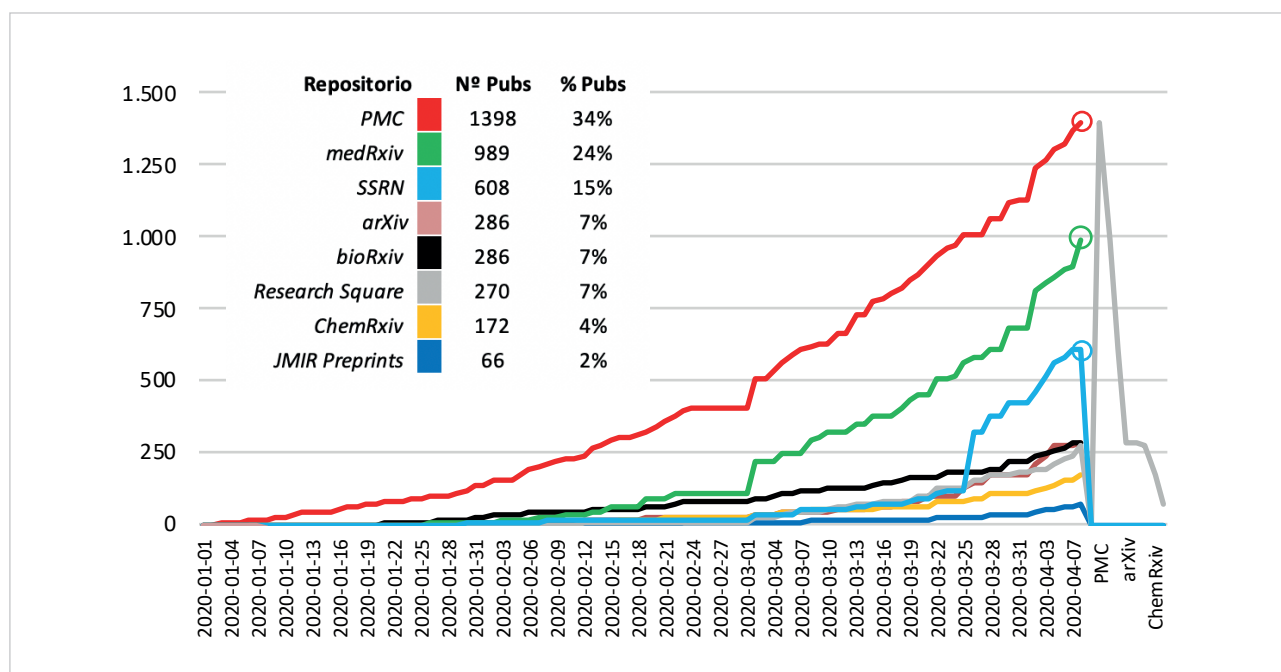


Gráfico 2. Evolución del número de publicaciones acumuladas sobre el Covid-19 en ocho repositorios

El siguiente repositorio con mayor cobertura es *medRxiv* con 998 (24%) publicaciones que empieza a publicar contenido de forma relevante a partir de marzo.

“*Dimensions* ha resultado ser un recurso más exhaustivo al integrar diversas fuentes”

La tercera fuente que contribuye a la producción sobre Covid-19 es el *SSRN* con 608 (15%) preprints.

Todas estas fuentes presentan crecimientos que se acercan al modelo exponencial.

Otros cinco repositorios (*arXiv*, *bioRxiv*, *Research Square*, *ChemRxiv* y *JMIR Preprints*) contribuyen con 1.080 publicaciones (27%). Todos ellos empezaron a mediados de marzo a publicar contenido de forma relevante y creciente.

Finalmente consideramos relevante ofrecer las cifras del acceso abierto en el frente de investigación del Covid-19 ya que son espectaculares y singulares. *Dimensions* indica que un total de 6.357 publicaciones están en acceso abierto, es decir el 67%. Si solo se consideran las publicaciones en revistas científicas se encuentra un total de 3.832 (el 60%). Esta cifra se eleva al 67% si solo se consideran las revistas de *PubMed* (2.843 de un total 4.291). En *Web of Science* el porcentaje de artículos en acceso abierto se incrementa al 97% ya que casi todas las grandes editoriales científicas ofrecen abiertos sus contenidos sobre el tema del Covid-19, tal y como indicamos en la introducción. En *Scopus* las cifras son también elevadas y el porcentaje de publicaciones en acceso abierto es del 84%. Esto indica que aproximadamente tres de cada cuatro publicaciones que aparecen en revistas sobre el Covid-19 están en acceso abierto a las que habría que sumar todos los preprints de los repositorios.

4. Consideraciones finales

En relación con los resultados obtenidos se ha determinado el ritmo de crecimiento que a nivel global es de 500 publicaciones diarias en *Dimensions*. En el caso de *PubMed* el crecimiento semanal se sitúa en torno a las 1.000 publicaciones.

“*PubMed* se mantiene como la gran base de datos de referencia”

Por tanto nos encontramos ante un fenómeno clásico bibliométrico de crecimiento exponencial ($R^2 = 0,92$). Desde el punto de vista de las bases de datos, los modelos más tradicionales de bases de datos se ven perjudicados ya que a veces tardan en incorporar los registros y puede que no sean lo ágiles e inmediatos que requiere un tema como el Covid-19. *Dimensions* ha resultado ser un recurso más exhaustivo al integrar diversas fuentes y por tanto lo convierten en una fuente más valiosa desde el punto de vista informativo. Asimismo, como no podía de otra manera, *PubMed* se mantiene como la gran base de datos de referencia con una respuesta en cobertura y acceso considerables.

En relación con el acceso abierto, en sus dos vertientes, está jugando un papel fundamental. Por un lado, los editores científicos se han volcado con la puesta disposición de los trabajos sobre Covid-19, alcanzándose porcentajes de acceso abierto inéditos en revistas. Los repositorios también están ocupando un lugar central, si bien no sustituyen a las revistas científicas, ya que las publicaciones de la modalidad ruta verde suponen sólo el 30% de las publicaciones. Este estudio ha contribuido, por tanto, a tener una imagen amplia sobre el volumen de la producción científica del Covid-19 y sus fuentes más relevantes. Hemos de terminar reseñando que el reto actual no sólo atañe al ámbito médico, sino que también exige una respuesta de los profesionales de la información ante una explosión de información sin precedentes que nos sitúa en el centro de la pandemia informativa.

“La actual explosión de información sitúa a los profesionales de la información en el centro de la pandemia informativa”

5. Bibliografía

Alba-Ruiz, Rubén (2020). “Covid-19, coronavirus pandemic: aproximación bibliométrica y revisión de los resultados”. *Zenodo*, 31 marzo.

<http://doi.org/10.5281/zenodo.3734062>

Chahrour, Mohamad; Assi, Sahar; Bejjani, Michael; Nasrallah, Ali A.; Salhab, Hamza; Fares, Mohamad Y.; Khachfe, Hussein H. (2020). “A bibliometric analysis of Covid-19 research activity: A call for increased output”. *Cureus*, v. 12, n. 3, e7357.

<http://doi.org/10.7759/cureus.7357>

Chen, Qingyu; Allot, Alexis; Lu, Zhiyong (2020). “Keep up with the latest coronavirus research”. *Nature*, v. 579, n. 7798, pp. 193.

<http://doi.org/10.1038/d41586-020-00694-1>

Hossain, Md Mahbub (2020). “Current status of global research on novel coronavirus disease (Covid-19): A bibliometric analysis and knowledge mapping”. *SSRN*, April 2nd.

<http://doi.org/10.2139/ssrn.3547824>

Kupferschmidt, Kai (2020). “Preprints bring ‘firehose’ of outbreak data”. *Science*, v. 367, n. 6481, pp. 963-964.

<http://doi.org/10.1126/science.367.6481.963>